Norwegian University of Life Sciences
School of Economic and Business

Philosophiae Doctor (PhD)
Thesis 2022:65

Seed security in the presence of climate shocks and socioeconomic inequality in Africa

Frøsikkerhet i nærvær av klimasjokk og sosioøkonomisk ulikhet i Afrika

Clifton Makate

Seed security in the presence of climate shocks and socioeconomic inequality in Africa

Frøsikkerhet i nærvær av klimasjokk og sosioøkonomisk ulikhet i Afrika

Philosophiae Doctor (Ph.D.) Thesis

Clifton Makate

Norwegian University of Life Sciences
School of Economic and Business

Ås (2022)

Thesis number: 2022:65
ISSN: 1894-6402
ISBN: 978-82-575-2017-5

Dedication

This work is dedicated to my family, friends, and parents.

Acknowledgments

The research presented in this thesis is the result of several contributions and sacrifices made by various individuals and institutions, and my significant gratitude goes to them all.

I extend my appreciation and deep thanks to my advisors Prof. Arild Angelsen (main advisor), Prof. Stein T. Holden, and Associate Prof. Ola T. Westengen (co-advisors), whose intensive and constructive comments on my work and life, in general, have shaped my career and made it possible to complete this work. I am very grateful for their tireless support, invaluable academic insights, and encouragement at all stages of my Ph.D. research.

My gratitude goes to the Norwegian University of Life sciences (NMBU) for awarding me such a competitive Ph.D. Research fellowship position to carry out this research and the funding from the Research Council of Norway (RCN) through the project: *Access to seeds: from Emergencies to Seed System development (ACCESS) (*project number RCN 288493) from which I carried out my research. I am also grateful to all the *ACCESS* project team members and partners for the constructive comments during project meetings and workshops that have helped me shape and improve my work. Many thanks to Dr. Munyaradzi Mutenje, Dr. Moti Jaleta, and Professor Ruth Haug for the comments and questions that helped me shape this research. Many thanks also to my colleague Teshome Hunduma Mulesa for the random and inciteful interactions and help with proofreading part of my work. I am also grateful to the World Bank (and partners) for making the Living Standards Measurement Study - Integrated Surveys on Agriculture (*LSMS-ISA*) data for Malawi, Tanzania, and Ethiopia available for this research.

My appreciation and thanks to Prof. Frode Alfnes, Kirsti Pettersen, Berit Pettersen, Lise Thoen, Reidun Aasheim, Hanne Marie Fischer, and all the other staff members at NMBU School of Economics and Business for their unwavering support. These people created an enabling environment that made my settling, studies, and stay in Norway run smoothly, and I greatly appreciate their efforts.

Lastly, I also extend my gratitude to my immediate family (Rutendo, Kayla, and Kiara Makate) and friends for their love and support that made it possible to work on this project.

Table of Contents

List of papers

Paper I: Clifton Makate, Arild Angelsen, Stein T. Holden, & Ola T. Westengen (2022). Crops in crises: shocks shape smallholders' diversification in rural Ethiopia*, World Development,* Vol. 159*,* 106054, pp 1-17, doi: https://doi.org/10.1016/j.worlddev.2022.106054

Paper II: Clifton Makate, Arild Angelsen, Stein T. Holden, & Ola T. Westengen (2022). *Rainfall shocks and inequality have heterogeneous effects on farmers' seed purchase decisions in East Africa (submitted to a journal)*.

Paper III: Clifton Makate, Arild Angelsen, Stein T. Holden, & Ola T. Westengen (2022). *Evolution of farm-level crop diversification and response to rainfall shocks in smallholder farming: Evidence from Malawi and Tanzania (submitted to a journal)*.

Paper IV: Clifton Makate, Arild Angelsen, Stein T. Holden, & Ola T. Westengen (2022). *Smallholder access to purchased seeds in the presence of pervasive market imperfections and rainfall shocks: Panel Data Evidence from Malawi and Ethiopia (submitted to a journal)*.

Summary

Access to diverse and well-adapted seeds is vital in helping farmers raise crop yields, agricultural incomes, and food security, reducing poverty and inequalities, and improving human well-being. The realization of benefits attached to seed as an asset in agro-based communities requires that households are seed secure. Household seed security as a concept requires that farmers and farming communities have ready access to adequate quantities of quality seed and planting materials of crop varieties, adapted to their agro-ecological conditions and socioeconomic needs, at planting time, at all times. The access dimension of seed security needs better empirical and theoretical underpinning to realize the concept's potential for short-term and long-term seed sector development - a policy goal. This thesis contributes to this scholarly literature. The main objective is to investigate opportunities and constraints to seed access and utilization in smallholder farming in the context of increasing climate variability, socioeconomic inequality, and pervasive transaction costs that characterize seed markets, focusing on three African countries: Malawi, Tanzania, and Ethiopia. Four independent but related empirical research papers address this objective.

The first paper focuses on Ethiopia and household's behavioral responses in their local and improved seed use and crop diversification decisions to recent exposure to covariate climate shocks and idiosyncratic household shocks. The second paper focuses on seed purchase, an important dimension for understanding seed access, and explores the influence of previous exposure to drought shocks, gender, and wealth endowments on the likelihood and extent of purchasing seeds of key crops in Malawi, Tanzania, and Ethiopia. Paper three addresses the evolution of crop diversification and response to rainfall shocks in Malawi and Tanzania. Lastly, paper four analyses the dynamic nature of transaction costs in seed markets that can facilitate (or constrain) seed access through purchase from available markets in Malawi and Ethiopia.

A few key conclusions and implications emanate from this research. Crop diversification, sourcing seeds off-farm through purchase, and improved seed varieties are strategies for buffering against production risk in smallholder agriculture. However, socioeconomic disadvantages and recurrent rainfall (drought) shocks make households more seed insecure and make adaptation more difficult. Pervasive transaction costs characterize access to seed through the market and reduce household seed security. However, accumulating market knowledge, experience, and networks helps smallholder farmers marginally reduce transaction costs in access to purchased seeds over time, supporting seed diversity and adaptation to rainfall shocks. The conclusions from this research support the call for the promotion of integrated approaches to seed system development.

Besides using farmer-saved seeds and local seed varieties, farmers also use improved seeds and seeds purchased through both formal and informal channels. As formal channels represent the most likely primary source for new varieties, policy should address the inequality in access to such channels through supply-side measures such as increasing the production of affordable quality-controlled seeds, plus reducing barriers to input market access and demand-side measures such as social protection programs, enhanced extension, and information access. At the same time, informal seed systems continue to be the backbone of the seed systems farmers use, and seed policies and regulations should enable the co-existence of formal and informal systems. An integrated seed systems approach supported with policies that will reduce inequities in accessing seed from commercial sources will improve the accessibility of improved and local seed varieties and serve the poor and vulnerable groups.

Sammendrag

Tilgang til mangfoldige og veltilpassede frø er avgjørende for å hjelpe bønder med å øke avlingene, landbruksinntektene og matsikkerheten, redusere fattigdom og ulikheter og forbedre menneskers velvære. Realisering av fordeler knyttet til frø som eiendel i agrobaserte samfunn krever at husholdningene er frøsikre. Husholdningenes frøsikkerhet som konsept krever at bønder og jordbrukssamfunn har klar tilgang til tilstrekkelige mengder kvalitetsfrø og plantematerialer av avlingsvarianter, tilpasset deres agroøkologiske forhold og sosioøkonomiske behov, til enhver tid. Tilgangsdimensjonen for frøsikkerhet trenger bedre empirisk og teoretisk forankring for å realisere konseptets potensial for kortsiktig og langsiktig frøsektorutvikling - et politisk mål. Denne avhandlingen bidrar til denne vitenskapelige litteraturen. Hovedmålet er å undersøke muligheter og begrensninger for frøtilgang og utnyttelse i småbruk i sammenheng med økende klimavariabilitet, sosioøkonomisk ulikhet og gjennomgripende transaksjonskostnader som preger frømarkedene, med fokus på tre afrikanske land: Malawi, Tanzania og Etiopia. Fire uavhengige, men beslektede empiriske forskningsartikler tar for seg dette målet.

Den første artikkelen fokuserer på Etiopia og husholdningens atferdsresponser i deres lokale og forbedrede frøbruks- og avlingsdiversifiseringsbeslutninger for nylig eksponering for kovariate klimasjokk og idiosynkratiske husholdningssjokk. Den andre artikkelen fokuserer på frøkjøp, en viktig dimensjon for å forstå frøtilgang, og utforsker påvirkningen av tidligere eksponering for tørkesjokk, kjønn og rikdomsbegavelser på sannsynligheten og omfanget av å kjøpe frø av viktige avlinger i Malawi, Tanzania og Etiopia. Papir tre tar for seg utviklingen av avlingsdiversifisering og respons på nedbørssjokk i Malawi og Tanzania. Til slutt analyserer papir fire den dynamiske karakteren av transaksjonskostnader i frømarkeder som kan lette (eller begrense) frøtilgang gjennom kjøp fra tilgjengelige markeder i Malawi og Etiopia.

Noen få sentrale konklusjoner og implikasjoner kommer fra denne forskningen. Avlingsdiversifisering, sourcing frø off-farm gjennom kjøp, og forbedrede frø varianter er strategier for buffering mot produksjonsrisiko i småbruket landbruk. Sosioøkonomiske ulemper og tilbakevendende nedbørssjokk (tørke) gjør imidlertid husholdningene mer frø usikre og gjør tilpasning vanskeligere. Gjennomgripende transaksjonskostnader preger tilgangen til frø gjennom markedet og reduserer husholdningenes frøsikkerhet. Å samle markedskunnskap, erfaring og nettverk hjelper imidlertid småbønder marginalt med å redusere transaksjonskostnadene i tilgangen til kjøpte frø over tid, og støtter frømangfold og tilpasning til nedbørssjokk. Konklusjonene fra denne forskningen støtter oppfordringen om å fremme integrerte tilnærminger til utvikling av frøsystemer.

I tillegg til å bruke bondelagrede frø og lokale frøsorter, bruker bøndene også forbedrede frø og frø kjøpt gjennom både formelle og uformelle kanaler. Ettersom formelle kanaler representerer den mest sannsynlige primærkilden for nye varianter, bør politikken adressere ulikheten i tilgangen til slike kanaler gjennom tilbudssidetiltak som å øke produksjonen av rimelige kvalitetskontrollerte frø, samt redusere barrierer for tilgang til inngangsmarkedet og etterspørselssidetiltak som sosiale beskyttelsesprogrammer, forbedret utvidelse og informasjonstilgang. Samtidig er uformelle frøsystemer fortsatt ryggraden i frøsystemene bøndene bruker, og frøpolitikk og forskrifter bør muliggjøre sameksistens av formelle og uformelle systemer. En integrert frøsystemtilnærming støttet med retningslinjer som vil redusere ulikheter i tilgang til frø fra kommersielle kilder, vil forbedre tilgjengeligheten av forbedrede og lokale frøvarianter og betjene de fattige og sårbare gruppene.

1. Introduction

1.1. Seeds and livelihoods

Seed is a fundamental asset for most rural households in Sub Sahara Africa, where crop production is a crucial source of livelihood. Seed, as a key input in agricultural production, serves multiple functions, including as a conduit for delivering new technology (i.e., new crop varieties) to enhance the resilience of crop production systems to multiple challenges (e.g., climate and socioeconomic shocks), and increase production and productivity (Walther et al. 2002; Lin 2011; McGuire and Sperling 2013; Hufford et al. 2019). With access to a wide variety of seeds, farmers can choose crop varieties well-suited to local conditions. Farmers may also diversify their portfolios and reduce production risk by growing a larger range of crops of local and or improved crop varieties. Therefore, access to seeds is considered an essential aspect of seed, food, and livelihood security in the wake of both acute and chronic stress situations.

The several functions of seed make it a vital resource in all forms of agriculture production (e.g., organic, industrial, smallholder agroecological farming) (Lammerts van Bueren et al. 2018; Brooker et al. 2021) and its access at various levels (household or national) relates to the attainment of multiple sustainable development goals (SDGs). Access to a diversity of plant genetic resources, including seeds by farmers, directly contributes to poverty reduction (SDG1), improvement of food and nutrition security (SDG2), and helps in coping and adaptation by farmers to environmental stress (SDG13). In addition, access to a fundamental livelihood asset such as seeds addresses a nexus of other sustainable development goals, including reducing inequalities in society (e.g., by raising income amongst the poor) (SDG10) and promoting business development and productive employment (SDG8). A number of untapped business opportunities are present in the seed sector in SSA (Langyintuo et al. 2010; Kassie et al. 2013; Haug et al. 2016), which can create productive employment. Hence, exploring the opportunities and constraints of improving household seed access and security can help achieve important policy goals and motivates this thesis.

1.2. Seed security and seed systems farmers use

The fruition of the benefits attached to seed as an asset in farming communities requires enhancing farmers' seed security. The FAO defines seed security as "ready access by rural households, particularly farmers and farming communities, to adequate quantities of quality seed and planting materials of crop varieties, adapted to their agroecological conditions and socioeconomic needs, at planting time, under normal and abnormal weather conditions." (FAO 1998; 187). Analogous to food security, seed security depends on five dimensions (Table 1): (i) seed availability, (ii) seed access, (iii) seed quality, (iv) seed varietal preferences/suitability, and (v) and resilience (stability of the seed system in the context of shocks)(FAO 1998; FAO and ECHA 2015; FAO 2016).

Table 1: Key dimensions for understanding household seed security

Availability	Is there seed available to the farmer for sowing at the sowing time? Seed availability refers to the farmer's supply of seed from all sources (own saved seed, social networks, local markets, the formal seed sector, and seed aid suppliers).
Access	Does the farmer have physical, social (social network), and economic access to seed? Can the farming household buy, barter, or borrow seed? Seed access refers to the ability of the farming household to acquire seed through exchange, barter, borrowing, or using their power in social networks, socioeconomic status, or influence.
Varietal suitability	Do farmers access seeds of crop varieties with desirable characteristics that suit their multiple needs? Varietal suitability points to seed agronomic and quality traits that meet farmers' needs and preferences. Examples of traits include resistance to pests, diseases, rainfall stress, yield advantage, fodder production, marketability, and culinary and related traits (e.g., taste, aroma, storability, ease of processing).
Quality	Is seed adapted (genetically) to stress?; Is the seed free of contaminants and healthy?; can it germinate? Seed quality is a technical parameter that includes several attributes such as varietal purity, physical purity, germination, and health
Resilience (stability)	Can the household's seed system adapt, resist, and recover from stresses and shocks which threaten the integrity of the household's seed security? A resilient seed system is a seed system where a farmer always has adequate access to enough preferred and adapted seeds to meet their multiple needs in both good and bad cropping seasons.

Notes: Information in the table is adapted from various sources, including (FAO 1998; FAO and ECHA 2015; FAO 2016).

Seed access is one of the most crucial factors behind chronic and acute seed (in)security (Sperling 2002; Nordhagen and Pascual 2013; Sperling 2020). This thesis focuses on understanding seed access in three African countries (Malawi, Ethiopia, and Tanzania) and exploring opportunities and constraints to access and household seed security.

When farmers access seeds, they operate within a seed system. By the term “seed system”, we refer to the organizations, institutions, actors, and activities involved in developing, producing, disseminating, and using seed (Almekinders et al. 1994; Tripp 1997; Louwaars and de Boef 2012). The seed systems farmers use include formal, informal, and emerging intermediate systems. *Formal* seed systems include public and private sector institutions and a series of activities along the seed value chain, including conservation of germplasm in gene banks, development of new varieties, crop variety release, registration, seed production, and dissemination to farmers. *Informal* seed systems are based on farmers saving their own seeds and involve farmers’ seed selection, production, storage, dissemination, and use (Almekinders et al. 1994; Almekinders and Louwaars 2002; Coomes et al. 2015). The *intermediate* seed system comprises market-oriented farmer groups who produce and disseminate (sell) non-certified seeds of both improved and local varieties (Kansiime and Mastenbroek 2016; Mulesa et al. 2021). The distinction among the three seed systems is not clear-cut. It is not only seeds of traditional varieties that circulate in informal and or intermediate seed systems, but also a considerable share of seeds of varieties bred in formal breeding programs from which seeds are saved from their own harvest, exchanged in social networks, or purchased in local markets (Coomes et al. 2015; McGuire and Sperling 2016; Sperling et al. 2020). However, the core characteristics that distinguish formal seed systems from the rest include varietal identity and purity and seed certification for quality (i.e., seeds of optimal physical, physiological, and sanitary quality) (van Gastel et al. 2002; Tripp 2006; McGuire and Sperling 2016). In the informal and intermediate systems, quality declaration and control work is done by farmers themselves. For instance, in the intermediate system, quality declaration of seed is a simplified certification scheme in which seed-producing farmers are responsible for seed quality without formal inspection by regulatory authorities except when training and monitoring farmers’ seed production and processing (FAO 2006).

This thesis draws from seed systems and seed security literature to analyze and comprehend seed access, choice, and use in smallholder farming in Ethiopia, Malawi, and Tanzania in the context of climate variability, shock exposure, and pervasive market imperfections that characterize seed markets. Drawing from seed systems and seed security literature in Paper I, the thesis focuses on the use and extent of use of local and improved seed varieties for selected crops (maize and wheat) and crop diversification in Ethiopia and explores how covariate climate shocks and idiosyncratic household socioeconomic shocks motivate (or constrain) choice of seeds and level of crop enterprise diversification. Paper II focuses on seed purchasing for both local and improved varieties and for selected key crops in Malawi, Ethiopia, and Tanzania and explores how previous drought shock exposure (as a natural experiment), gender, and household wealth endowments explain household seed purchase decisions. In papers III & IV, the thesis focuses more on the dynamics of seed access and use decisions and explores how lagged rainfall shock exposure (drought and flood shocks), transaction costs, knowledge, and experience acquired from past access decisions influence subsequent seed access decisions. Paper III focuses on studying the evolution of crop diversification decisions in Malawi and Tanzania under recurrent rainfall shock exposure and learning how past crop diversification decisions, shocks, and resource limitations (abundance) motivate subsequent crop diversification decisions and adaptation to rainfall shocks. Likewise, in paper IV, the focus is on the evolution of off-farm seed sourcing through purchase in Malawi and Ethiopia and explores how transaction costs, rainfall shocks, community market access, and other factors influence subsequent seed purchasing and household seed security.

The approach taken in this dissertation is to explain access to both local and improved seed varieties (paper I and II) and diversification (paper III and IV) and not only access to improved varieties, as is common in literature. This is important to understand seed security as certified seeds of improved varieties from formal seed systems still only contribute a small share of seeds used by smallholder farmers (Coomes et al. 2015; McGuire and Sperling 2016), although recent trends show an increasing trend (Sheahan and Barrett 2017). Comprehending constraints and opportunities associated with sourcing seed from formal and informal seed systems gives a complete overview of farmers' seed security. Also, smallholder farmers in marginal and heterogeneous environments often maintain large amounts of agrobiodiversity to meet many objectives (Bellon 1996; Wood 1997; Love and Spaner 2007; Jarvis et al. 2011). Hence, by studying the evolution of farmers' crop diversification and other seed sourcing decisions (e.g., purchase), we better understand how farmers' management of traditional varieties, their seed systems, and knowledge and experience in markets (both formal and informal) contributes to seed diversity and security over time.

1.3. Objectives, research questions, and significance of the thesis

This research aims to understand opportunities and constraints to seed access and utilization in smallholder farming in selected African countries in the context of increasing climate variability, socioeconomic inequality, and impaired market access. This research generates evidence from three African countries: Malawi, Ethiopia, and Tanzania. To meet the objectives, the thesis tries to answer a few interrelated questions, which are addressed in the four papers:

(1a) How do exposure to covariate and idiosyncratic shocks influence the types of seeds used by farmers and crop diversification in rural Ethiopia?

(1b) How do household diversity in asset wealth endowments, land size holding, and access to social safety nets mediate the influence of covariate shock exposure on seed use and diversification decisions in rural Ethiopia?

(2a) How do drought shocks, household endowments (household farm size, non-land asset wealth), and gender correlate with household seed purchasing decisions in selected East and Southern African countries?

(2b) Does the impact of past drought shock exposure on seed purchasing decisions vary with the households' wealth and gender?

3a) Do recent (past) exposure to drought or flood shocks, long-term rainfall variability, transaction costs, and knowledge and experience from past diversification increases crop diversification as adaptation in Malawi and Tanzania?

3b) Are households better endowed with assets (land and non-land household assets) more likely to capitalize on experience from past diversification decisions and past shock exposure to intensify crop diversification and help them deal with recurrent rainfall shocks, unlike their poorer counterparts?

4a) Is there persistent state-dependency in household seed purchase decisions leading to selective access to purchased seeds in Malawi and Ethiopia?

4b) Do lagged positive and negative rainfall deviations, average long-term climate (rainfall and temperature), and community market access enhance seed purchasing in Malawi and Ethiopia?

By addressing the above questions, this research contributes to the seed systems and seed security literature on understanding the dynamics in seed access and use amongst the rural poor, particularly how covariate climate shocks, idiosyncratic household socioeconomic shocks, market imperfections, and resource endowment inequalities shape seed access and use options. By using large and representative panel household surveys from three countries, this research generates results that improves our understanding of seed access and utilization in studied countries, which may not be the case when small, district-based cross-sectional surveys are used as common in literature. In addition, by using panel data, the thesis reveals important dynamics instead of static relationships, generated by simple cross-sectional surveys.

Relying on panel data also enhances rigor in empirical approaches applied to study seed access as it allows: careful treatment of endogeneity, unobserved household heterogeneity, and initial household conditions that may influence smallholder household's seed use decisions across space and time (Finkel 1995; Wooldridge 2005; Hsiao 2007; Wooldridge 2010). Overall, relying on large and representative panel household surveys (complemented with longitudinal climate data) in this thesis enhances precision in estimating results (internal validity) and external validity, and should thus serve as an input for evidence-based policy decision making. Results from this thesis are hence relevant for policymakers, development practitioners, and all other stakeholders aiming to enhance seed security for adaptation to shocks and better livelihoods in Sub-Sahara Africa.

The research enhances comprehension of seed access and seed use trends in smallholder farming using large national livelihood data sets, which can inform policies targeting scaling of seed access and security interventions. By revealing heterogeneous responses of farmer seed choice, use, and diversification to covariate and household-specific shocks, results are relevant for coping and adaptation to shocks. Besides, the research unravels the constraints posed by market imperfections in limiting access to seeds which is relevant for policies to reduce barriers to market access in SSA. Furthermore, understanding heterogeneity in seed use options and response to shocks, market imperfections, and other factors is relevant for crafting pro-poor policy interventions to enhance resilience not only of the average farmer but also of the most vulnerable.

A snapshot of the thesis that highlights the key research questions, hypotheses, theoretical frameworks, data, empirical approaches, and key findings is given in Table 2.

Table 2:A snapshot of the thesis

Paper	Research question(s)	Hypothesis	Theory of change	Data and country of focus	Methods	Key findings
I	(i) Does exposure to covariate and idiosyncratic shocks influence the types of seeds used by farmers and the extent of crop diversification in rural Ethiopia? (ii) How does diversity in wealth endowments and access to social safety nets mediate the influence of covariate shock exposure on seed use and diversification?	(i) Exposure to adverse rainfall shocks increases the use of improved seeds and crop diversification through both push and pull factors. (ii) Farmers in disadvantaged positions (e.g., the poor) to a lesser degree than better-off farmers use improved varieties and/or diversify their cropping portfolio post-exposure	State-Contingent Framework (SCF) of Chambers and Quiggin (2000), seed security theory	Living Standard Measurement Survey (LSMS) panel data for rural Ethiopia 2012-2016 combined with historical monthly weather data from WorldClim and NASA's Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2).	Correlated Random Effect (CRE) logit, Tobit, and Poisson models	(i) Drought and temperature shocks increase improved seed use (ii) Long-term rainfall, flood shocks, and temperature shocks encourage crop diversification (iii) Recurrent drought exposure enhances local seed use while deterring improved seed use (iv) Drought shocks among farmers with small farm sizes and low asset endowments reduce improved seed use and diversification but increase local seed use
II	(i) How do drought shocks, household asset endowments, and gender influence household seed purchasing decisions in East Africa? (ii) Does the impact of past drought shock exposure on seed purchasing decisions vary with the households' wealth and gender?	(i) Exposure to drought shocks in prior seasons increases the likelihood and extent of purchasing seeds in the following season (ii) Households better endowed with assets and male-led households (in input decision making) are likely to purchase seeds post-drought shock exposure	State-Contingent Framework (SCF), seed security theory	LSMS panel data sets for Malawi (2010-2019), Tanzania (2010-2015), and Ethiopia (2016-2018), and historical monthly weather data from WorldClim.	Cragg's double-hurdle models	(i) Seed purchasing is a common and essential component of seed access in studied countries (ii) Past exposure to drought shocks encourage seed purchasing in the following season. (iii) Relatively more affluent households and male-led households are more likely to purchase improved seeds post-drought shock exposure
III	(i) Do rainfall shocks (drought, flood shocks, long-term rainfall variability), knowledge, transaction costs, and experience from past diversification increase crop diversification as adaptation in Malawi and Tanzania? (ii) Are households better endowed with assets more likely to capitalize on experience from past diversification decisions and past shock exposure to intensify crop diversification and help them deal with recurrent rainfall shocks, unlike their poorer counterparts?	(i) Lagged rainfall shocks increase subsequent crop diversification. (ii) Crop diversification decisions are state-dependent. (iii) Households better endowed with assets capitalize on the experience gained from past diversification decisions and use it to intensify crop diversification and achieve multiple goals, including dealing with climate risk, unlike their poorer counterparts.	Dynamic transaction costs theory, seed security theory, and State-Contingent Framework (SCF)	LSMS household panel data for Malawi (2010-2019) and Tanzania (2008-13) combined with historical weather data from WorldClim.	Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Tobit and Poisson models	(i) Smallholder farmers in Malawi and Tanzania intensify on-farm crop diversification in response to increasing rainfall variability and drought shocks. (ii) Crop diversification decisions are state-dependent, implying past crop diversification enhances later diversification. (iii) Household asset endowments help farmers exploit opportunities from experience and intensify crop production to help them adapt to rainfall shocks.
IV	(i) Is there persistent state-dependency in household seed purchase decisions leading to selective access to purchased seeds in Malawi and Ethiopia? (ii) Do lagged rainfall shocks, average long-term climate, and community market access enhance seed purchasing in Malawi and Ethiopia?	(i) There is persistent state dependency in the smallholder farmer seed purchasing decisions leading to selective access to purchased seeds. (ii) Lagged rainfall shocks and long-term climate variability enhance seed purchase decisions. (iii) Seed purchasing decisions increase with market access within the community.	Dynamic transaction costs theory. State-Contingent Framework (SCF), seed security theory	LSMS household panel data for Malawi (2010-2019), and Ethiopia (2012-16) combined with historical weather data from WorldClim.	Dynamic Random Effects (DRE) Probit and Tobit models	(i) Pervasive transaction costs characterize seed markets in Malawi and Ethiopia, leading to selective access to purchased seeds. (ii) Seed purchasing also responds to long-term and short-term measures of climate variability, household resource endowments, and community market access.

2. Conceptual framework

The predominant conceptual framework used in this dissertation to explain farmer seed use decisions under risk and uncertainty and other factors is the State-Contingent Framework (SCF) of Chambers and Quiggin (2000).

2.1. Insights from the State-Contingent Framework

The SCF originated from Arrow (1953) and Debreu (1952) in the context of general equilibrium theory and the analysis of decisions under risk and uncertainty (Chambers and Quiggin 2000; Quiggin and Chambers 2006). The SCF has been used previously to study agriculture technology adoption decisions under risk and uncertainty, including the adoption of improved seeds (Holden and Quiggin 2017; Katengeza et al. 2019a; Gebru et al. 2021), cash crops (Gebru et al. 2021), integrated soil fertility management technologies (Katengeza et al. 2019b), and land rental market participation (Gebru et al. 2019; Tione and Holden 2021a), to mention a few examples.

In the SCF, farming households endowed with assets, labor, and land who aim to maximize crop production utility based on beliefs about the likelihood and production outcomes under different states of nature make state-contingent decisions (Chambers and Quiggin 2000; Quiggin and Chambers 2006; Holden and Quiggin 2017). Farmers make input decisions before weather conditions for that season are revealed, based on their beliefs, preferences, and expectations, and these decisions influence the production outcomes, which subsequently form the basis for input decisions in the following year (Chambers and Quiggin 2000; Quiggin and Chambers 2006; Holden and Quiggin 2017). Over time, exposure to different states of nature helps the farmer build more realistic expectations about the performance of alternative farming practices and technologies (including seeds) that may influence adoption and adaptation processes. In other words, households gain experience over time that helps them shape their subjective production risk assessment, farming input choices, and consumption decisions, *ex-ante* and *ex-post* the production period (Quiggin and Chambers 2006; Dercon and Christiaensen 2011).

Thus, the SCF allows us to study how producers may allocate input resources to manage risks, given their constraints, preferences, and other factors, e.g., the price of inputs (Chambers and Quiggin 2000; Quiggin and Chambers 2006; Holden and Quiggin 2017). The characteristics of input resources and state-contingent outputs are best described in terms of complementarity and substitutability relationships. An increase in the probability of a less favorable state of nature (e.g., drought or flood) will lead to an increase in the share of risk substituting inputs in the input mix for a given expected output or vice versa. In the context of this thesis, farmers' urge to adapt to increased climate risks associated with climate change will, *ceteris paribus*, increase the likelihood of adopting risk-substituting crop varieties or other seed use options, e.g., diversity.

The SCF allows for mixtures of technology or seed, such as those arising from partial adoption of new technology (e.g., new seed). On the one hand, partial adoption of seed types may yield benefits from diversification (e.g., using both improved and local varieties). On the other hand, partial adoption may reflect constraints to adoption, including high costs of new technologies, access constraints leading to selective access to technology in input markets, and or heterogeneous environments that make technology performance and choice more complex (Holden and Quiggin 2017).

In analyzing farmers' response to risks and uncertainty in their seed choice, use, and diversification decisions, this thesis assumes that the vulnerability of households is closely associated with their resource poverty. Their most important resources are land, labor, and durable household assets. Land-, labor-, asset-poor households are therefore assumed to be more vulnerable to shocks. Also, this research assumes that it is more difficult to use social networks to protect oneself against covariate shocks than against idiosyncratic household-specific shocks, making interventions such as the safety net programs more important as protection against covariate shocks such as droughts or floods. Furthermore, drought and flood shocks have direct effects on the performance of the crops and varieties grown and on, market prices and the availability of essential commodities. Improved varieties may or may not perform better than the local varieties in different environments and under different states of nature. The effects of interactions between multiple risk sources are complex and ambiguous.

Shocks may, hence, alter the household's farming activities in heterogeneous ways. For instance, the literature distinguishes between *ex-post* risk coping mechanisms (what farmers do after exposure to shocks) and what they do before exposure (*ex-ante* risk management) (Angelsen and Dokken 2018). Characteristics of the rural settings such as over-reliance on agriculture, lack of functional insurance markets, and the dire consequences of a bad season (Rose 2001; Dercon 2005) complicate both *ex-post* and *ex-ante* responses to shocks. Households may switch from selling food in years with good rainfall and becoming net buyers in years with poor rainfall. Covariate risk implies that such rainfall shocks occur simultaneously to households in large geographical areas with the consequence that most of them are net sellers in good years, and most are net buyers of food in bad years. Therefore, poor market integration leads to low food prices when they are net sellers and high food prices when they are net buyers.

Following the SCF, this dissertation utilizes climate variability and weather shocks as a natural experiment and tests the influence of climate variability and recent rainfall shocks (e.g., drought shocks) on seed choice (improved vs. local varieties), seed sourcing through purchase (for both improved and local varieties), and diversification of crop production. Climate variability and shock variables make key variables used to explain seed use, sourcing, and diversification decisions throughout the thesis (papers I to IV).

2.2. Insights from other theoretical frameworks

Farmers' seed use, sourcing and diversification decisions may not be influenced by production factors alone but also by consumption and related factors. Household production decisions are separable (independent) from consumption decisions if markets are perfect. However, in most parts of SSA, markets are imperfect, which means farmers' production and consumption decisions are inseparable. For that reason, we draw insights from other theoretical frameworks such as the Sustainable Livelihoods Framework (SLF) and Agriculture Household Models (AHM) (Singh et al. 1986; De Janvry et al. 1991; Behrman 2000) (briefly described below) and explore how other factors, such as asset endowments, household characteristics, may drive or condition seed use and sourcing decisions as a response to shocks.

The sustainable livelihoods framework (SLF) of Carney (1998), Scoones (1998), and Ellis (2000) has been used to study the economics of agricultural decisions and outcomes including climate change vulnerability and adaptation (e.g., Pandey et al. (2017)), and livelihoods and poverty reduction in rural settings (Ellis and Mdoe 2003; Ade Freeman et al. 2004; Ellis and Freeman 2004). The SLF talks about dynamics in interactions of various components, including household resources (assets), livelihood activities, institutional context, vulnerability contexts (i.e., exposure and vulnerability to external shocks and internal changes), and outcomes (Carney 1998; Scoones 1998; Ellis 2000). Assets (physical, natural, human, social and financial) are used by households in their on-farm and off-farm activities to make a living while contextual factors (e.g., social relations and institutions) and exposure to covariate (e.g., climate shocks), and idiosyncratic shocks (e.g., changes in household composition or asset base) directly influence how these assets influence livelihood strategies and outcomes. Consequently, the combinations, and recombination of household assets, institutions that govern access and control of assets, and shocks determine production relations in smallholder farming (Binswanger and Rosenzweig 1986). The net effects of these interactions are reflected in in a range of livelihood outcomes, including income, food security, and poverty reduction, amongst other measures of quality of life (Carney 1998; Scoones 1998; Ellis 2000). The relationships in the SLF are dynamic as outcomes in one season impact assets and livelihood strategies in subsequent seasons.

Likewise, agricultural household models (Singh et al. 1986; De Janvry et al. 1991; Behrman 2000) can be used to study agricultural decisions and outcomes where households are both producers and consumers in imperfect markets. Semi-commercial farms that produce multiple crops and or livestock make up a large part of the agricultural sector in developing economies, and these farms often combine two fundamental units of economic analysis (the household and the firm/farm). These agricultural household models can be independent (recursive) or dependent (simultaneous) (Singh et al. 1986). Recursive agriculture household models treat production and consumption decisions as independent, implying that farm households can be modeled as pure profit maximizers. However, household production and consumption decisions are inseparable (especially with imperfect markets), indicating that a production or profit maximization model would not adequately describe the decision-making process (Singh et al. 1986; De Janvry et al. 1991; Caviglia-Harris 2004). Non-separability of production and consumption decisions, in other words, implies that asset distribution and consumption needs may significantly impact production decisions and the management of land and labor (Caviglia-Harris 2004).

Overall, production and consumption-related factors are important when analyzing smallholder farmer technology adoption decisions. For instance, the smallholder farmer's decision to seed choice or diversify crop production could be driven by the need to respond to market imperfections common in SSA (Ellis 2000; Dercon and Christiaensen 2011; Alobo Loison 2015). Due to market imperfections in SSA, market access is not uniform because households may face different transaction costs (Renkow et al. 2004; Barrett 2008). These prevalent transactions costs emanate from policies, institutions, and social factors that influence the degree of information asymmetry and access to productive resources (Fafchamps 2004; Holden et al. 2010; Ricker-Gilbert and Chamberlin 2018; Gebru et al. 2019; Tione and Holden 2021b). To assess the possible influence of high transaction costs emanating from imperfect factor markets (seed markets) in papers III and IV, we draw from intertemporal production relations (Binswanger and Rosenzweig 1986) and dynamic transaction costs models (Holden et al. 2007; Holden et al. 2010) and study evolution of diversification and seed purchase decisions and explore how previous participation (proxying experience and knowledge in markets that may reduce transaction costs over time) influence subsequent decisions.

2.3. Sketching a conceptual framework

The central aim of this thesis hinges on understanding seed access, opportunities, and constraints to seed access and resilience in selected east and southern African countries (Malawi, Tanzania, and Ethiopia) in the context of increasing weather variability, shock exposure, socioeconomic inequality, and impaired market access. The four papers that make the dissertation explore in-depth links between these key factors generating evidence from the studied countries.

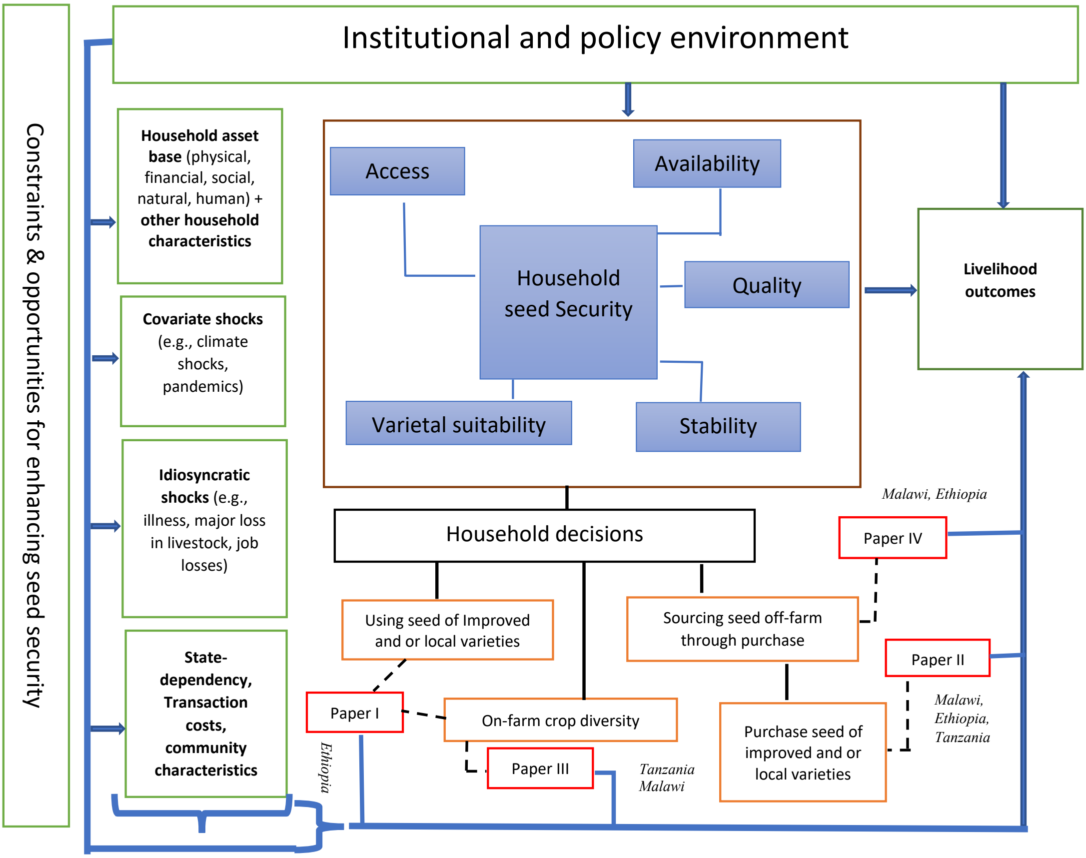

Figure 1: Constraints and opportunities to enhancing household seed security

Figure 1 presents a simplified framework for understanding household seed security, including how seed security elements (access, quality, resilience, availability, and varietal suitability) interact with various factors from the policy & institutional framework to shocks (covariate and idiosyncratic), household asset characteristics, transaction costs, community characteristics, other farm characteristics, and seed security links to livelihood outcomes (e.g., productivity and food security). The linkages presented in Figure 1 are neither unquestionable nor exhaustive. Still, they highlight the key sources of constraints and opportunities to household seed security studied in the four papers that make up the dissertation. The relationships are dynamic, as seed security outcomes (e.g., access) in one season may impact livelihoods, household characteristics (e.g., asset build-up), and access to seeds in subsequent seasons.

Dependent variables used in the analysis are the decisions made by households on seed use by type (local and improved), seed purchasing by type (local and improved varieties), and crop diversification proxy household seed security (mainly access). By analyzing the use of local and improved seeds used by farmers (as in paper 1), the purchase of seeds of local and improved varieties (as in papers II and IV), and the diversification of crop species at the farm-level (as in paper I and III) we comprehend farmers access to seeds from available formal and informal seed sources. Also, by analyzing responses in household seed access variables (use, purchase, and diversification) to previous shock exposure, transaction costs, and other factors, we can learn about the use of different varieties and cropping strategies by farming households in coping and adapting to shocks. Also, by analyzing the responses to shocks in seed access variables by farmers in different socioeconomic strata (by asset endowments), we can derive implications of results for the resilience of farmer seed systems to shocks and other forces.

3. Data and Methods

3.1. Data sources

3.1.1. Household survey data

Longitudinal household survey data and historical climate data are used in the thesis. Household socioeconomic data for Malawi, Tanzania, and Ethiopia comes from multiple rounds of the rich and representative Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA), available through the World Bank[1]. Collection of these data is done through the LSMS-ISA team working with national governments. For our three study

[1] These data are available publicly at the following site: https://microdata.worldbank.org/index.php/catalog/lsms/?page=1&ps=15&repo=lsms

countries, these data sets are usually termed Ethiopia Socioeconomic Survey (ESS), Malawi Integrated Household Survey (IHS), and Tanzania National Panel Survey (TNPS). The LSMS-ISA data sets contain comprehensive information on agricultural activities and household socioeconomic conditions. These data cover both rural and urban households, but this thesis uses only data from rural households with information on engagement in agriculture and seed access and use.

The exact information collected have changed over survey rounds, and some variables important for the thesis are available in some rounds and not in others. Hence, the selection of data (and survey rounds) for the papers that make up this dissertation is not entirely uniform across countries and in all manuscripts. In paper I, which focuses on seed use and diversification in Ethiopia, the study derives a balanced household panel from the first three rounds of ESS (1, 2, and 3), which is then used to test hypotheses in that paper. However, in Paper II, the aim was to comprehend seed purchasing trends for both improved and local seed varieties in all studied countries. Survey rounds that captured seed purchasing practices by seed type were selected. For Ethiopia, data from the two latest cross-sectional rounds of the ESS data (ESS3-2016 and ESS4-2018) was used. In addition, survey data from three rounds of the Malawi IHS (3rd 2010/11, 4th 2016/17, and 5th 2019/2020) are used, while for Tanzania, three rounds of the TNPS data from TNPS 2(2010/11, 3(2012/13), and 4(2014/15), respectively) are used. Manuscript III studies the evolution(dynamics) of crop diversification in Malawi and Tanzania and response to rainfall shocks and uses balanced panel data for both countries. For Malawi, the balanced panel data is derived from the Malawi short panel that spans eight years (from 2010/11-2018/19) and comprises four rounds of data conducted in 2010, 2013, 2016, and 2019. For Tanzania, a balanced panel data set was made from the first three rounds of the TNPS (1, 2, and 3) and used to comprehend dynamics in crop diversity and response to shocks. Lastly, paper IV studied the dynamics and evolution of seed purchasing decisions in Ethiopia and Malawi using the same household balanced panel data as in paper I (for Ethiopia) and paper III (for Malawi).

3.1.2. Weather data

In addition to the LSMS-ISA household survey data, historical monthly weather data from WorldClim (Masarie and Tans 1995; Harris et al. 2014; Fick and Hijmans 2017) is used to define historical climate variables (precipitation and maximum temperature), climate variability (temperature and rainfall anomalies), and or lagged climate and shock variables (e.g., drought or flood). The LSMS-ISA household data provide the approximate locations (longitude and latitude) of clusters (villages) from which interviewed households reside. This location (georeferenced) data is used to extract historical monthly climate data for over 30 years (1980-2018) combined with household data for analysis. The WorldClim data is available at a fine(high) spatial resolution of 2.5 minutes (approximately ~21 km^2) and is suitable for the study. The WorldClim data are rationalized by the Climatic Research Unit, University of East Anglia, using WorldClim 2.1 for bias correction (Masarie and Tans 1995; Harris et al. 2014; Fick and Hijmans 2017). Climate variables used throughout the thesis are derived using these data.

In addition, an alternative climate data source is used to assess the robustness of results to the choice of weather data. Properties of weather that characterize data from different sources, such as the selection of weather stations, bias correction methods used, spatial resolutions of data, and imputation of missing data, among other factors, may drive results (Auffhammer et al. 2013; Letta et al. 2018). This dissertation therefore also uses data from NASA's Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al. 2017), to define weather variables and shocks in robustness analyses. After obtaining similar patterns in climate data and analysis results (from using WorldClim and NASA data) in paper I, the thesis proceeded with using WorldClim data as the main data source for climate variables throughout the dissertation (paper II-IV).

3.2. Empirical estimation strategies

Empirical strategies used in this thesis are motivated by several factors: (i) working with limited dependent (censored or count) variables, (ii) the need to deal with unobserved household heterogeneity, (iii) addressing possible attrition bias, (iv) uncovering dynamic as opposed to static relationships, (v) utilizing natural experiments (natural events) to estimate causal effects, and (vi) the need to perform heterogeneity analysis, to mention a few key issues. A brief discussion on some of the issues in terms of their effects and how they are addressed in this thesis is given below; detailed information is given in papers (I-IV) that make up the thesis.

3.2.1. Limited dependent variables

Censored outcome variables

Seed access and use variables used as dependent variables throughout the thesis are censored (in some way) as not all farmers within the sample adopt a particular technology. For local seeds, improved seed, diversification, and seed purchasing variables, we have adopters (farmers using the technology) and non-adopters (those not using the technology). Accordingly, the intensity of adoption is only observed for adopters, and non-adopters will have zeros. As discussed in Wooldridge (2010), having censored dependent variables requires that the model applied is appropriately expressed as a corner solution. To deal with such censoring, the Tobit estimator (Tobin 1958), Cragg Double Hurdle (DH) models (Cragg 1971), and Heckman selection models (Heckman 1979) are alternatives. The Tobit estimator may be used with censored dependent variables since zeros (non-adoption) represent household choice and not missing data due to incidental truncation. However, the Tobit estimator is restrictive in that it assumes that for a given technology, the decision to adopt and the degree of adoption are determined by the same process. Alternatively, Cragg DH and Heckman models give more flexibility and impose weaker assumptions by allowing adoption and intensity of adoption to be a two-stage decision. Within the Cragg DH models, the decision to adopt (the first hurdle) is estimated using a probit estimator, while the degree or intensity of adoption (the second hurdle) is estimated using a truncated regression model. The Heckman selection estimator also relaxes the restrictive assumptions of the Tobit estimator by having a first stage selection equation for the adoption decision and a second stage outcome equation for the degree of adoption. However, the Heckman selection estimator is more demanding in that for it to be identified, the exclusion restriction must be satisfied. At least one exogenous variable in the selection equation that significantly explains the adoption decision must be dropped in the second stage (Heckman 1979; Cameron and Trivedi 2005) to satisfy the exclusion restriction. Throughout the thesis chapters (paper I-IV), the Tobit, Cragg DH, and Heckman selection estimators are used as alternatives.

Count outcome variables

Count outcomes are also nonlinear outcome variables that require appropriate treatment in regression modeling. Regression models for count data are nonlinear, with many properties and special features, intimately connected to discreteness and non-linearity as with other limited dependent variable models such as probit or logit (Cameron and Trivedi 2005). Poisson regression and negative binomial regression are alternatives to dealing with count outcomes. In a Poisson regression model, the probability of a count is determined by a Poisson distribution, and the mean of the distribution is a function of the independent variables. The conditional mean of the outcome should be equal to the conditional variance. However, in practice, the conditional variance may exceed the conditional mean, a feature usually called overdispersion. The Negative Binomial Regression model is robust to overdispersion as it allows the variance to exceed the mean (Gardner et al. 1995; Cameron and Trivedi 2005). The choice between the Poisson and Negative binomial models requires testing for dispersion using residual plots or performing likelihood ratio tests to determine whether there is a statistically significant difference in the fit of the two regression models. In papers I and III, various indices to measure crop diversification, including the crop count index, are used where strategies described here are applied to choose appropriate count data regression techniques to model crop count. More details in that respect are given in the respective papers.

3.2.2. Dealing with Unobserved Heterogeneity

The presence of household heterogeneity that influences adoption (our outcome variables) but is otherwise unobserved is another challenging aspect that this thesis had to deal with. The main challenge that unobserved heterogeneity (if not addressed) can bring is that of introducing selection bias as some households (because of underlying characteristics, e.g., ability and motivation) are more likely to adopt a particular technology than other households (Wooldridge 2010). The standard panel data method to deal with unobserved household heterogeneity will be to apply household Fixed Effects (FE), which allow arbitrary correlation between unobserved household time-invariant factors and observed household variables. However, the household FE applies when the dependent variables are linear, which is not the case with seed use variables used in this thesis (which are nonlinear). To avoid incidental variables problems that FE introduces when applied to nonlinear models, the thesis adopts the Correlated Random Effects (CRE) framework, first suggested by Mundlak (1978) and Chamberlain (1982), which is equivalent to using household FE with continuous dependent variables. The CRE framework applies to nonlinear models, including Probit, Logit, Tobit, Double Hurdle, and count data models (Wooldridge 2010). In line with the CRE approach, the assumption that the unobserved household heterogeneity can be replaced with its linear projection onto the time averages of all household level regressors (Mundlak 1978; Chamberlain 1982) is made. The CRE is advantageous in that it also relaxes the strong assumption of no correlation between unobserved household heterogeneity and observed variables required in standard Random Effects (RE) models (Wooldridge 2010). More information on how the CRE framework is applied to address unobserved household heterogeneity in the analysis is detailed in the respective papers that make up the thesis (e.g., in papers I and IV).

3.2.3. Attrition bias

When working with longitudinal data, attrition is a common problem, which is the loss of sample members between the first and subsequent waves of data collection (Fitzgerald et al. 1998; Wooldridge 2010). To handle the possible attribution bias effect in estimation, the thesis follows the following steps in testing and controlling for attrition bias. First, estimation of probit attrition models is done to assess whether attrition was random and hence ignorable with a dummy variable (1=yes) for households not observed in the follow-up surveys, and zero otherwise, using household characteristics at baseline as explanatory variables. From the attrition probit results, and by assessing the significance of explanatory variables, it was evident that some household characteristics were significant in explaining the probability of dropping out, indicating that attrition was not random, which could lead to bias. In that case, where attrition is not random, regression estimates will need adjustment to correct for attrition bias.

To correct attrition bias, the second step taken was constructing an Inverse Mills Ratio (IMR) from the attrition probit models. The IMR constructed becomes a time-invariant variable in the balanced panel data used as households retain the same value of IMR across panel rounds. Third, the constructed IMR is used to test and control for the potential attrition bias effect by including it as an additional explanatory variable in panel regression models. The inclusion of the IMR in panel regression models tests and controls for possible effects of attrition bias. Finer details on testing and controlling for attrition bias in estimation are detailed in manuscripts that make up the thesis (e.g., papers I, III, and IV).

3.2.4. Uncovering dynamic relationships

Using household panel data comes with numerous benefits as compared to cross-sectional data. Some notable advantages of using panel data as opposed to cross-sectional data are that panel data are more informative (have more variability, more degrees of freedom, and less collinearity), they allow the study of dynamic as opposed to static effects, they give information on the time ordering of events and also permit careful control of individual unobserved heterogeneity which is a key problem (leading to confounding) in non-experimental research (survey data) (Finkel 1995; Hsiao 2007). In addition, the availability of panel data also allows the estimation of dynamic nonlinear panel models (e.g., dynamic Tobit and Probit models) (Finkel 1995; Wooldridge 2005; Hsiao 2007), which possibly captures the influence of past decisions on subsequent decisions (state-dependency effects). Besides, dynamic panel models also control for initial unobserved conditions that may influence seed use and diversification decisions across time and space (Finkel 1995; Wooldridge 2005; Hsiao 2007). In papers III and

IV, dynamic nonlinear panel data models are used to test state dependency and hence the possible influence of knowledge and experience and transaction costs in driving subsequent seed use decisions (e.g., purchasing seed) and crop diversification.

3.2.5. Using natural experiments (exogenous natural events) to establish causal effects

Natural experiments allow the study of the effects of uncontrolled natural events, such as exposure to drought or flood shocks, pandemics, or uncontrollable famines where randomization is not feasible. The key element of natural experiments is that the change in exposure is caused by factors or shocks outside the researcher's control and that manipulation of exposure by the researcher is not possible which allows the identification of intervention and control groups (Craig et al. 2012). The 2021 Nobel Laureates David Card, Joshua Angrist, and Guido Imbens, through their work, demonstrate how natural experiments can be used to answer central questions for society, such as the effect of education on earnings, how the minimum wage affects employment, how the labor market is affected by immigration, and clarifies what conclusions can be drawn from using the "natural experiment" research approach (Angrist and Keueger 1991; Card and Krueger 1994; Imbens and Angrist 1994). Their work has revolutionized empirical research in the economic sciences and motivates this thesis to explore the possible influence of uncontrollable events such as climate variability and shocks on farming decisions, including seed use, purchasing, and diversification practices. Accordingly, this dissertation explores the causal influence of exogenous climate variables measuring variability and shock exposure to explain different dimensions of seed security (e.g., access, stability) in the studied countries.

3.2.6. Heterogeneity analysis

As explained in the conceptual framework, this thesis assumes that households' vulnerability to shocks is closely associated with resource poverty. An immediate implication of this assumption is that poorer households (in terms of land and non-land assets) are more vulnerable to shocks. This thesis hence explores heterogeneity analysis in studied relationships to understand whether households' responses to climate shocks (in terms of seed use decisions) vary considerably with household asset (land and non-land) endowments. Heterogeneity analysis is achieved through sample splitting[2] and interaction effects. Further details on heterogeneity analysis are given in the papers I, II, and III.

[2] Splitting samples by indicator variables of resource endowments (high vs low) and running separate regressions in sub-groups of farmers.

4. Main findings

This section summarizes the main findings from the research. It highlights the broader research questions addressed, methods used, and the key results and implications. Elaborate findings and discussions are detailed in the papers (that make up this thesis.

4.1. Do climate shocks influence seed choice, use, and on-farm crop diversification in rural Ethiopia? (Paper I)

Covariate and idiosyncratic risks are a central part of livelihoods in Sub-Saharan Africa. Agriculture is a key pillar in rural livelihoods, and exposure to both types of shocks is common and affects access to agricultural inputs and therefore coping and adaptation strategies. Therefore, understanding how farmers cope and adapt to shocks is important for policy responses to improve their seed, food, and livelihood security. Paper I evaluates the influence of lagged covariate and idiosyncratic shock exposure on local and improved crop variety use and crop diversification practices in rural Ethiopia. A balanced household panel data set compiled from three rounds (2012, 2014, and 2016) of the Living Standards Measurement Study-Integrated surveys on Agriculture (LSMS-ISA) for Ethiopia, combined with historical monthly weather data, is used to answer policy-relevant questions. More specifically, paper I answer two main research questions: (i) How do exposure to covariate and idiosyncratic shocks influence the types of seeds used by farmers and the extent of crop diversification? (ii) How does household diversity in asset wealth endowments, land size holding, and access to social safety nets mediate the influence of covariate shock exposure on seed use and diversification decisions in rural Ethiopia? The study applies correlated random effects models, which control for time-invariant household unobservables in a similar way as household fixed effects do when continuous dependent variables are used.

The paper yields several interesting findings. First, lagged drought and temperature shocks and historical mean rainfall enhance improved seed use. Second, lagged flood and temperature shocks and historical mean rainfall enhance crop diversification. Third, recurrent drought exposure significantly reduces overall agricultural activity. Idiosyncratic shocks (e.g., losing formal employment and livestock) only minimally explain seed use and crop diversification decisions when compared to covariate rainfall shocks.

Heterogeneity analysis reveals that drought shock exposure among farmers with small farm sizes and low asset endowments reduces improved seed use and diversification but increases local seed use. Overall, the results imply that farmers develop weather expectations from previous weather conditions and influence production decisions in subsequent seasons and that poorly endowed households in terms of assets are more vulnerable to negative shocks such as droughts. Findings from the study are robust to various sensitivity checks and are relevant for policy responses aiming to strengthen smallholders' ability to cope with and adapt to shocks. Policies addressing social inequality and supporting farmer seed systems are necessary to enhance seed security under shocks in rural Ethiopia.

4.2. How do drought shocks and socioeconomic heterogeneity prompt seed purchase in East Africa? (paper II)

Paper II focuses on seed purchase, an important dimension for understanding seed access, long considered one of the most crucial factors behind chronic and acute seed (in) security. Seed purchasing enables the farmer to respond to factors that result in chronic and temporary seed insecurity (e.g., varietal deterioration with time (quality) and depleted farmer saved seed stocks through droughts (availability)) and enables them to exploit opportunities associated with accessing new seed (e.g., growing new crops and accessing drought-tolerant crop varieties (access)). This paper, therefore, focuses on seed purchasing (for both local and improved varieties), which is vital for comprehending seed security. Using data from three African countries (Malawi, Ethiopia, and Tanzania), it tests the hypothesis that exposure to drought shocks in prior seasons increases the likelihood and extent of purchasing seeds in the following season. In addition, and to understand the interactions between climatic factors and farm size, and other wealth factors influencing seed purchasing, the paper addresses the questions: (i) How do household farm size, non-land asset wealth, and gender correlate with seed purchasing decisions? (ii) Does the impact of drought shock exposure on seed purchasing decisions vary with the households' wealth and gender of prime decision maker on input acquisition? The paper uses multiple rounds of large and representative Living Standard Measurement Study (LSMS) data sets, available from the World Bank, combined with historical monthly weather data (rainfall and temperature) and rigorous econometric techniques to answer research questions.

Highlights from the results portray that, on average, lagged drought shock exposure increases seed purchasing for both improved and local seeds in Malawi and Tanzania while encouraging (discouraging) local (improved) seed purchases in Ethiopia. In all three countries, farmers better endowed with household assets increase seed purchasing, particularly for improved seeds, after a drought shock exposure. In addition, smaller farm sizes and low asset wealth endowments in all study countries are significant deterrents for buying seeds in the market, particularly improved seeds. The overall implication of the findings is that farmers respond to previous drought shocks by purchasing new seeds to exploit opportunities associated with accessing new seeds (e.g., accessing seeds with drought tolerance/resistance traits) and responding to depleted farmer-saved seed stocks (farmer seed saving is limited in drought-hit years). Policies need to support both formal and informal seed systems and address inequalities in access to seed from formal seed channels to achieve seed and food security under elevated climate risk.

4.3. How does crop diversification evolve and respond to rainfall shocks over time? Evidence from Malawi and Tanzania (paper III)

Heavy reliance on one or a few crops makes seed systems more vulnerable to climate change. Moreover, low crop diversity in food systems may hurt the environment, economies, and human health. This paper focuses on the evolution of crop diversification in smallholder farming in Malawi and Tanzania and answers three interrelated questions: (i) Does recent exposure to short-term drought and flood shocks and exposure to long-term rainfall variability increase crop diversification? (ii) Are crop diversification decisions state-dependent, i.e., can past crop diversification explain later crop diversification decisions? (iii) Are households better endowed with assets (land and household assets) more likely to: capitalize on the experience gained from past diversification decisions and use it to diversify production over time to achieve multiple objectives, including dealing with climate risk, unlike their poorer counterparts?

In order to answer these questions, the paper uses balanced household panel data for rural farmers from the LSMS-ISA in Tanzania and Malawi combined with historical weather data. Using multiple indices of farm-level crop diversity (count and Simpson index), the study applies household correlated random effects and dynamic random effects estimation methods that control for unobserved heterogeneity in household crop diversification decisions plus unobservable initial conditions that may influence crop diversification across space and time.

The results reveal that smallholder farmers in Malawi and Tanzania respond to short-term drought shocks and long-term rainfall variability by intensifying on-farm crop diversification and that crop diversification decisions are state-dependent, implying past crop diversification enhances later diversification. Knowledge and experience from past diversification (among other factors) gradually reduce transaction costs in implementing subsequent crop diversification. On the one hand, knowledge, linkages, and experience in formal markets marginally reduce transaction costs and enhance subsequent access to seeds from available markets. On the other hand, crop diversification supports *in-situ* agrobiodiversity conservation, which is also an essential source of seed and planting materials for farmers. Hence, when farmers have access to off-farm seeds and farmer saved seeds, it supports crop diversification and adaptation to rainfall shocks over time.

Heterogeneity analysis reveals that land and non-land asset (household wealth) endowments help farmers implement diversified cropping portfolios over time to help them deal with rainfall shocks. Relatively better-off farmers with abundant land and non-land assets are more likely to achieve crop diversification after a drought exposure and as an adaptation to future expected shocks. The study findings support policies that ensure and promote access to a diversity of affordable, well-adapted crop seeds that meet farmers' needs and preferences in Malawi and Tanzania and thereby support crop diversification and improve resilience under rainfall uncertainty and shocks.

4.4. How do pervasive transaction costs in seed markets and increased rainfall uncertainty constrain seed access through purchase in Malawi and Ethiopia? (paper IV)

As addressed in paper II, seed purchase is an avenue that helps smallholder farmers source seed off-fam and complement on-farm seed sources, consequently reinforcing household seed security and adaptation to rainfall shocks. However, seed purchase through available markets is complicated by pervasive market imperfections that characterize many markets in the developing world, including SSA. Smallholder farmers may face dynamically variable transaction costs with imperfect factor markets that are poorly integrated (coordinated), less developed, and spatially dispersed. Such costs can significantly influence decisions by households on whether to participate or not to participate in the market; transaction costs raise the price effectively paid by buyers and lower the price effectively received by sellers, creating a price range within which some households may find it unprofitable either to sell or to buy. In seed markets and on the demand side, such transaction costs may include the costs of searching and obtaining information on production and consumption traits of the seed of different crops and or varieties, costs of searching and locating them, and negotiation costs.

Paper IV uses balanced household panel data from Malawi and Ethiopia and investigates how access to purchased seeds is constrained (or facilitated) by state dependency, rainfall shocks, and other factors. State dependency in seed markets implies that market participants capitalize on their experience and established networks gained through repeated engagements in the market to identify trading partners, which is not the case with new entrants without such experience and networks. The paper compares Malawi and Ethiopia, two countries with contrasting features in terms of the policy framework governing the development of seed systems that farmers use. Using dynamic Probit and Tobit random-effects models, the papers assess how nonlinear transaction costs in seed markets, rainfall shocks, and other factors influence access to off-farm seed while controlling for unobserved household heterogeneity that might influence participation in the seed market by smallholder farmers.

Findings from the study reveal nonlinear effects of lagged seed purchase decisions on subsequent decisions with strong initial effects (weakening over time). For instance, initial maize seed purchase decisions are associated with between 11-13% (1 kg) and 21-27% (2 kg) higher probability (intensity/household) of purchase in later rounds in Malawi and Ethiopia, respectively. The state dependency effects causing selective access to seeds are more pronounced in Ethiopia than in Malawi. Seed purchase from available markets hence gives an advantage to smallholder farmers with experience and other benefits that may come with experience (e.g., established networks) compared to new entrants. Further, results reveal that seed purchase decisions also respond to climate variability and shocks. For instance, lagged drought shocks enhance subsequent maize purchase decisions in Malawi and Ethiopia. Historical average rainfall and temperature enhance maize seed purchase decisions in both countries.

Overall, the paper points to state dependency on the demand side of the seed market, leading to selective access to purchased seeds. Further, seed purchase in smallholder farming is a liquidity and risk-dependent input choice. In order to enhance access to seed through purchase and support adaptation to rainfall shocks, policy efforts need to continue targeting reducing transaction costs and other barriers to entry into seed markets.

5. Limitations

The study approach in this dissertation is not without limitations. First, it relies on quasi-experimental data where household cropping activities' data is partly self-reported, including the classification of improved vs. local crop varieties. While local varieties are commonly understood as traditional varieties (aka "landraces"), farmers sometimes refer to locally developed improved varieties as local and sometimes also refer to exotic improved varieties as local after "recycling" seeds as farm-saved seeds for a few seasons (Westengen et al. 2014). In fact, even national and international agricultural research organizations classify improved wheat varieties recycled for more than five seasons as local (Yirga et al. 2013). The point in this study is, however, not to assess the performance of different types of crops but to understand how a diversity of crop varieties are used as coping and adaptation strategies, thus, the self-reported categories improved, and local are useful proxies for diversity below the species level.

Second, the data we use allows comprehension of past shock exposure and vulnerability effects on current farmer actions (ex-post), not directly what they do before exposure (ex-ante risk management). However, studying the impacts of past exposure on current farmer practices shed light on future exposure to shocks and farmers' responses in coping with them. Further, the data used is representative and involves repeated measurements over multiple seasons, allowing more accurate inference of model parameters and comprehension of dynamic as opposed to static relationships-mainly explored in literature.

Third, our analysis relies on combining household data with rainfall and temperature data extracted using location data (longitude and latitude) which are randomly offset within a 5 km distance for confidentiality reasons. The implication is that we did not have access to exact locations where households interviewed reside but only approximate GPS coordinates (locations) of clusters (villages) from which households interviewed reside. However, this is less of a concern given that the climate data source(s) we rely on are of a higher resolution. Besides, covariate temperature and rainfall shocks do not vary much over short distances.

Fourth, although using household panel data offers more advantages than challenges in econometric estimation, dealing with short panels (i.e., 3 to 4 rounds) have some limitations. For instance, short panels spanning a few years allow the estimation of only short to medium-term relationships between shocks and household seed security indicators and not necessarily on the long-term relationships of studied phenomena.

Fifth, the approach taken in the thesis enriches our understanding of the effects of climate variability[3] on studied phenomena (household seed use decisions) but not necessarily on the

[3] According to the IPCC, Climate variability refers to variations in the mean state and other statistics (such as standard deviations, the occurrence of extremes, etc.) of the climate at all spatial and temporal scales beyond that of individual weather events (IPCC 2012; 557).

effects of climate change[4]. Rainfall and temperature shocks used in the thesis reflect mainly on the variance in the current climate (from the long-term mean), whereas climate change is a long term run phenomenon in which other factors such as adaptation and non-linearities could significantly alter relationships (IPCC 2012; Dell et al. 2014; Letta et al. 2018). Hence the results presented in the thesis are better interpreted as the effects of climate variability and not that of climate change.

Despite the noted concerns, results from this research give relevant insights for policy formulation and implementation.

6. Main conclusions and policy implications

6.1. Key conclusions

The findings of the thesis suggest a few general policy-relevant conclusions and implications.

First, smallholder farmers in Ethiopia still rely mostly on local crop varieties than improved varieties, with improved variety use showing a slightly increasing trend over time. Crop production is also highly diversified in Ethiopia compared to Malawi and Tanzania. Also, a significant share (ranging between 40-50%) of farmers across countries are sourcing seeds off-farm through purchase, and the rates of purchasing seeds have increased over time.

Second, seed choice and use decisions and crop diversification respond to rainfall variability and shocks. In Ethiopia, recent exposure to droughts enhances the use of improved varieties. Also, recent exposure to droughts enhance seed purchasing for both improved and local crop varieties in Malawi and Tanzania and mainly for local varieties in Ethiopia. In Ethiopia, flood shocks enhance crop diversification, while drought shocks and long-term rainfall variability enhance crop diversification decisions in Malawian Tanzania. The overall implication is that crop diversification, sourcing seeds off-farm through purchase, and the use of improved seed varieties yield a proper strategy for buffering production risk in smallholder agriculture and, hence, rural development.

Third, households better endowed with assets are more resilient to rainfall shocks. Households with relatively better land and non-land asset endowments increase the use of improved seeds,

[4] Climate change is defined by the IPCC as a change in the state of the climate that can be identified (e.g., by using statistical tests) by changes in the mean and/or the variability of its properties and that persists for an extended period, typically decades or longer (IPCC 2012; 557).

purchase seeds particularly improved seeds, and increase crop diversification post-exposure to rainfall shocks. On the contrary, households poorly endowed with assets intensify local variety use post-drought shock exposure. Also, women's decision-making on input acquisition is associated with lower (higher) chances of purchasing improved (local) seeds following drought shock exposure. The overall implication is that socioeconomic disadvantages and rainfall shocks (e.g., droughts) make households more seed insecure over time.

Fourth, seed purchase and crop diversification decisions are state-dependent. This result points to the importance of knowledge, experience, and networks gained from past decisions to reduce transaction costs in subsequent access to purchased seeds and implement crop diversification. A further implication is that pervasive and nonlinear transaction costs characterize access to seed through the market and consequently reduce household seed security over time. However, the problem is likely to reduce over time for households that can break the first hurdle of entering the market because of established social networks, experience, and market linkages that may marginally reduce transaction costs and improve subsequent access to purchased seeds. For crop diversification, the state-dependency confirms the importance of not only accumulating knowledge, experience, and networks in reducing transaction costs in access to seeds over time. Also, the state-dependency confirms the contributions of past crop diversification decisions to on-farm agrobiodiversity (e.g., through on-farm seed saving), which further supports diversification to meet a multitude of farmer needs over time, including responding to rainfall shocks.

6.2. Policy implications

The conclusions from this research support the call for the promotion of integrated approaches to seed system development. Besides using farmer saved seeds and local seed varieties, farmers also use improved seeds and seeds purchased through both formal and informal channels. Access to off-farm seed through purchasing helps improve the resilience of farmer seed systems to climate risk. As formal channels represent the most likely primary source for new varieties, government policies should address the inequality in access to such channels through supply-side measures such as increasing the production of affordable quality-controlled seeds and demand-side measures such as social protection programs. At the same time, informal seed systems continue to be the backbone of the seed systems farmers use, and seed policies and regulations should enable the co-existence of formal and informal systems. An integrated seed systems approach supported with policies that will reduce inequities in accessing seed from commercial sources will improve the accessibility of improved and local seed varieties and serve the poor and vulnerable groups.

Also, policy efforts should continue targeting reducing transaction costs and other entry barriers into formal and informal seed markets to improve access to seeds. For instance, continued development and upgrading of road infrastructure and agricultural support services such as rural financing and extension are possible interventions. Lower transaction costs will increase the effective demand for purchased seed and enhance farmers' seed security over time.

Given that smallholder farmers require diverse and locally adapted seeds, working on capacitating smallholder farmers to enhance their skills to achieve successful diversification and conserve locally adapted crop genetic resources on-farm (*in situ*) will complement supply-side efforts. For instance, well-designed pro-poor extension services that can (i) capacitate farmers in perfecting on-farm seed saving as a strategy to ensure future access to seed and (ii) give them information on how to implement different crop combinations to buffer against different climate risks (e.g., droughts or floods) at the lowest possible costs are worthwhile options.

References

Ade Freeman H, Ellis F, Allison E (2004) Livelihoods and Rural Poverty Reduction in Kenya Development Policy Review 22:147-171 doi:https://doi.org/10.1111/j.1467-7679.2004.00243.x

Almekinders CJ, Louwaars NP (2002) The importance of the farmers' seed systems in a functional national seed sector Journal of new Seeds 4:15-33

Almekinders CJ, Louwaars NP, De Bruijn G (1994) Local seed systems and their importance for an improved seed supply in developing countries Euphytica 78:207-216

Alobo Loison S (2015) Rural Livelihood Diversification in Sub-Saharan Africa: A Literature Review The Journal of Development Studies 51:1125-1138 doi:10.1080/00220388.2015.1046445

Angelsen A, Dokken T (2018) Climate exposure, vulnerability and environmental reliance: a cross-section analysis of structural and stochastic poverty Environment and Development Economics 23:257-278 doi:10.1017/S1355770X18000013

Angrist JD, Keueger AB (1991) Does compulsory school attendance affect schooling and earnings? The Quarterly Journal of Economics 106:979-1014

Arrow K (1953) Le rôle des valeurs boursiers pour la repartition la meillure des risques. 53 NRS, Paris Cahiers du Seminair d'Economie

Auffhammer M, Hsiang SM, Schlenker W, Sobel A (2013) Using weather data and climate model output in economic analyses of climate change Review of Environmental Economics and Policy 7:181-198

Barrett CB (2008) Smallholder market participation: Concepts and evidence from eastern and southern Africa Food policy 33:299-317

Behrman JR (2000) Basic economic models and econometric tools. In: Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study.

Bellon MR (1996) The dynamics of crop infraspecific diversity: A conceptual framework at the farmer level Economic botany 50:26-39

Binswanger HP, Rosenzweig MR (1986) Behavioural and material determinants of production relations in agriculture The Journal of Development Studies 22:503-539

Brooker RW, George TS, Homulle Z, Karley AJ, Newton AC, Pakeman RJ, Schöb C (2021) Facilitation and biodiversity–ecosystem function relationships in crop production systems and their role in sustainable farming Journal of Ecology 109:2054-2067

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge university press,

Card D, Krueger AB (1994) American economic association The American Economic Review 84:772-793

Carney D (1998) Implementing the sustainable rural livelihoods approach Sustainable rural livelihoods: What contribution can we make 3:27

Caviglia-Harris JL (2004) Household production and forest clearing: the role of farming in the development of the Amazon Environment and Development Economics 9:181-202

Chamberlain G (1982) Multivariate regression models for panel data Journal of econometrics 18:5-46

Chambers RG, Quiggin J (2000) Uncertainty, production, choice, and agency: the state-contingent approach. Cambridge University Press,

Coomes OT et al. (2015) Farmer seed networks make a limited contribution to agriculture? Four common misconceptions Food Policy 56:41-50

Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods Econometrica: Journal of the Econometric Society:829-844
Craig P et al. (2012) Using natural experiments to evaluate population health interventions: new Medical Research Council guidance Journal of Epidemiology and Community Health 66:1182-1186 doi:10.1136/jech-2011-200375
De Janvry A, Fafchamps M, Sadoulet E (1991) Peasant household behaviour with missing markets: some paradoxes explained The Economic Journal 101:1400-1417
Debreu G (1952) A social equilibrium existence theorem Proceedings of the National Academy of Sciences 38:886-893
Dell M, Jones BF, Olken BA (2014) What do we learn from the weather? The new climate-economy literature Journal of Economic Literature 52:740-798
Dercon S (2005) Risk, poverty and vulnerability in Africa Journal of African Economies 14:483-488
Dercon S, Christiaensen L (2011) Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia Journal of development economics 96:159-173
Ellis F (2000) Rural livelihoods and diversity in developing countries. Oxford university press,
Ellis F, Mdoe N (2003) Livelihoods and Rural Poverty Reduction in Tanzania World Development 31:1367-1384 doi:https://doi.org/10.1016/S0305-750X(03)00100-1
Ellis F, Freeman HA (2004) Rural Livelihoods and Poverty Reduction Strategies in Four African Countries The Journal of Development Studies 40:1-30 doi:10.1080/00220380410001673175
Fafchamps M (2004) Market institutions in sub-Saharan Africa. Cambridge, MA: MIT Press,
FAO Developing seed security strategies and programmes for food security in developing countries. In, Rome, Italy, 1998. Proceedings of the international workshop on Seed Security for Food security, 30 November to 1 December 1997. FAO,
FAO (2006) Quality declared seed (QDS) system vol 185. Food and Agriculture Organization of the United Nations, Rome, Italy
FAO (2016) Seed Security Assessment: A Practitioner's Guide; Food and Agriculture Organization of the United Nations (FAO). Food and Agriculture Organization of the United Nations, Rome, Italy
FAO, ECHA (2015) Building Capacity for Seed Security Assessments: Household Seed Security Concepts and Indicators, Discussion Paper. FAO, and European Commission Humanitarian Aid. .
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas International journal of climatology 37:4302-4315
Finkel SE (1995) Causal analysis with panel data. vol 105. Sage,
Fitzgerald J, Gottschalk P, Moffitt R (1998) An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics Journal of Human Resources:251-299
Gardner W, Mulvey EP, Shaw EC (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models Psychological bulletin 118:392
Gebru M, Holden ST, Tilahun M (2019) Tenants' land access in the rental market: evidence from northern Ethiopia Agricultural Economics 50:291-302
Gebru M, Holden ST, Alfnes F (2021) Adoption analysis of agricultural technologies in the semiarid northern Ethiopia: a panel data analysis Agricultural and Food Economics 9:1-16
Gelaro R et al. (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2) Journal of Climate 30:5419-5454

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset International journal of climatology 34:623-642

Haug R, Hella JP, Nchimbi-Msolla S, Mwaseba DL, Synnevag G (2016) If technology is the answer, what does it take? Development in Practice 26:375-386

Heckman J (1979) Sample selection bias as a specification error. Econometrica 47:153–161

Holden ST, Quiggin J (2017) Climate risk and state-contingent technology adoption: shocks, drought tolerance and preferences European Review of Agricultural Economics 44:285-308

Holden ST, Deininger K, Ghebru HH (2007) Impact of land certification on land rental market participation in Tigray region, Northern Ethiopia Northern Ethiopia (October 1, 2007)

Holden ST, Otsuka K, Place FM (2010) The emergence of land markets in Africa: Impacts on poverty, equity, and efficiency. Routledge,

Hsiao C (2007) Panel data analysis—advantages and challenges Test 16:1-22

Hufford MB, Teran JCBMy, Gepts P (2019) Crop Biodiversity: An Unfinished Magnum Opus of Nature Annual Review of Plant Biology 70:727-751 doi:10.1146/annurev-arplant-042817-040240

Imbens G, Angrist J (1994) Estimation and identification of local average treatment effects Econometrica 62:467-475

IPCC (2012) Glossary of terms: A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). In: Field CB et al. (eds) Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp 555-564

Jarvis DI, Hodgkin T, Sthapit BR, Fadda C, Lopez-Noriega I (2011) An heuristic framework for identifying multiple ways of supporting the conservation and use of traditional crop varieties within the agricultural production system Critical Reviews in Plant Sciences 30:125-176

Kansiime MK, Mastenbroek A (2016) Enhancing resilience of farmer seed system to climate-induced stresses: Insights from a case study in West Nile region, Uganda Journal of Rural Studies 47:220-230 doi:https://doi.org/10.1016/j.jrurstud.2016.08.004

Kassie G, MacRobert J, Setimela P, Erenstein O, Mwangi W, Shiferaw B (2013) Political and economic features of the maize seed industry in southern Africa Agrekon 52:104-127

Katengeza SP, Holden ST, Lunduka RW (2019a) Adoption of drought tolerant maize varieties under rainfall stress in Malawi Journal of Agricultural Economics 70:198-214 doi: https://doi.org/10.1111/1477-9552.12283

Katengeza SP, Holden ST, Fisher M (2019b) Use of Integrated Soil Fertility Management Technologies in Malawi: Impact of Dry Spells Exposure Ecological Economics 156:134-152 doi:https://doi.org/10.1016/j.ecolecon.2018.09.018

Lammerts van Bueren ET, Struik PC, van Eekeren N, Nuijten E (2018) Towards resilience through systems-based plant breeding. A review Agronomy for Sustainable Development 38:42 doi:10.1007/s13593-018-0522-6

Langyintuo AS, Mwangi W, Diallo AO, MacRobert J, Dixon J, Bänziger M (2010) Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private–public intervention to promote growth Food Policy 35:323-331

Letta M, Montalbano P, Tol RS (2018) Temperature shocks, short-term growth and poverty thresholds: Evidence from rural Tanzania World Development 112:13-32

Lin BB (2011) Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change BioScience 61:183-193 doi:10.1525/bio.2011.61.3.4
Louwaars NP, de Boef WS (2012) Integrated Seed Sector Development in Africa: A Conceptual Framework for Creating Coherence Between Practices, Programs, and Policies Journal of Crop Improvement 26:39-59 doi:10.1080/15427528.2011.611277
Love B, Spaner D (2007) Agrobiodiversity: Its value, measurement, and conservation in the context of sustainable agriculture Journal of Sustainable Agriculture 31:53-82
Masarie KA, Tans PP (1995) Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record Journal of Geophysical Research: Atmospheres 100:11593-11610
McGuire S, Sperling L (2013) Making seed systems more resilient to stress Global Environmental Change 23:644–653 doi:http://dx.doi.org/10.1016/j.gloenvcha.2013.02.001
McGuire S, Sperling L (2016) Seed systems smallholder farmers use Food Security 8:179-195
Mulesa TH, Dalle SP, Makate C, Haug R, Westengen OT (2021) Pluralistic Seed System Development: A Path to Seed Security? Agronomy 11:372
Mundlak Y (1978) On the Pooling of Time Series and Cross Section Data Econometrica 46:69-85 doi:10.2307/1913646
Nordhagen S, Pascual U (2013) The impact of climate shocks on seed purchase decisions in Malawi: implications for climate change adaptation World Development 43:238-251
Pandey R, Jha SK, Alatalo JM, Archie KM, Gupta AK (2017) Sustainable livelihood framework-based indicators for assessing climate change vulnerability and adaptation for Himalayan communities Ecological Indicators 79:338-346 doi:https://doi.org/10.1016/j.ecolind.2017.03.047
Quiggin J, Chambers RG (2006) The state-contingent approach to production under uncertainty Australian Journal of Agricultural and Resource Economics 50:153-169
Renkow M, Hallstrom DG, Karanja DD (2004) Rural infrastructure, transactions costs and market participation in Kenya Journal of development economics 73:349-367
Ricker-Gilbert J, Chamberlin J (2018) Transaction costs, land rental markets, and their impact on youth access to agriculture in Tanzania Land Economics 94:541-555
Rose E (2001) Ex ante and ex post labor supply response to risk in a low-income area Journal of development economics 64:371-388
Scoones I (1998) Sustainable rural livelihoods: a framework for analysis
Sheahan M, Barrett CB (2017) Ten striking facts about agricultural input use in Sub-Saharan Africa Food Policy 67:12-25 doi:https://doi.org/10.1016/j.foodpol.2016.09.010
Singh I, Squire L, Strauss J (1986) A survey of agricultural household models: Recent findings and policy implications The World Bank Economic Review 1:149-179
Sperling L (2002) Emergency seed aid in Kenya: some case study insights on lessons learned during the 1990s Disasters 26:329-342
Sperling L (2020) Seed security response during COVID-19: building on evidence and orienting to the future Food Security:1-5
Sperling L, Gallagher P, McGuire S, March J, Templer N (2020) Potential Seed Traders: The Backbone of Seed Business and African Smallholder Seed Supply Sustainability 12:7074
Tione SE, Holden ST (2021a) Can rainfall shocks enhance access to rented land? Evidence from Malawi Agricultural Economics
Tione SE, Holden ST (2021b) Non-convex transaction costs and land rental market participation in Malawi Land Economics:062620-060092R

Tobin J (1958) Estimation of relationships for limited dependent variables Econometrica: journal of the Econometric Society:24-36
Tripp R (1997) New seed and old laws: regulatory reform and the diversification of national seed systems. Intermediate Technology Publications Ltd (ITP),
Tripp R (2006) Strategies for seed system development in sub-Saharan Africa: A study of Kenya, Malawi, Zambia, and Zimbabwe. SAT eJournal 2:1-58
van Gastel TJG, Gregg BR, Asiedu EA (2002) Seed Quality Control in Developing Countries Journal of New Seeds 4:117-130 doi:10.1300/J153v04n01_09
Walther G-R et al. (2002) Ecological responses to recent climate change Nature 416:389-395
Westengen OT, Ring KH, Berg PR, Brysting AK (2014) Modern maize varieties going local in the semi-arid zone in Tanzania BMC Evolutionary Biology 14:1 doi:10.1186/1471-2148-14-1
Wood D (1997) The conservation of agrobiodiversity on-farm: questioning the emerging paradigm Biodiversity & Conservation 6:109-129
Wooldridge JM (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity Journal of applied econometrics 20:39-54
Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press,
Yirga C, Mohammad A, Kassie M, Groote Hd, Mebratu T, Jaleta M, Shiferaw B (2013) Analysisof Adoption and Diffusion of Improved Wheat Technologies inEthiopia. Ethiopian Institute of Agricultural Research (EIAR) and CIMMYT,

Paper I

World Development 159 (2022) 106054

Contents lists available at ScienceDirect

World Development

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

Crops in crises: Shocks shape smallholders' diversification in rural Ethiopia

Clifton Makate [a,*], Arild Angelsen [a], Stein Terje Holden [a], Ola Tveitereid Westengen [b]

[a] School of Economics and Business, Norwegian University of Life Sciences, P.O.Box 5003, 1432 Ås, Norway
[b] Department of International Environment and Development Studies, Noragric, Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway

ARTICLE INFO

Article history:
Accepted 22 July 2022

Keywords:
Shock exposure
Seed use
Crop diversity
Smallholder farmers
Ethiopia

ABSTRACT

Crop diversity plays a central role in smallholder farmers' ability to cope with and adapt to shocks. Shifting crop varieties and diversifying the crop portfolio are common risk reduction strategies. This paper addresses the influence of covariate climate shocks and idiosyncratic socioeconomic shocks on crop variety use and crop species diversification by smallholder farmers using nationwide balanced panel data (2011/12, 2013/14, & 2015/16) from rural households in Ethiopia combined with village-level historical monthly rainfall and temperature data. We apply correlated random effects models, which control for time-invariant household unobservables. Past exposure to drought shocks increased the use of improved seed varieties in general and for wheat, while long-term average rainfall and lagged flood shocks enhance crop species diversity. Lagged temperature shocks increase improved seed use and crop species diversity. However, recurrent drought exposure and exposure to relatively more severe drought shocks significantly reduced overall agricultural activity. Idiosyncratic shocks, to a much lesser degree, influenced seed use and crop diversification decisions compared to covariate drought shocks. Heterogeneity analysis revealed that drought shock exposure on farmers with less than average farm sizes and other assets - compared to those better-off – increased their relative reliance on local seed use, reduced crop diversification, and reduced improved seed use. The results are robust to various sensitivity checks. Our findings are relevant for policy responses aiming to strengthen smallholders' ability to cope with and adapt to shocks: farmers' seed-based risk reduction strategies rely on access to seeds from both formal and informal seed systems, but policies addressing economic inequality are needed to enhance access to improved seeds and crop diversity for resource-poor socioeconomic groups.

1. Introduction

Seeds are essential assets in smallholder farmers' portfolio of coping and adaptation strategies during periods of environmental and socioeconomic stress. With access to a wide variety of seeds, farmers can choose crop varieties suited to local conditions. Farmers may also diversify their portfolios and reduce production risk by growing a broader range of local and improved crop varieties. Therefore, access to seed is considered an essential aspect of seed, food, and livelihood security in the wake of both acute and chronic stress situations (Bezner, 2022; Howden et al., 2007; McGuire & Sperling, 2013; Mortimore & Adams, 2001; Sperling, 2020).

Livelihoods in Sub-Saharan Africa are vulnerable to both covariate and idiosyncratic shocks. Covariate shocks universally affect many households living in the same geographic location (e.g., climate shocks and epidemics), while idiosyncratic shocks affect specific households and one household's experience is not related to the experience of neighboring households, such as illness, death, or loss in employment (Dercon 2004, 2005; Pradhan & Mukherjee, 2018). Agriculture is a key pillar in rural livelihoods, and exposure to both types of shocks are common and affects access to agricultural inputs and thereby coping and adaptation strategies. For instance, the COVID-19 pandemic has disrupted access to key inputs by farmers (including seed) and increased logistical, administrative, and transaction costs for farmers (Sperling, 2020). The pandemic has thus added to already struggling agri-food systems in the region. Understanding how farmers cope and adapt to shocks is important to develop evidence-based policy responses to improve their seed, food, and livelihood security.

* Corresponding author.
E-mail addresses: clifton.makate@nmbu.no (C. Makate), arild.angelsen@nmbu.no (A. Angelsen), stein.holden@nmbu.no (S.T. Holden), ola.westengen@nmbu.no (O. T. Westengen).

https://doi.org/10.1016/j.worlddev.2022.106054
0305-750X/© 2022 The Author(s). Published by Elsevier Ltd.

A meta-study for the Intergovernmental Panel on Climate Change Assessment Report (IPPC AR5) tested the relative adaptation effect of a range of on-farm adaption measures and found cultivar adjustment to be one of the most effective methods (Challinor et al., 2014; IPCC 2014). The most recent IPCC report (AR6) furthermore emphasizes the adaptation potential in crop diversification (IPCC 2022). Other studies have found greater crop diversity to be associated with enhanced livelihood outcomes, including higher temporal food production stability at both the household (Asfaw, Scognamillo, Caprera, Sitko, & Ignaciuk, 2019; Bozzola & Smale, 2020; Di Falco, Bezabih, & Yesuf, 2010; Makate, Wang, Makate, & Mango, 2016; Mulwa & Visser, 2020) and national levels (Renard & Tilman, 2019). However, for such crop-based adaptation methods to be effective, farmers need to be seed secure.

Seed security entails that farmers have access to quality seeds of well-adapted varieties that meet their needs and preferences (FAO 2018; Sperling, 2020). Farmers access seeds through seed systems, which encompass the chains of actors, institutions and activities involved in the development, distribution, and use of seeds (Almekinders, Louwaars, & De Bruijn, 1994; Sperling, Cooper, & Remington, 2008). In developing countries, the formal seed system delivering seeds of improved varieties released by plant breeders and certified by seed inspection authorities supplies only a small share of the total volume of seeds used while informal sources such as seed saving from own harvest, sourcing through social networks and local markets supply the bulk of seeds used (Louwaars & de Boef, 2012). Thus, crop-based adaptation to climate change relies on well-adapted varieties (technology) as well as well-functioning seed systems (institutions).

The uptake of improved varieties' remains low in many parts of SSA, although it has increased over time (Sheahan & Barrett, 2017). In their review of adoption studies Acevedo et al. (2020) have shown that unavailability of improved seeds, inadequate information, lack of complementary farming inputs, and high seed prices are common barriers to adoption of climate-resilient crops. Moreover, farmer preferences for traits not present in modern varieties could explain some of the low adoption rates (Fisher et al., 2015), but adoption studies rarely explore the reason for growing other varieties than the improved ones. This is a research gap since improved varieties, and formal seed systems supply only a small share of the seeds used by smallholders in developing countries (Coomes et al., 2015). The study of local crop varieties (i.e., varieties of local origin, selected by farmers) has typically been the domain of a branch within crop science. This literature on genetic resources and seed systems has shown that local varieties sourced through informal seed systems play a role in the coping behavior and adaptation to various stressors (Abay, Waters-Bayer, & Bjørnstad, 2008; Mekbib, 2007) and that farmers often mix local and improved varieties to serve different needs and to minimize risks in their households (Bellon & Hellin, 2011; Westengen, Ring, Berg, & Brysting, 2014). There is thus a need for more knowledge about how smallholder farmers' seed use, including seeds of both local and improved varieties, is influenced by shocks.

Smallholder farmers may respond to shocks by diversifying their livelihood strategies both on– and off-farm (Asfaw et al., 2019; Morton, 2007; Mulwa & Visser, 2020). In a risky context with imperfect input and output (food) markets, it is assumed that low-income families can minimize their exposure to future shocks by diversifying their activities and growing enough food for subsistence (Fafchamps, 1992; Kurosaki & Fafchamps, 2002). Economic theories that study farmer behavior under risk, such the state-contingent theory of adaptation by Chambers and Quiggin (2000) are useful for explaining farmers' responses to previous shock exposure. The state-contingent theory of adaptation assumes that farming households make production decisions to maximize anticipated utility of returns in different states of nature, e.g., states with and without climate shocks (Holden & Quiggin, 2017). Therefore, production risks and shocks, farmers' perceptions of those risks based on shock experiences as well as risk preferences, influence farming decisions. Emerging studies that incorporate climate shocks and risk attitudes and behavior have shown that lagged climate shock exposure leads to higher uptake of improved (drought-tolerant) varieties, and more so by more risk-averse farmers (Holden & Quiggin, 2017; Katengeza, Holden, & Lunduka, 2019). Employing the state-contingent theory of adaptation, this study evaluates the influence of lagged shock exposure on variety use and crop diversification practices in rural Ethiopia. We analyze a panel data set compiled from three rounds (2011/12, 2013/14, and 2015/16) of the Living Standards Measurement Study-Integrated surveys on Agriculture (LSMS-ISA) for Ethiopia. More specifically, this study aims to answer the following research questions: (i) How does exposure to covariate and idiosyncratic shocks influence the types of seeds used by farmers and the extent of crop diversification? (ii) How does household diversity in asset wealth endowments, land size holding, and access to social safety nets mediate the influence of covariate shock exposure on seed use and diversification decisions in rural Ethiopia?

The rest of the article is organized as follows. Section 2 briefly discusses the Ethiopian context, while section 3 lays out the conceptual framework. Section 4 describes the methodology and presents descriptive statistics. Section 5 presents the results, while section 6 discusses them. Section 7 concludes and presents policy implications.

2. The Ethiopian context

Agriculture is a key source of employment and income in low and middle-income countries, including Ethiopia. Due to environmental and cultural diversity and heterogeneity, Ethiopia is the centre of origin and diversity of various food crops, and farmers also today grow multiple crops for both consumption and commercial purposes (Dessie, Abate, Mekie, & Liyew, 2019). Food production is dominated by smallholders, as they cultivate approximately 96 % of the total area devoted to food production (Taffesse, Dorosh, & Gemessa, 2012). There are two main rainy seasons (Meher and Belg) and hence two cropping seasons. The Meher season is the most important season for crop production, with more than 90 % of total cereal production. Five major cereal crops are at the core of Ethiopia's agriculture and food production economy: teff, maize, sorghum, wheat and barley.

The current Ethiopian seed policy promotes an integrated seed sector development that recognizes the complementarity between the country's different seed systems (MoA 2019). The national seed policy and Pluralistic Seed System Development Strategy, (released in 2013 and adopted in 2017) (MoA and ATA 2017), provides the legal basis for the co-existence of formal and informal seed systems. It also includes provisions to support interventions in both formal and informal systems and promote an emerging 'intermediate' system. The intermediate seed system has grown considerably under the new strategy and includes Seed Producer Cooperatives (SPC) producing Quality Declared Seeds (QDS) of improved varieties (Sisay, Verhees, & van Trijp, 2017). Informal seed systems provide the bulk of the seeds used by farmers in the country (Thijssen, Bishaw, Beshir, De Boef, & (eds)., 2008), but for some crops, including vegetable seeds, hybrid maize and wheat, the formal system supplies a significant share of the certified seeds of improved varieties (Alemu & Bishaw, 2015; Erenstein & Kassie, 2018).

Ethiopia, as with many developing regions is not spared for recurrent shock exposure. Common shocks in history include covariate weather shocks (drought, flood, and other weather shocks), covariate economic shocks (price shocks in input and output markets), conflicts, and idiosyncratic shocks such as illness, death, family break-ups, loss of formal employment, loss of livestock to theft and predation (Dercon, 2004; Porter, 2012). Several shocks have been experienced in different parts of Ethiopia for the study period (2011–2016) and afterward. According to the International disaster database (EM-DATA[1]), major recent shocks include: a major drought of 2011 which was experienced in most parts of the country (e.g., Dire Dawa, Gambela, Harari, Oromia, SNNP, Somali and Addis Ababa) and affected approx. 1 million people; the El Nino drought of 2015/2016 seasons (experienced in Afar, Somali, Oromia, Amhara, and SNNP) which affected about 10.2 million people; flash floods (in Wolayita district in SNNP region, and Bale district in Oromia region) which affected close to half a million people in 2016. More recent examples include the locust outbreak which started in November 2019 (experienced in Afar, Amhara, Oromia, Somali, Tigray regions), the ongoing Covid-19 pandemic, and the civil war (since November 2020).

Both covariate and idiosyncratic shocks affect livelihoods and are usually linked to a reduction in assets, fall in incomes, and a significant reduction in consumption. However, smallholder farming households usually find it easier to cope with idiosyncratic household shocks than to covariate weather shocks (Dercon, 2005; Nguyen, Nguyen, & Grote, 2020).

Since 2005, there has been growing political momentum around social protection and cushioning of the most vulnerable from the impacts of shocks in Ethiopia. Safety net programs such as the Productive Safety Net Program (PSNP) introduced in 2005 have been very important for household food security, in particular in areas with chronic food insecurity. These programs represent the main source of insurance against shocks and household food insecurity and include food-for-work, cash-for-work, and free food distribution outside the main growing season for eligible households and communities. The PSNP program was designed to serve three main purposes: (a) smoothing food consumption for the poor and food-insecure through food or cash transfer during periods of stress, (b) cushioning household asset depletion due to shocks and other socioeconomic stressors, and (c) building community assets using the public works component (food or cash-for-work) that has been focused on building village and feeder roads(Debela, Shively, & Holden, 2021; Dejene & Cochrane, 2021).

Moreover, until recently economic progress in Ethiopia has reduced poverty and enhanced resilience to shocks. High economic growth, combined with continued population growth, has resulted in a rapid rural transformation process in Ethiopia, with fast-growing rural towns and larger cities and diversification of the economy (Bezu & Holden, 2014; Holden & Tilahun, 2020; Masters et al., 2013). Also, farm sizes have reduced over time, resulting in agriculture intensification (Masters et al., 2013).

3. Conceptual framework

Farmers' seed-related adoption decisions under risk may be analyzed within the state-contingent framework of Chambers and Quiggin (2000). Within this framework, smallholder farmers make input decisions before weather conditions are revealed. Production decisions under uncertainty are made to maximize average utility of returns in different states of nature (Holden & Quiggin, 2017). We assume that the vulnerability of households is closely associated with their resource poverty. Their most important resources are their availability of land and labor endowments relative to their consumption needs. Land- and labor-poor households are therefore assumed to be more vulnerable to shocks. We also assume that it is more difficult to use social networks to protect oneself against covariate shocks than against idiosyncratic shocks, making interventions such as the safety net programs more important as protection against covariate shocks such as droughts. Furthermore, drought shocks have direct effects on the performance of the crops and varieties grown and on market prices and the availability of essential commodities. Improved varieties may or may not perform better than the local varieties in different environments and under different states of nature.

Shocks may alter the household's farming activities in heterogeneous ways. The literature distinguishes between *ex-post* risk coping mechanisms (what farmers do after exposure to shocks) and what they do before exposure (*ex-ante* risk management) (Angelsen & Dokken, 2018). Characteristics of the rural settings such as over-reliance on agriculture, lack of functional insurance markets, and the dire consequences of a bad season (Dercon, 2005; Rose, 2001) complicate both *ex-post* and *ex-ante* response to shocks. Households may switch from selling food in years with good rainfall and becoming net buyers in years with poor rainfall. Covariate risk implies that such rainfall shocks occur simultaneously to households in large geographical areas with the consequence that most of them are net sellers in good years, and most are net buyers of food in bad years. Therefore, poor market integration leads to low food prices when they are net sellers and high food prices when they are net buyers. Holden and Shiferaw (2004) found that the indirect price effects were stronger than the direct production loss effects of such shocks in Ethiopia. Households may resort to the selling of livestock and assets as a coping mechanism after shock exposure and, they may engage in the diversification of income portfolios to prepare themselves for future shocks (Dercon & Christiaensen, 2011; Dercon, 2005). For instance, Gebregziabher and Holden (2011) found that in Tigray, Ethiopia, when households exhaust selling their assets, they distress rent out their land after shock exposure to get urgent cash. Gebru, Holden, and Alfnes (2021) used household panel data to study the adoption of improved wheat and drought-tolerant teff in northern Ethiopia and found that higher rainfall in the previous year was associated with more adoption of drought-tolerant teff.

This paper focuses on understanding smallholder farmers' behavioral responses in their seed use decisions to climate variables, particularly previous shock exposure. Different crop varieties may perform differently with and without shocks, and farmers exposed to shocks are likely to discover and learn the different benefits associated with different varieties. We consider long-term climate variables, lagged idiosyncratic and covariate shocks as our main test variables.

We, however, take cognizant that the behavior of farmers and their preferences will be related to resource endowments (wealth, education) and other household characteristics. Therefore, we control for household resource endowments, such as household tropical livestock units, farm size, access to productive safety nets, and asset wealth. We also control household characteristics, such as the number of literate household members, household dependency ratio, age, gender, and marital status of the household head.

Following the literature on agricultural household modelling (De Janvry, Fafchamps, and Sadoulet (1991)), sustainable livelihoods literature (Ellis, 2000) and the agricultural adoption literature (Acevedo et al., 2020; Takahashi, Muraoka, & Otsuka, 2020), farmer's decisions to choose a given farming practice or technology also market-related factors. For instance, responding to market imperfections and failure (resulting in large price bans between selling and purchasing prices), farmers may grow a combination

[1] The EM-Data is a global database on natural and technological disasters(shock s), capturing the occurrence of disasters in the world from 1900 to the present. The EM-DAT is maintained by the Centre for Research on the Epidemiology of Disasters (CRED) at the School of Public Health of the Université catholique de Louvain located in Brussels, Belgium. The data is accessible online via the following link(https://www.emdat.be/).

of varieties or diversify crop production to cover the household's consumption needs (Alobo Loison, 2015). Hence, we control for variables that proxy market access, including distance to markets and distance to nearest paved road.

The impact of shocks is likely to be heterogeneous on farmers with different vulnerability levels, which possibly shape 'farmers' responses to seed variety use and diversification decisions. Poorer farmers (farmers less endowed with assets) and farming households who lack formal insurance options tend to be more vulnerable to shock exposure (Dercon & Christiaensen, 2011; Dercon, 2005). The behavioral impact of shocks on disadvantaged households often takes the form of adopting low-risk activities as risk management strategies at the expense of lower mean returns and incomes. To test for heterogeneity in the behavioral response to shocks, we also test the effect of interaction effects. We consider (i) land size inequality, (ii) access to social safety nets, and (iii) asset wealth inequality in assessing conditioned impacts of drought shocks. This study, therefore, seeks to answer the research questions by testing the following hypotheses:

First, we hypothesize that past exposure to adverse rainfall shocks increases the use of improved seeds and crop diversification through both push and pull factors. Past exposure to rainfall shocks can affect households' ability to produce and save their own seed, thus acting as a push factor increasing their propensity to access improved seed through the formal seed systems and/or diversify the crop portfolio in the following year. On the other hand, past exposure to drought shocks may also promote learning on the performance of varieties hence pulling them towards improved varieties and more diverse cropping to adapt farming to future shock exposure.

*Second, we hypothesize that farmers in disadvantaged positions (*e.g.*, the poor) to a lesser degree than better-off farmers use improved varieties and/or diversify their cropping portfolio post-exposure*. Farmers in disadvantaged positions (i.e., those with poor assent endowments) and those without access to formal insurance options are more vulnerable to shock exposure (Dercon & Christiaensen, 2011; Dercon, 2005).. Hence, we expect poorer farmers to be more likely to be pushed towards less costly crop use options (e.g., use of local seeds) as *ex-post* risk management strategies.

4. Data and methods

4.1. Data

The study uses of a rich panel data set from three rounds of the Ethiopian Socioeconomic Survey (ESS) combined with monthly weather data (rainfall and temperature) for the period 1980 to 2017. The ESS is administered by the Ethiopian Central Statistical Agency in collaboration with the World Bank's Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) project. Three-panel rounds of the data for Ethiopia are publicly available on the World Bank website.[2] We construct a three-year balanced household panel of 2 398 rural households interviewed successively in three-panel rounds (2011/12, 2013/14, 2015/16). The three-year household panel for Ethiopia started with 3 969 households, of which 3 466 (87 %) were rural in 2011/12. We trace rural households successively interviewed in all three rounds, with consistent household identification information, and usable data on agricultural activities, including seed use information to construct a balanced panel. The ESS data is representative at the national and regional level for rural areas and the four largest regions in Ethiopia: Oromia, Amhara, Tigray and Southern Nations, Nationalities, and People's Region (SNNP) (Aguilar, Carranza, Goldstein, Kilic, & Oseni, 2015).

[2] The data sets are publicly available on https://surveys.worldbank.org/lsms.

The historical climate data are from WorldClim (Fick & Hijmans, 2017; Masarie & Tans, 1995), and were used to define historical rainfall and temperature variables and lagged shock variables. We link survey data with historical climate data by using geo-referenced data available at the Enumeration Area (EA) level, which is the smallest sampling unit (the village) for LSMS-ISA data. Further details on LSMS-ISA data descriptions and on how we processed weather data, generated lagged shock variables, and merged it with household-level data are given as part of the Supplementary material (appendix).

4.2. Model and variable specification

We model the farmers' decisions to select and or adopt crop varieties and diversify cropping portfolios using limited dependent variable models (Wooldridge, 2010). We assume the farmer aims to maximize overall welfare from their decisions, which implies that they choose seed type and farming practices that maximize anticipated utility or minimize production risks subject to constraints. Farmers' seed and diversification decisions are based on several factors as given in the conceptual framework and these may include weather expectations for that season, resources available, and characteristics of the farming technology or practice (Ding, Schoengold, & Tadesse, 2009; Katengeza, Holden, & Lunduka, 2019). To evaluate the impact of shock exposure on farming household's seed use decisions and diversification, we, therefore, apply appropriate limited dependent variable models within the Correlated Random Effects (CRE) approach, which controls for time-invariant household unobservables in a similar way as household fixed effects do when continuous dependent variables are used (Wooldridge, 2019). For seed use (dummies) and intensity of use (quantities), we use CRE logit and Tobit models, while for crop count and Simpson indices of crop diversification we use CRE Poisson and Tobit models respectively. We share more details on the CRE approach including its merits in the next section (Model estimation and justification).

Farmers' decisions to use improved or local seed are modeled, first as binary decision variables, and second as a censored outcome variable measuring the intensity of use as shown in Equations (1) and (2).

Binary decision variables (logit model):

$$P(Q_{it} = 1|C_{vt}, S_{it}, H_{it}, YR_t, \sigma_i) \\ = \mathrm{F}(\theta_0 + \theta_1 C_{vt} + \theta_2 S_{it} + \theta_3 H_{it} + \theta_4 YR_t + \sigma_i + \varepsilon_{it}), \tag{1}$$

Censored outcome variables (Tobit model):

$$Q_{it} = \max(0, \gamma_0 + \gamma_1 C_{vt} + \gamma_2 S_{it} + \gamma_3 H_{it} + \gamma_4 YR_t + \sigma_i + \epsilon_{it}) \tag{2}$$

Count outcome variables (Poisson model):

$$E(Q_{it}|C_{vt}, S_{it}, H_{it}, YR_t, \sigma_i) = \sigma_i \exp(\beta_0 + \beta_1 C_{vt} + \beta_2 S_{it} + \beta_3 H_{it} \\ + \beta_4 YR_t + \sigma_i + \varrho_{it}) \tag{3}$$

For seed use type decisions, Q_{it} is the dependent variable and represents different values for the use and intensity of use decisions. In the first stage, seed use estimation (use decision) Q_{it} is a dummy variable equal to one if household i used improved (local) seed in year t, and zero otherwise. This practice is done in general for all crops (improved seed and local seed) and specifically for maize and wheat, which are important cereal crops in the Ethiopian basket of food crops (Rashid & Minot, 2010). For the intensity of local and improved seed use, Q_{it} is measured as the quantity of local and improved seed used by the household (self-reported), respectively. Seed use intensity variables are all log-transformed to reduce heteroscedasticity and make our data more normally distributed.

For crop diversification, we use the count and the Simpson indices of diversity and model the respective crop outcome variables as shown in equations (2) and (3). The crop count index measures the number of cultivated crops (richness), and it is based on the assumption that all crops contribute equally to the household crop portfolio, which is not often the case (Tesfaye & Tirivayi, 2020). The Simpson index overcomes the weaknesses of the count index as it measures not only richness but the relative abundance of each species (evenness).

C_{vt}, and S_{it} are respectively, vectors of covariate and idiosyncratic shock variables. In the vector of idiosyncratic shocks (S_{it}), we include major loss of livestock and loss of formal employment by a household member in the recent past. In the vector (C_{vt}), we include objective measures of covariate climate shocks. We follow related studies, for example Katengeza, Holden, & Lunduka (2019),[3] and measure one and two-year lag measures of climate shock exposure in the *Meher* season. The *Meher* season is the most important season for agricultural production, with more than 90 % of total cereal production in Ethiopia (Taffesse et al., 2012). We follow studies by Michler, Baylis, Arends-Kuenning, and Mazvimavi (2019), and Ward and Shively (2015) and define temperature and rainfall shocks as normalized deviations in a single season's climate variable (rainfall and temperature) from the expected seasonal climate variable, as defined by its historical average. We define rainfall and temperature shocks accordingly as follows:

a) $Rainshock_{vt} = \left[\frac{rain_{vt} - \overline{rain}_v}{\sigma_{rain_v}}\right]$, where $Rainshock_{vt}$ is a rainfall shock measure for a cluster(village) (v), in the year (t), and $rain_{vt}$ is the observed amount of rainfall for the defined period (season), $\overline{rain}_v$ is the average seasonal rainfall for the village(v) over the 38 years (1980–2017), and, σ_{rain_v} is the standard deviation of rainfall during the same period.

We follow the same approach and define temperature shocks as follows:

b) $Tempshock_{vt} = \left[\frac{temp_{vt} - \overline{temp}_v}{\sigma_{temp_v}}\right]$b), where $Tempshock_{vt}$ is a temperature shock measure for a cluster (village) (v), in the year (t), and $temp_{vt}$ is the observed temperature for the defined period (season), $\overline{temp}_v$ is the average seasonal temperature for the village(v) over the 38 years (1980–2017), and, σ_{temp_v} is the village-level standard deviation of temperature during the same period.

These two measures are symmetric in the way that higher than normal rainfall or temperature having have the same effects – just with the opposite sign - as lower rainfall or temperature. Given our interest in testing for the influence of drought shocks (negative Z-scores) we split the rainfall shock variable in (a) into positive and negative rainfall deviations (Z-scores) and term the negative Z-scores drought shock. Our measure of drought shock is hence defined and split as follows:

c)

$$Droughtshock_{vt} = \left\{ \left[\frac{rain_{vt} - \overline{rain}_v}{\sigma_{rain_v}}\right] if\ rain_{vt} < \overline{rain}_v,\ and\ 0\ otherwisec \right),$$

where σ_{rain_v} is the village-level standard deviation of the cumulative rainfall for the months May-September over the 38-year period from 1980 to 2017. The resultant drought shock will have negative rainfall Z-scores ranging from - x to 0 and is summarized in Figure 2. To facilitate direct and more intuitive interpretations of results on the influence of drought shocks (negative Z-scores) on seed use and diversification decisions, in all our regressions we take the absolute value of the negative drought shocks measured as Z-scores. For all the shock measures in (a, b, and c), we measure 1 and 2-year lags from the reference season. We specifically define all the climate variables for two periods: the *Meher* season and the early season of the *Meher* season (May to July). We use the latter to test for early season shocks in our regressions, given that such shocks can have more drastic effects on crop production (Elagib, 2015). We first test for the effects of general temperature and rainfall shock variables, and then we specifically test for drought shocks.

We merge shock variables to household data based on the year (reference season) in which agricultural data for households was collected. We also include the historical mean of rainfall and temperature (1981–2017) of the early season for the *Meher* season in all our regressions. We include temperature variables in our regression to avoid potential omitted variable bias if we exclude temperature, given that crop production responds both to rainfall and temperature. The vectors for covariate shocks (C_{vt})(rainfall and temperature shocks) and idiosyncratic shocks (S_{it}) (losing livestock & formal employment) represent our key "treatment" variables in a natural experiment approach. Hence, we treat them as exogenous variables and discuss their impacts rather than only assess their correlations with seed use and diversification decisions.

We control for other household socioeconomic variables (H_{it}) in our seed use and diversification equations, including household wealth variables (e.g., agricultural asset index[4] and farm size), human capital variables (e.g., education), access to social safety nets (e.g., Productive Safety Net (PNSP) program), and other field related characteristics. We control for additional covariates mainly as a robustness check to our main findings. The vector YR_t represents year dummies, and the year 2011/12 is used as the reference. Finally, σ_i, captures individual household time-invariant effect while $\varepsilon_{it}, \epsilon_{it}$, and ϱ_{it} are the idiosyncratic error terms.

4.3. Model estimation and justification

Parameters in equations (1), 2, and 3 are estimated using the correlated random effects (CRE) model, as proposed by Mundlak (1978) and Chamberlain (1984). In line with the CRE approach, we assume that the unobserved heterogeneity can be replaced with its linear projection onto the time averages of all household level regressors (Chamberlain, 1982; Mundlak, 1978). Hence, in estimating equations (1), 2, and 3, we add the means (across years) of variables in the vector of socioeconomic variables (H_{it}) as additional controls. The CRE approach is preferred over the traditional random effects (RE) model because it relaxes the stringent exogeneity assumption of the RE approach by allowing an arbitrary correlation between the unobserved effect or household-specific heterogeneity (σ_i) and the explanatory variables. CRE also avoids the incidental parameters problems associated with fixed effects in models with limited dependent variables (Wooldridge, 2019). As highlighted in Wooldridge (2010), the CRE can be applied to commonly used models, such as unobserved effects probit, Tobit,

[3] Katengeza et al. (2019b) uses the state-contingent theory to explain decision-situations and decisions in such recursive models and how risk and risk perceptions influence decisions.

[4] To come up with the household asset wealth index, we combine information on household ownership of durable non-land assets (e.g., agricultural equipment and machinery) captured in Ethiopia LSMS-ISA data to create the household asset wealth index, using Principal Components Analysis(PCA) (Filmer & Pritchett, 2001).

and count models (Wooldridge 2010, 2019). Average Partial Effects (APEs) are presented to help interpret the economic and not just the statistical significance of variables.

This study, therefore, models the binary use decisions (i.e., use of local seed variety and improved seed varieties for all crops and specifically for maize and wheat), using a CRE logit estimator and report odds ratios. The decision on seed use intensity (amount of local or improved seed used per household) is modeled using a CRE Tobit estimator to account for those who do not use the seed variety. We run separate regressions for the two crop diversification indices. For crop count (richness), CRE Poisson regression is used,[5] while for the Simpson index, a CRE Tobit estimator is used to account for left censoring on the index.[6] In running our model specifications, we first estimate simple models where we control only for the test variables of interest (S_{it}, C_{vt}), and then secondly, we add additional controls as a robustness check.

4.4. Heterogeneity analysis

There is heterogeneity in Ethiopia in agro-ecological conditions and cropping patterns (Beyene, Gibbon, & Haile, 2006). More so, farming households are diverse in resource endowments (land labor and capital) and access to markets, government support, and other institutional services. We, therefore, assess the conditioned impacts of shocks. We do this by using interaction terms of covariate drought shock variables and indicator variables for (i) low agricultural asset endowments (elaborated below), (ii) households with less than average land size holdings (elaborated below), and (iii) households with access to social safety nets. We define low asset wealth (farm size) endowments as dummy variables (1 = yes) for households in the bottom 40 % of the sample asset wealth index (farm size) distribution. We start by defining five quintile categories (1 (=lowest), to 5 (highest)) for each variable (asset wealth and farm size), and then assign one to households with quintile categories 1 and 2, and 0 otherwise. Our indicator variables hence measure relative household asset endowments. The model specification involving interaction terms takes the following form:

$$Q_{it} = \eta_0 + \eta_{1*}(C_{vt} \times Div\_int) + \eta_2 S_{it} + \eta_3 H_{it} + \eta_4 YR_t + \sigma_i + \mu_{it} \quad (4)$$

where ($C_{vt} \times Div\_int$) is the interaction between the covariate shock variable and the indicator variable for relevant household characteristics described prior. We test for interaction effects in all our dependent variables (seed use and intensity, and crop diversification indices). η_{1*} is the vector of coefficients linked to interaction terms between indicators of socioeconomic diversity and covariate shock exposure. We consider covariate drought shocks (lagged drought shock variables) only in the analysis of interaction effects. The interaction effects of shock exposure and household diversity indicators are performed in three separate equations (one for each indicator variable). However, we take cognizant of the fact that access to safety nets is non-random. Hence, our results on the interaction effects of rainfall shock exposure and access to safety nets should be interpreted cautiously. As much as we can control for the unobserved heterogeneity at the household level for a set of time-varying household socioeconomic variables (H_{it}), we cannot fully account for unobserved time-varying characteristics at the household level, which are potentially correlated with the allocation or access to productive safety nets.

4.5. Robustness checks

We explore the robustness of our results by: (i) controlling for additional covariates (in addition to key test variables), (ii) using alternative econometric estimation methods, (iii) using a different weather data source, and (iv) testing and controlling for possible attrition bias.

We run our main results with and without additional controls. To assess the consistency of the primary study outcomes, we also apply the conditional mixed process (CMP) framework proposed by Roodman (2011). The underlying rationale is that we often want to jointly estimate two or more equations with linkages among their error terms. For instance, equations for local and improved seed varietal use could have correlated errors, as farmers can use both local and improved varieties as complementary strategies. Also, seed use decisions and the decision to diversify could have correlated error terms as the use of different crop varieties relates to diversification. The CMP adopted here is based on Zellner (1962) concept of the seemingly unrelated regression (SUR) estimator. Its main advantage is that if there are meaningful correlations between error processes of individual equations for seed use decisions, SUR estimates take account of these correlations and yield more efficient estimates than those derived from single equations. We also estimate seed use and intensity decisions in alternative CRE Craggit Double Hurdle models[7] (Cragg, 1971), and crop count and Simpson indices of diversification using CRE negative binomial, and CRE fractional probit models[8] (Wooldridge, 2010) as robustness analyses.

Besides, we also reproduce our main tables using a different weather data source. It may be possible that properties of weather data may drive results used, such as the selection of weather stations, bias correction methods used, spatial resolutions of data, imputation of missing data, among other factors (Auffhammer, Hsiang, Schlenker, & Sobel, 2013; Letta, Montalbano, & Tol, 2018). For robustness analysis, we use data from NASA's Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al., 2017), to define weather variables and shocks.

Lastly, we also check the robustness of our main results to possible attrition bias in the analyzed panel. First, we estimate an attrition probit model with a dummy dependent variable for households not observed in the follow-up survey 2013/14, using household characteristics at baseline as explanatory variables. Second, we construct an Inverse Mills Ratio (IMR) from the attrition Probit models. Third, following the procedure of the Heckman model, we use the constructed IMR to test and control for the potential attrition bias effect by including it as an additional explanatory variable in our correlated random effect models. Adjusting for attrition bias in all our equations does not alter our main conclusions, showing that our findings are robust to attrition bias.

We present results from the various robustness checks in the Supplementary material (Table B-W).

[5] An alternative to model count data would be the negative binomial model, however, the Poisson model is used because it is robust to both over and under dispersion which is not the case with the negative binomial model. The negative binomial model is robust only with over-dispersion (Gardner, Mulvey, & Shaw, 1995).

[6] About 6% of farmers in the pooled sample had zero (0) values for diversification for the Simpson crop diversification index.

[7] Craggit Double Hurdle (DH) models are used as an alternative modelling framework for seed use decisions (use and intensity). In the DH models the first hurdle involves estimating a probit model that determines the probability that the farming household uses a certain seed type (1=yes; 0= no), while the second hurdle involves estimating a truncated regression model to determine the intensity of seed use.

[8] Given that the Simpson Index (SI) can be read as fractions, defined, and observed only on an interval scale of $0 \leq SI \leq 1$, we can also model the SI index of crop diversification, using a fractional probit estimator (Papke & Wooldridge, 2008). We hence implement the CRE fractional probit models as an alternative to CRE Tobit model of crop diversification using the Simpson index. Fractional regression models such as the fractional probit implement quasi-maximum likelihood estimators to constrain the predicted value between zero and one (Papke & Wooldridge, 2008; Wooldridge, 2011).

5. Potential study limitations

Our study's approach is not without limitations. First, we rely on self-reported data on household cropping activities, including the classification of improved and local varieties. While local varieties are commonly understood as traditional varieties (aka 'landraces'), farmers sometimes refer to locally developed improved varieties as local and sometimes also refer to exotic improved varieties as local after 'recycling' seeds as farm-saved seeds for a few seasons (Westengen, Jeppson, & Guarino, 2013). In fact, even national and international agricultural research organizations classify improved wheat varieties recycled more than five seasons as local (Yirga, Mohammad, Kassie, & Groote, 2013). The point in our study is, however, not to assess the performance of different types of crops but to understand how a diversity of crop varieties are used as coping and adaptation strategies, thus the self-reported categories improved, and local are useful proxies for diversity below the species level. Second, our data allows us to understand the impact of past shock exposure and vulnerability on current farmer actions (*ex-post*), not what they do before exposure (*ex-ante* risk management). However, we believe that studying the impacts of past exposure on current farmer practices can shed light on future exposure to shocks and farmers' responses in coping with them.

The strength of the data used is that it is representative, covering the same households over multiple seasons. Hence, we can understand the responses of farmer's seed use and diversification decisions to shocks in Ethiopia from large data sets, and we can understand the dynamics of the effect as opposed to static effects mainly explored in literature. Hence, we feel that our study gives relevant insights for policy despite noted possible concerns.

5.1. Descriptive statistics

5.1.1. Outcome variables by year

Households mainly rely on local seed varieties in their crop production, as shown in Table 1. Use rates for local seed varieties range from 97 to 99 % of the households over the three seasons. A considerable proportion - about 20 % - of the farmers also use improved seed. From 2011/12 through 2015/16, the use of improved seeds increased modestly from 18 % to 21 %. In terms of the quantity of seed used per household (averaged for all crops), much higher amounts of local seeds are used compared to improved seeds. On average, 16 to 17 kg of improved seeds is used per year per household in the studied period. More than four times as much local seeds were used on average.

Crop diversification is also common, with, on average, rural households growing about 8 different crops in a given year (season). The Simpson index of crop evenness also shows high crop diversification levels in rural Ethiopia, as the average index ranges from 0.71 to 0.73.

Maize is an important cereal for Ethiopians, and 61 to 64 % of our sampled rural households grow maize. In the middle panel of Table 1, descriptive statistics for local and improved varietal use for maize growers are reported. Like for other crops, rural farmers rely mostly on local maize seed. About 90 % of the farmers used local maize seed in 2011/12, but the share decreased slightly to 85 and 84 % in 2013/14 and 2015/16. This is mirrored by an increase in improved maize seeds, from 19 % of the households in 2011/12 to 24 % (2013/14) and 25 % (2015/16). On average, about three times more local maize seeds (58 kg) are used compared to improved seeds (19 kg). Thus, although local seeds dominate, maize cultivation has a relatively higher use of improved seeds compared to other crops.

Wheat is also an important cereal grown in selected high potential areas in Ethiopia. About 26–27 % of the sampled farmers grow wheat (Table 1). Among wheat growers, the use of improved seed has been between 10 and 13 %, with no clear time trend. Over 91 % of the wheat growers relied on local wheat varieties. The heavy reliance on local varieties for wheat growers is also underscored by the much higher average quantities of local wheat seeds (123–147 kgs) than for improved seeds (15–19 kgs). Wheat has less use of improved varieties compared to maize. The ratio of local to improved wheat varietal use (for wheat growers) is about 8, thus twice the average across crops.

5.1.2. Key explanatory variables by year

Table 2 presents descriptive statistics for key explanatory variables selected to explain seed use decisions and diversification. From Table 2, we can see that, on average, the one-year lag for rainfall shock is dominated by flood shocks (positive Z-scores), while the 2-year lag is dominated by drought shocks

Table 1
Descriptive statistics of selected outcome variables used in the analysis.

Variable definitions	2011/12	2013/14	2015/16
	Mean(s.d.)	Mean(s.d.)	Mean(s.d.)
All crops (N = 2398)			
Improved seed use (1 = yes; 0 = otherwise)	0.18(0.39)	0.21(0.41)	0.21(0.41)
Quantity of improved seeds used per household	16.13(67.30)	16.40(54.68)	17.65(68.39)
Local seed use (1 = yes; 0 = otherwise)	0.97 (0.16)	0.98(0.15)	0.99(0.12)
Quantity of local seeds used per household	60.78(107.78)	78.59(127.01)	69.96(111.36)
Grow maize (1 = yes; 0 = otherwise)	0.64(0.48)	0.61(0.49)	0.62(0.49)
Grow wheat (1 = yes; 0 = otherwise)	0.26(0.44)	0.27(0.45)	0.27(0.44)
Number of crops grown per household	8.77(4.77)	8.66(4.66)	8.48(4.63)
Simpson index of crop diversity	0.73(0.21)	0.72(0.21)	0.71(0.22)
Maize growers (N = 1539)			
Improved maize seed (1 = yes; 0 = otherwise)	0.19(0.39)	0.24(0.43)	0.25(0.43)
Quantity of improved maize seed used per household	17.69(74.08)	18.34(55.89)	21.04(73.20)
Local maize seed use (1 = yes; 0 = otherwise)	0.90(0.30)	0.85(0.36)	0.84(0.37)
Quantity of local maize seeds used per household	54.62(96.18)	61.47(106.55)	59.07(102.14)
Wheat growers (N = 628)			
Improved wheat seed (1 = yes; 0 = otherwise)	0.12(0.33)	0.13(0.34)	0.10(0.29)
Quantity of improved wheat seed used per household	18.79(82.86)	14.60(53.10)	19.30(90.37)
Local wheat seed use (1 = yes; 0 = otherwise)	0.90(0.29)	0.91(0.29)	0.94(0.23)
Quantity of local wheat seeds used per household	123.39(151.03)	147.24(177.15)	149.61(150.84)

Notes: Summary statistics are not weighted, standard deviations (s.d) in parentheses.

Table 2
Descriptive statistics of explanatory variables(test variables) used in the analysis.

Variables	2011/12	2013/14	2015/16
	mean	mean	mean
Rainfall shock 1-year lag (Z-score)	0.053	0.183	0.027
Rainfall shock 2-year lag (Z-score)	−0.450	−0.028	0.360
Temperature shock 1-year lag (Z-score)	0.416	0.117	0.826
Temperature shock 2-year lag (Z-score)	2.177	0.594	0.797
Livestock loss in the previous year (1 = yes) †	0.067	0.038	0.071
Formal employment loss (off-farm) by a household member last year(1 = yes) †	0.007	0.006	0.008
Observations	2398	2398	2398

Notes: summary statistics are not weighted, † denotes dummy variable: Shock variables shown in the table are for the period May-July (early season of the Meher season).

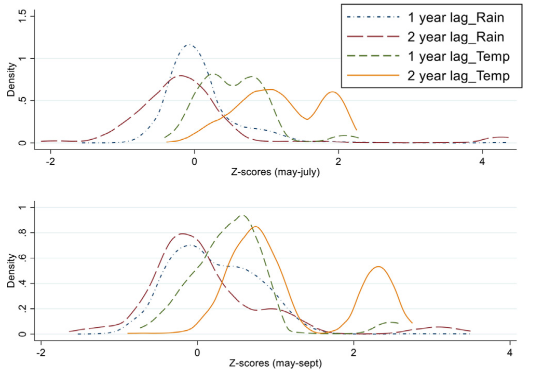

Figure 1. Distribution of rainfall and temperature shocks for the Meher season (may-sept) and early season (may–July) of the Meher in the pooled sample.

(negative z-scores). In terms of drought shocks, we also see that, on average, the drought shock is severe for the two-year lag compared to the one year-lag (Figure 2). For temperature shocks, positive Z-scores dominate for both 1 and 2-year lags, suggesting an overall increase in temperatures. We show the distribution of rainfall and temperature shocks (Z-scores) in Figure 1.

On average, 6 % of households in the pooled sample lost some of their livestock due to death or theft (1-year lag); about 6.7, 3.8, and 7.1 % of farmers lost their livestock in 2011/12, 2013/4, and 2015/16. Also, losing formal employment (1-year lag) was only experienced by about 1 % of respondents in the pooled sample and all survey years (Table 2).

Descriptive statistics for other rainfall and temperature measures considered, including long-run mean rainfall and long-run mean temperature, are shown together with other control variables considered in the Supplementary material.

6. Results

This section presents the main findings from our regression analyses. In Tables 3-6 we report results from seed use equations while in tables 7-8 we report results from crop diversification equations. The results presented are estimated within the Correlated Random Effects (CRE) framework. We first present naïve regression results (where we include only the key treatment variables of interest (C_{vt}, S_{it}) and year dummies (YR_t) but without control variables (H_{it})). Second, we show results where we control for additional controls (and their means across years). For brevity we only report coefficients of our key test variables.

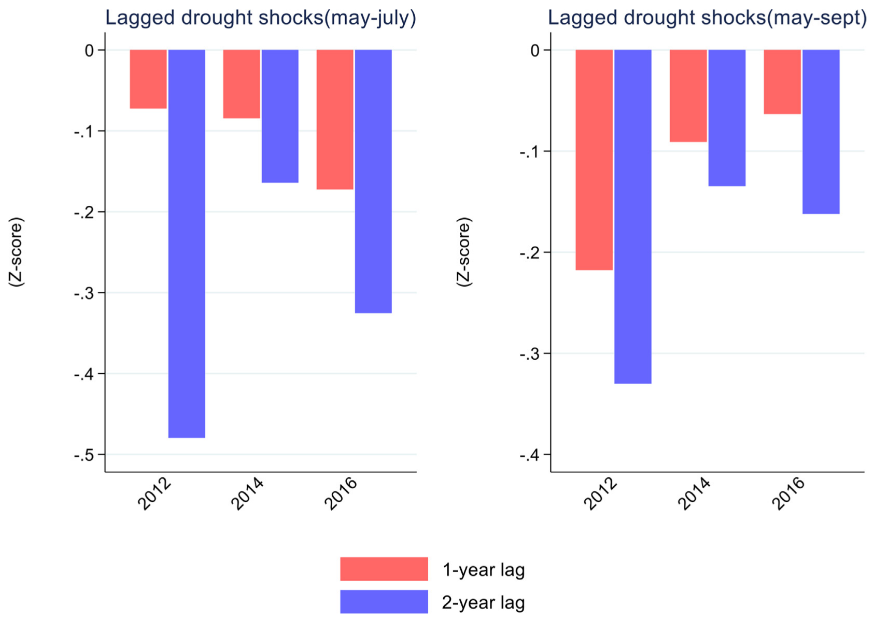

Figure 2. Drought shocks (rainfall shortage) for the Meher season and early season (May-July) of the Meher season in the pooled sample.2012,2014, and 2016 represents 2011/12, 2013/14, and 2015/16 survey rounds of Ethiopia Socioeconomic Survey (ESS).

Table 3
Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia.

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Rainfall shortage 1-year lag	3.645***	2.233***	0.636	0.316***
	(1.1541)	(0.5564)	(0.3659)	(0.0999)
Rainfall shortage 2-year lag	0.354***	−1.738***	4.965***	−0.054
	(0.0548)	(0.2736)	(2.4025)	(0.0432)
Temperature shock 1-year lag	1.016	0.084	0.391***	−0.213***
	(0.0980)	(0.1717)	(0.0771)	(0.0283)
Temperature shock 2-year lag	1.836***	1.093***	4.652***	0.138***
	(0.2797)	(0.2704)	(0.8771)	(0.0396)
Historical mean temperature (1980–2017)	0.959*	−0.114***	0.927**	−0.091***
	(0.0223)	(0.0434)	(0.0314)	(0.0100)
Historical mean rainfall (1980–2017)	1.003***	0.005***	1.000	0.001***
	(0.0002)	(0.0004)	(0.0004)	(0.0001)
Livestock loss†	0.904	−0.061	1.919	0.021
	(0.1804)	(0.3487)	(0.9172)	(0.0597)
Job loss (off-farm)†	0.998	0.101	0.587	0.036
	(0.4923)	(0.8609)	(0.5247)	(0.1639)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				
Rainfall shortage 1-year lag	2.740***	1.723***	0.448	0.265***
	(0.8786)	(0.5491)	(0.2594)	(0.0929)
Rainfall shortage 2-year lag	0.418***	−1.414***	3.790***	0.014
	(0.0659)	(0.2722)	(1.8145)	(0.0414)
Temperature shock 1-year lag	1.097	0.221	0.471***	−0.196***
	(0.1064)	(0.1701)	(0.0928)	(0.0269)
Temperature shock 2-year lag	1.662***	0.858***	4.432***	0.104***
	(0.2521)	(0.2643)	(0.8464)	(0.0374)
Historical mean temperature (1980–2017)	0.959*	−0.124***	0.937	−0.104***
	(0.0239)	(0.0449)	(0.0381)	(0.0078)
Historical mean rainfall (1980–2017)	1.003***	0.005***	1.000	0.0001**
	(0.0003)	(0.0005)	(0.0004)	(0.0001)
Livestock loss†	0.962	0.041	1.324	−0.037
	(0.1928)	(0.3428)	(0.6348)	(0.0557)
Job loss (off-farm)†	0.875	−0.063	0.972	0.072
	(0.4314)	(0.8468)	(0.9032)	(0.1532)
Year dummies	Yes	Yes	Yes	Yes
Other controls + their means across years	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors at EA level in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT = intensity. APE = Average partial effects, OR = odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity of use (INT) equations using Correlated Random Effects logit and Tobit, respectively. †denotes a dummy variable.

6.1. Impact of shocks on seed use decisions

6.1.1. Full sample

.This paper focuses on how shocks affect access and the use of local and improved seeds and crop diversification. Table 3 and Table 4 show that the one-year lag of drought shock increases the use of improved seeds in general and improved wheat in particular. However, the two-year lag of drought shock exposure is shown to have a contrasting effect, as it reduces the chances and intensity of improved seeds and improved maize use while increasing the chances and intensity of using local seeds and local maize. The contrasting effects of the two lags of drought shock seem to suggest that it is the most recent drought shocks (1-year lag) that trigger adaptive behavioral responses by farmers in their seed use decisions and that the relatively more intense and distant drought shocks (e.g., the 2-

Table 4
Impact of shocks on maize and wheat seed use decisions in rural Ethiopia.

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Rainfall shortage 1-year lag	1.428	0.336	0.546	0.063	7.369***	7.002***	0.343	−0.625**
	(0.6650)	(0.6111)	(0.2568)	(0.1701)	(4.4500)	(2.1868)	(0.2497)	(0.2974)
Rainfall shortage 2-year lag	0.193***	−2.206***	7.104***	0.435***	1.011	−0.139	0.928	−0.289
	(0.0451)	(0.3083)	(1.9030)	(0.0733)	(0.3877)	(1.3331)	(0.4103)	(0.1829)
Temperature shock 1-year lag	1.200	0.352*	1.211	0.029	1.276	0.846	0.878	−0.020
	(0.1924)	(0.2120)	(0.2112)	(0.0552)	(0.2646)	(0.7315)	(0.2159)	(0.0964)
Temperature shock 2-year lag	1.379	0.536*	1.019	0.053	2.794**	3.541***	0.573	−0.368***
	(0.3163)	(0.3147)	(0.2283)	(0.0747)	(1.1492)	(1.3541)	(0.2254)	(0.1279)
Historical mean temperature (1980–2017)	0.781***	−0.380***	1.188***	−0.015	1.086*	0.289*	0.887**	−0.072***
	(0.0332)	(0.0589)	(0.0471)	(0.0159)	(0.0473)	(0.1531)	(0.0453)	(0.0227)
Historical mean rainfall (1980–2017)	1.004***	0.006***	0.998***	−0.000	1.000	−0.001	1.000	0.000
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0004)	(0.0002)
Livestock loss†	1.191	0.378	1.255	0.240**	0.664	−1.246	1.018	−0.097
	(0.3374)	(0.3704)	(0.4006)	(0.1093)	(0.2828)	(1.4527)	(0.4637)	(0.1828)
Job loss(off-farm)†	0.951	−0.075	2.040	0.363	1.167	0.177	0.573	−0.373
	(0.6442)	(0.8715)	(1.5992)	(0.2946)	(1.3884)	(4.2408)	(0.6956)	(0.5576)
Models with additional controls								
Rainfall shortage 1-year lag	0.987	−0.078	0.640	−0.010	5.603**	5.510**	0.911	−0.200
	(0.4659)	(0.6032)	(0.3057)	(0.1637)	(3.7960)	(2.2085)	(0.7994)	(0.2737)
Rainfall shortage 2-year lag	0.226***	−1.935***	6.049***	0.433***	1.270	0.920	0.845	−0.055
	(0.0540)	(0.3047)	(1.6576)	(0.0723)	(0.5297)	(1.3455)	(0.4168)	(0.1733)
Temperature shock 1-year lag	1.227	0.337	1.157	0.035	1.312	0.893	0.808	−0.125
	(0.1961)	(0.2071)	(0.2004)	(0.0539)	(0.2933)	(0.7263)	(0.2177)	(0.0916)
Temperature shock 2-year lag	1.195	0.289	1.034	−0.002	2.799**	3.042**	0.634	−0.201*
	(0.2718)	(0.3044)	(0.2311)	(0.0730)	(1.2516)	(1.3378)	(0.2759)	(0.1210)
Historical mean temperature (1980–2017)	0.770***	−0.407***	1.224***	0.002	1.044	0.104	0.904*	−0.100***
	(0.0360)	(0.0626)	(0.0526)	(0.0158)	(0.0515)	(0.1578)	(0.0541)	(0.0201)
Historical mean rainfall (1980–2017)	1.004***	0.005***	0.998***	−0.000***	1.000	−0.002	1.000	−0.001***
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0005)	(0.0002)
Livestock loss†	1.318	0.479	1.101	0.146	0.760	−0.677	0.902	−0.084
	(0.3770)	(0.3639)	(0.3529)	(0.1050)	(0.3495)	(1.4601)	(0.4557)	(0.1690)
Job loss†	0.846	−0.142	3.020	0.536*	1.307	0.666	0.653	−0.129
	(0.5792)	(0.8580)	(2.4233)	(0.2831)	(1.5805)	(4.0002)	(0.8213)	(0.5173)
Other controls + their means across years	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors at EA level; * p < 0.10, ** p < 0.05, *** p < 0.01; INT = intensity. APE = Average partial effects, OR = odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable.

year lag) limit use of improved seeds while increasing use of local seed varieties.

Also, we found that the probability and intensity of using improved seeds increase with historical mean rainfall. In contrast, the chances of using local maize and wheat varieties decrease with historical mean rainfall (Table 3 and Table 4). These results could point to the fact that areas with higher rainfall historically (agronomically favorable areas) have higher use of improved seeds while more marginal areas have less.

Furthermore, we see that temperature shocks (both one and two-year lags) are positively associated with improved seed use (including improved wheat and maize). Lagged temperature shock variables do not show consistent effects on local seed use decisions. Additionally, the historical mean temperature is negatively associated with the intensity of both improved and local variety use (Table 3 and Table 4).

Additionally, Table 4 shows the loss of a formal job by a household member to enhance the chances and intensity of using local maize varieties. We also observe that in most of our models, our

Table 5
Interaction effects of drought shocks and household diversity indicators on household seed use decisions (all crop model) in rural Ethiopia.

	Improved seed		Local seed	
	Use	INT	Use	INT
Rainfall shortage (interactions)	(OR)	(APE)	(OR)	(APE)
Rainfall shortage 1-year lag × Rainfall shortage 2-year lag	0.254***	−2.226***	0.187***	0.143**
	(0.0682)	(0.4495)	(0.0903)	(0.0652)
All other baseline controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194
Small farm size				
Rainfall shortage 1-year lag × LFS	2.325	1.271	0.205**	0.283**
	(1.2013)	(0.9405)	(0.1366)	(0.1442)
Rainfall shortage 2-year lag × LFS	0.592**	−0.822**	3.291**	0.026
	(0.1279)	(0.3832)	(1.7196)	(0.0557)
All other baseline controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194
Asset poor households				
Rainfall shortage 1-year lag × poor	1.241	0.314	0.739	0.344***
	(0.5489)	(0.7733)	(0.5687)	(0.1236)
Rainfall shortage 2-year lag × poor	0.496***	−1.204***	2.738*	−0.110**
	(0.0949)	(0.3322)	(1.5747)	(0.0491)
All other baseline controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194
Received Social Safety Nets				
Rainfall shortage 1-year lag × SSN	8.602***	4.070***	0.421	0.368*
	(5.8657)	(1.1738)	(0.5232)	(0.1941)
Rainfall shortage 2-year lag × SSN	0.520	−0.958	2.408	0.230**
	(0.2132)	(0.7073)	(2.5710)	(0.1020)
All other baseline controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors at EA level in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT = intensity. APE = Average partial effects, OR = odds ratios. LFS = indicator variable for Low farm size; SSN = Indicator variable for having received Social Safety Nets; Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

test variables' crude effects (effects without additional controls) are comparable to the adjusted effects (effect after controlling for additional controls), indicating that our results are robust to adding additional controls.

6.1.2. Heterogeneity analysis

We perform heterogeneity analysis using interaction terms to understand the conditioned effects of lagged drought shock exposure. We start by interacting one and two-year lags of drought shocks and assess the influence of recurrent drought exposure on our dependent variables. We then perform interaction effects analysis of indicator variables for household socioeconomic diversity (small farm size, low agricultural asset endowment, and access to social safety nets) with lagged drought shocks and report results in Tables 5-6. We only show an extract of results from the interaction effects analysis for brevity. Results show that recurrent drought shock exposure discourages the use of improved varieties while enhancing the use of local seed varieties (Table 5 and Table 6).

The results also show that drought shock exposure for households with less than average farm sizes significantly reduces their chances of using improved seeds and improved maize and increases their chances of using local seed varieties in general and specifically for maize and wheat.

Further, results show that drought shock exposure for smallholder farming households in the low-agricultural asset category significantly increases their chances and intensity of local seed variety use in general (Table 5) and particularly for local maize (Table 6) and reduces chances of using improved seed (in general) and improved maize.

Also, interacting lagged drought shock exposure with the reception of Social Safety Nets (SSN) reveals that access to SSN with exposure enhances the use and intensity of improved seeds and improved maize and the intensity of local seed (in general) and local wheat. However, when interacting with a more severe and distant drought shock (2-year lag), social safety nets significantly reduce the use and intensity of improved varieties.

6.2. Impact of shocks on crop diversification decisions

6.2.1. Full sample

The impact of shocks on crop diversification decisions is shown in Table 7. We report results on the two indices considered: crop count (richness) and the Simpson index (evenness). We show both crude (without additional controls) and adjusted (with additional controls) effects of climate variables and shocks on crop diversification decisions in Table 7.

Table 7 shows that historical mean rainfall is positively associated with crop diversity, both in terms of species richness and evenness while historical mean temperatures have the opposite effect on crop diversity. However, drought shocks are associated with a reduction in the number of crops grown (Table 7), and flood shocks experienced in the recent past encourage crop

Table 6
Interaction effects of shocks and household diversity indicators on maize and wheat seed use decisions in rural Ethiopia.

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Rainfall shortage (interactions)								
Rainfall shortage 1-year lag × Rainfall shortage 2-year lag	0.286***	−1.521***	1.416	0.314***	0.396**	−2.644*	2.091	0.416**
	(0.1193)	(0.5224)	(0.5709)	(0.1149)	(0.1697)	(1.3715)	(1.0290)	(0.1731)
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	4479	4479	4479	1937	1937	1937	1937
Low farm size								
Rainfall shortage 1-year lag × LFS	2.355	0.364	0.349	−0.364	4.061	4.909	0.295	−0.299
	(1.8819)	(1.0917)	(0.2668)	(0.2697)	(4.3822)	(3.6207)	(0.3489)	(0.5147)
Rainfall shortage 2-year lag × LFS	0.218***	−1.932***	5.873***	0.364***	0.925	−0.021	2.645*	0.291
	(0.0774)	(0.4539)	(2.2895)	(0.1030)	(0.4369)	(1.5419)	(1.5246)	(0.2024)
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	4479	4479	4479	1937	1937	1937	1937
Asset poor households								
Rainfall shortage 1-year lag × poor	0.695	−0.367	0.999	0.165	1.486	1.133	1.135	−0.043
	(0.4352)	(0.8265)	(0.6498)	(0.2202)	(1.3811)	(2.9627)	(1.3198)	(0.3451)
Rainfall shortage 2-year lag × poor	0.433***	−1.167***	3.351***	0.197**	0.688	−0.888	1.435	0.231
	(0.1180)	(0.3521)	(1.0985)	(0.0846)	(0.2913)	(1.3495)	(0.7106)	(0.1673)
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	4479	4479	4479	1937	1937	1937	1937
Received Social Safety Nets								
Rainfall shortage 1-year lag × SSN	7.395*	3.562**	0.054***	−0.597*	5.767	5.152	0.836	−0.051
	(8.4166)	(1.5184)	(0.0560)	(0.3592)	(6.4709)	(3.7186)	(1.2045)	(0.5291)
Rainfall shortage 2-year lag × SSN	0.283*	−1.554*	2.377	0.479**	0.433	−2.443	2.696	0.700***
	(0.2082)	(0.9411)	(1.6449)	(0.1941)	(0.2922)	(2.1521)	(2.2696)	(0.2544)
All other baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors at EA level; * p < 0.10, ** p < 0.05, *** p < 0.01; INT = intensity. APE = Average partial effects, OR = odds ratios. LFS = indicator variable for Low farm size, SSN = Indicator variable for having received Social Safety Nets, Improved and local varieties for both maize and wheat is first measured as dummy variables for use and secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively. †denoted a dummy variable.

diversification (Supplementary material Table J). Besides, historical mean temperature discourages crop diversification.

Also, we found that crop diversification is positively associated with temperature shocks. Further, livestock loss within the household is negatively associated with crop diversification, while job loss within the household is positively associated with the number of crops grown (Table 7).

Results also show that controlling for additional variables does not significantly alter the interpretation of results on the impact of covariate and idiosyncratic shocks on crop diversification. Our results are hence robust to the addition of additional controls.

6.2.2. Heterogeneity analysis

We also assessed the impacts of recurrent drought shock exposure and how heterogeneity in household socioeconomic conditions influences the impact of shocks on diversification. The results (Table 8) show recurrent drought shock exposure to reduce the number of crops grown. Also, interacting drought exposure on households with low farm size discourages crop diversification. Besides interacting, lagged drought shocks with an indicator variable for asset-poor households negatively associates with the number of crops grown by the household.

Further, we see that drought shock exposure on households who accessed social safety nets promotes crop diversification (Table 8).

7. Discussions

Our study evaluated (i) whether past exposure to covariate shocks and idiosyncratic shocks significantly influence seed variety use and diversification of cropping portfolios, and (ii) whether effects of shocks are heterogeneous by households' land size holding inequality, agricultural asset endowment inequality, and access to social safety nets. We discuss key findings for each of these two research questions below.

7.1. Impact of shocks on seed use decisions and crop diversification

Several findings emerged from our analyses. First, drought shock exposure increases the likelihood of farmers using improved seeds, in particular improved wheat, and reduces the likelihood of local wheat seed use. We learned that the most recent drought shocks (1-year lag) are more influential on crop seed use compared to more distant drought shocks (2-year lag). Relatively more severe and long-term drought shocks (2-year lag), and recurrent drought shock exposure (1-year lag*2-year lag) reduce the likelihood and intensity of using improved seeds while enhancing the likelihood and intensity of using local seeds. Also, lagged temperature shocks enhance the likelihood and intensity of using improved seeds in general and for wheat and maize. Second, recurrent drought shock exposure discourage crop diversification, while flood shocks and temperature shocks promote crop diversification. Third,

Table 7
Impact of shocks on crop diversification decisions in rural Ethiopia.

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Rainfall shortage 1-year lag	−0.033	0.005	−0.061*	−0.001
	(0.0321)	(0.0136)	(0.0321)	(0.0132)
Rainfall shortage 2-year lag	0.038***	0.006	0.012	−0.005
	(0.0144)	(0.0059)	(0.0147)	(0.0059)
Temperature shock 1-year lag	−0.022**	−0.005	−0.006	−0.004
	(0.0091)	(0.0038)	(0.0091)	(0.0037)
Temperature shock 2-year lag	−0.006	0.018***	−0.021	0.016***
	(0.0138)	(0.0054)	(0.0138)	(0.0053)
Historical mean temperature (1980–2017)	−0.035***	−0.017***	−0.018***	−0.010***
	(0.0047)	(0.0012)	(0.0044)	(0.0012)
Historical mean rainfall (1980–2017)	0.0005***	0.0002***	0.0004***	0.0002***
	(0.000047)	(0.000014)	(0.000045)	(0.000013)
Livestock loss†	0.009	−0.021**	0.001	−0.021**
	(0.0191)	(0.0082)	(0.0191)	(0.0080)
Job loss(off-farm)†	0.093*	0.002	0.078	−0.005
	(0.0529)	(0.0222)	(0.0527)	(0.0216)
Year dummies	Yes	Yes	Yes	Yes
Other controls + their means across years	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors at EA level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; APE = Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

idiosyncratic household shocks have a less significant role in explaining seed use and crop diversification decisions when compared to covariate rainfall (or temperature) shocks.

The finding that improved seed use is positively associated with experiencing drought shock in the previous season can be explained by both push and pull factors. First, given that smallholder farmers in Ethiopia rely mainly on informal seed sources, including farm-saved seeds, farmer to farmer seed exchange, and local markets (Thijssen et al., 2008), prior exposure to a bad season probably reduces seed supply from the informal sources. Hence, past exposure to a bad season pushes farmers to move from their default position (use of local seeds) towards using off-farm sourced improved seeds. In the longer term, exposure to drought in a rural economy dependent on rain-fed agriculture is likely to intensify poverty (Dercon, 2004) and this may explain why we found 2-year lag drought shocks to reduce crop diversification, and recurrent drought shock exposure (drought shock 1 and 2-year lag interactions) to reduce improved seed use and enhance local seed use. Liquidity constraints following a bad season (or even worse: recurrent bad seasons) can be severe amongst smallholder farmers. Hence, acquiring improved seed from the formal seed system and implementing a diversified crop portfolio become more expensive and outside their reach if the households experience intensified poverty.

Second, the choice of improved varieties at the expense of local varieties after exposure to drought shocks could also reflect the pull factor of learning. Farmers may have learned from their past experiences that improved varieties perform better under low rainfall conditions, and this might also explain the increased use of improved varieties at the expense of local varieties when faced with drought shocks. In such a case, farmers may increase the use of improved varieties as a form of insurance to future anticipated shocks. For instance, Katengeza, Holden, & Lunduka (2019) found that past exposure to drought shocks improves the probability of using improved drought-tolerant maize varieties in Malawi. Based on the same data, Holden and Quiggin (2017) found that more risk-averse farmers were more likely to adopt such varieties as well as local maize at the expense of other improved varieties. In addition, past shock exposure enhanced the use of drought-tolerant maize and discouraged the use of local maize. Preferences may therefore interact with learning through exposure to shocks in the adaptation process. This idea is further supported by literature that alludes to the fact that rural households switch from their business-as-usual practices to practices that increase their mutual insurance to shocks to better cope with shocks (Angelsen & Dokken, 2018; Takasaki, 2011).

The positive effects of lagged flood shocks on crop diversification show more opportunities than constraints associated with abundant rainfall. Abundant rainfall in the previous year may translate into good harvests, which could relax liquidity constraints and lead to higher farming activity and crop diversification.

Furthermore, the finding that smallholder farmers' seed use and crop diversification decisions consistently respond to most recent shocks compared to long-term shocks likely reflect that smallholder farmers are more likely to build their weather expectations for the coming seasons based on their most recent weather shock experiences. Our results here are in line with previous studies (e.g., Katengeza, Holden, and Fisher (2019)) that found more recent weather shocks to be more influential in shaping farmers' weather expectations compared to more distant, long-term weather shocks. However, there could be more competing explanations for the contrasting effects of shocks (immediate vs distant shocks), leaving room for future research to explore the mechanisms that could lead to differential responses to immediate and distant weather shocks.

Idiosyncratic shocks minimally explain seed use and crop diversification decisions when compared to covariate rainfall shocks.

Table 8
Interaction effects of shocks and household diversity indicators on crop diversification decisions in rural Ethiopia.

	Crop diversification indices	
	Crop Count	Simpson Index
Rainfall shortage (interactions)		
Rainfall shortage 1-year lag × Rainfall shortage 2-year lag	−0.080***	0.008
	(0.0250)	(0.0093)
All other baseline controls	Yes	Yes
Observations	7194	7194
Small farm size		
Rainfall shortage 1-year lag × LFS	−0.277***	−0.048**
	(0.0574)	(0.0205)
Rainfall shortage 2-year lag × LFS	−0.012	0.013
	(0.0208)	(0.0079)
All other baseline controls	Yes	Yes
Observations	7194	7194
Asset poor households		
Rainfall shortage 1-year lag × poor	−0.170***	−0.022
	(0.0431)	(0.0177)
Rainfall shortage 2-year lag × poor	−0.024	−0.002
	(0.0175)	(0.0070)
All other baseline controls	Yes	Yes
Observations	7194	7194
Received Social Safety Nets		
Rainfall shortage 1-year lag × SSN	0.358***	0.048*
	(0.0742)	(0.0278)
Rainfall shortage 2-year lag × SSN	0.013	0.021
	(0.0375)	(0.0145)
All other baseline controls	Yes	Yes
Observations	7194	7194

In parenthesis are cluster robust standard errors at EA level; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; APE = Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Losing livestock assets and formal employment within the household minimally explains seed use and diversification decisions. Losing livestock assets and income from formal work reduces household income and asset endowments, further hurting farming investments. Livestock is an essential source of wealth and manure to fertilize the soil and draft power to cultivate the Land for farming households (Thornton & Herrero, 2015). Hence, losing livestock reduces the availability of crucial inputs, which may minimize crop diversification on the farm. However, smallholder farming households usually find it easier to cope with idiosyncratic household shocks than to covariate weather shocks (Dercon, 2005; Nguyen et al., 2020), which can explain why idiosyncratic shocks were less important in explaining seed use and diversification decisions.

7.2. Conditioned effects of drought shocks

Heterogeneity analyses show that drought exposure among farmers with small farm sizes and low agricultural assets reduces reliance on improved seeds, increases reliance on local varieties (in general and for maize and wheat), and reduces crop diversification. On the other hand, access to productive safety nets enhances the likelihood that the farmers will use improved seeds and diversify their crop portfolio following a drought.

Uninsured climate shocks usually lead to fluctuations in household welfare, and may lead to transient (temporary) poverty. This might, however, be avoided if effective safety nets are in place (Dercon, 2005). We found access to social safety nets to significantly alter the effects of drought shock exposure on seed use and diversification decisions. Access to social safety nets enables households to maintain agricultural activity and crop diversification and improve their use of improved seeds. However, relatively more intense drought shocks reduce improved seed use and crop diversification and enhance local seed use. Access to productive safety nets does not significantly alter this relationship for relatively more intense drought shocks. However, given that we cannot fully account for unobserved time-varying characteristics at the household level, which are potentially correlated with access to productive safety nets our results on the interaction effects of rainfall shock exposure and access to productive safety nets must be regarded as correlations and not implying any causal relations.

The adverse effects of recurrent drought shocks on cropping decisions could reflect the effects of temporal poverty induced by drought shock exposure. The effects of drought shock exposure are worse among poorer households (Deressa, Hassan, & Ringler, 2008). This notion possibly explains why farmers with less than average farm sizes and households lowly endowed with agricultural assets were found to intensify on local seed use and reduce diversification post-exposure to drought shocks.

Further, the use of improved varieties may appear a risky venture for the farmer when weather conditions are uncertain. This point to shock exposure also having behavioral impacts, where households faced with risks and with limited insurance substitutes are pushed towards risk management strategies that include low-risk activities (e.g., use of local seeds) but also with lower returns (Dercon 2005, 2002, 2004). This finding is in line with earlier studies: poor households who face production shocks become less likely to engage in beneficial activities that are considered risky (Dercon & Christiaensen, 2011; Gebremariam & Tesfaye, 2018).

8. Conclusions

Crop diversification and varietal change are important strategies for buffering production risk in smallholder agriculture and, hence, rural development (Asfaw et al., 2019; Bozzola & Smale, 2020; Di Falco et al., 2010; Katengeza & Holden, 2021; Tesfaye & Tirivayi, 2020). This study gives important insights into the drivers and constraints involved in farmers decision to use different types of seeds and to diversify their crop portfolio. The bulk of the seeds used by Ethiopian farmers are local. However, improved seed use for all crops and specifically for maize shows a slightly increasing trend over the study period. Furthermore, cropping portfolios at the household level are highly diversified in rural Ethiopia.

We found that more rainfall is associated with more use of improved seeds as well as higher crop diversity. Exposure to drought shocks increases households' use of improved seeds in general and specifically for wheat. However, one and two-year lags of drought shocks have heterogenous effects on seed use and diversification decisions. The most recent weather shocks (one-year lags) appear more influential than more distant weather shocks (two-year lags) in shaping farmers' weather expectations which influence seed use and diversification. Recurrent and severe drought shocks significantly reduce agricultural activity, including improved seed use and crop diversification. Besides, loss of livestock within the household reduces resources available and hence prospects to diversify. Also, losing off-farm work by a household member enhances reliance on local maize seed. Overall, shock exposure poses heterogeneous impacts on seed use and crop diversification in rural Ethiopia. Low-income households and those with less than average farm size significantly intensify local seed use and reduce reliance on improved seed use following covariate shocks. The implication is that socioeconomic disadvantages (e.g., poor asset endowments) and drought shocks make households more seed insecure.

The negative interaction between poverty and shock exposure for the crop-based adaptation activities is significantly lowered when households have access to social security nets. However, in Ethiopia, the productive safety net program only reaches 8 % of the Ethiopian population (Berhane, Gilligan, Hoddinott, Kumar, & Taffesse, 2014; Duru, 2016). In such context, one would expect that farmers have incentives to diversify crop and variety choices as a strategy to buffer risks. And, previous studies have found that crop diversification and improved seed use indeed directly enhance food security under and after shocks. However, our results indicate that recurrent exposure to adverse shocks reduces farming returns and intensifies liquidity constraints, and possibly enhances poverty, hindering farmers from effectively implementing adaptation actions to such shocks, thus hindering the realization of the positive welfare outcomes highlighted in the previous literature.

To avoid negative climate responses for agricultural development by smallholder farmers, up-scaling sustainable and affordable insurance and effective social safety nets is needed. Moreover, given the significance of both improved and local seeds in the face of shocks, farmer's seed systems must co-exist and work in harmony with efforts to increase access to both improved and local varieties. For the less land and asset endowed and those without access to the public social security program, local seeds are an essential part of their de facto safety nets. The informal seed systems supplying local seeds must thus at a minimum be allowed the legal space to exist, but they should also be considered an important entry point for supporting farmers' seed security through such measures as farmer-group seed production and decentralized seed quality control. Hence, our results lend support to Ethiopia's pluralistic seed system development strategy (which recognizes formal, informal, and intermediate seed systems) as an institutional approach to enhance farmers' adaptative capacity (MoA and ATA 2017; Mulesa, Dalle, Makate, Haug, & Westengen, 2021). If well implemented, the pluralistic seed system development strategy can improve farmers' chances to access sufficient quality seed of preferred crops and varieties both in normal seasons and post-shock exposure.

Crop diversity at both species and varietal level is key for adapting to the effects of climate change and other risks faced by smallholder farmers in SSA. We have shown that Ethiopian rural households indeed respond to shocks by making changes in their crop portfolios in subsequent seasons, but that the nature and intensity of those changes depend on their socioeconomic status and their access to social safety nets. Policy measures aimed at reducing vulnerability through increasing seed security must thus address the seed systems farmers rely on for access to these vital resources as well as social inequalities in seed access.

CRediT authorship contribution statement

Clifton Makate: Conceptualization. **Arild Angelsen:** Conceptualization, Supervision. **Stein Terje Holden:** Conceptualization, Supervision. **Ola Tveitereid Westengen:** Conceptualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research work is an output of a research project, "*Access to seeds: from Emergencies to Seed System development (ACCESS)*" which is fully supported by the Research Council of Norway, project number (RCN 288493). The authors are also grateful to the World Bank for making the Livelihood Standard Measurement survey data for Ethiopia available for this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.worlddev.2022.106054.

References

Abay, F., Waters-Bayer, A., & Bjørnstad, Å. (2008). Farmers' seed management and innovation in varietal selection: Implications for barley breeding in Tigray, northern Ethiopia. *AMBIO: A Journal of the Human Environment, 37*(4), 312–320.

Acevedo, M., Pixley, K., Zinyengere, N., Meng, S., Tufan, H., Cichy, K., et al. (2020). A scoping review of adoption of climate-resilient crops by small-scale producers in low-and middle-income countries. *Nature Plants, 6*(10), 1231–1241.

Aguilar, A., Carranza, E., Goldstein, M., Kilic, T., & Oseni, G. (2015). Decomposition of gender differentials in agricultural productivity in Ethiopia. *Agricultural Economics, 3*(46), 311–334.

Alemu, D., & Bishaw, Z. (2015). Commercial behaviours of smallholder farmers in wheat seed use and its implication for demand assessment in Ethiopia. *Development in Practice, 25*(6), 798–814.

Almekinders, C. J., Louwaars, N. P., & De Bruijn, G. (1994). Local seed systems and their importance for an improved seed supply in developing countries. *Euphytica, 78*(3), 207–216.

Alobo Loison, S. (2015). Rural Livelihood Diversification in Sub-Saharan Africa: A Literature Review. *The Journal of Development Studies, 51*(9), 1125–1138.

Angelsen, A., & Dokken, T. (2018). Climate exposure, vulnerability and environmental reliance: A cross-section analysis of structural and stochastic poverty. *Environment and Development Economics, 23*(3), 257–278.

Asfaw, S., Scognamillo, A., Caprera, G. D., Sitko, N., & Ignaciuk, A. (2019). Heterogeneous impact of livelihood diversification on household welfare: Cross-country evidence from Sub-Saharan Africa. *World Development, 117*, 278–295.

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

Bellon, M. R., & Hellin, J. (2011). Planting hybrids, keeping landraces: Agricultural modernization and tradition among small-scale maize farmers in Chiapas, Mexico. *World Development, 39*(8), 1434–1443.

Berhane, G., Gilligan, D. O., Hoddinott, J., Kumar, N., & Taffesse, A. S. (2014). Can social protection work in Africa? The impact of Ethiopia's productive safety net programme. *Economic Development and Cultural Change, 63*(1), 1–26.

Beyene, A., Gibbon, D., & Haile, M. (2006). Heterogeneity in land resources and diversity in farming practices in Tigray, Ethiopia. *Agricultural Systems, 88*(1), 61–74.

Bezner, K., R.,, T. Hasegawa, R. Lasco, I. Bhatt, D. Deryng, A. Farrell, H. Gurney-Smith, H. Ju, S. L.-C., F. Meza, G. Nelson, et al. (2022). Food, Fibre, and Other Ecosystem Products. . In [H.-O. Pörtner, D. C. R., M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)] (ed.) *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. In Press: Cambridge University Press. .

Bezu, S., & Holden, S. (2014). Are rural youth in Ethiopia abandoning agriculture? *World Development, 64*, 259–272.

Bozzola, M., & Smale, M. (2020). The welfare effects of crop biodiversity as an adaptation to climate shocks in Kenya. *World Development, 135* 105065.

Challinor, A., Watson, J., Lobell, D., Howden, S., Smith, D., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change, 4*(4), 287–291.

Chamberlain, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics, 18*(1), 5–46.

Chamberlain, G. (1984). Chapter 22 Panel data. In vol. 2 *Handbook of Econometrics*, pp. 1247-1318: Elsevier.

Chambers, R. G., & Quiggin, J. (2000). *Uncertainty, production, choice, and agency: The state-contingent approach*. Cambridge University Press.

Coomes, O. T., McGuire, S. J., Garine, E., Caillon, S., Mckey, D., Demeulenaere, E., et al. (2015). Farmer seed networks make a limited contribution to agriculture? Four common misconceptions. *Food Policy, 56*, 41–50.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 829–844.

De Janvry, A., Fafchamps, M., & Sadoulet, E. (1991). Peasant household behaviour with missing markets: Some paradoxes explained. *The Economic Journal, 101* (409), 1400–1417.

Debela, B. L., Shively, G. E., & Holden, S. T. (2021). Implications of food-for-work programs for consumption and production diversity: Evidence from the Tigray Region of Ethiopia. *Agricultural and Food Economics, 9*(1), 28.

Dejene, M., & Cochrane, L. (2021). Safety nets as a means of tackling chronic food insecurity in rural southern Ethiopia: What is constraining programme contributions? *Canadian Journal of Development Studies/Revue canadienne d'études du développement*, 1–19.

Dercon, S. (2002). Income risk, coping strategies, and safety nets. *The World Bank Research Observer, 17*(2), 141–166.

Dercon, S. (2004). Growth and shocks: Evidence from rural Ethiopia. *Journal of Development Economics, 74*(2), 309–329.

Dercon, S. (2005). Risk, poverty and vulnerability in Africa. *Journal of African Economies, 14*(4), 483–488.

Dercon, S., & Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics, 96*(2), 159–173.

Deressa, T., Hassan, R. M. & Ringler, C. (2008). *Measuring Ethiopian farmers' vulnerability to climate change across regional states*: Intl Food Policy Res Inst.

Dessie, A. B., Abate, T. M., Mekie, T. M., & Liyew, Y. M. (2019). Crop diversification analysis on red pepper dominated smallholder farming system: Evidence from northwest Ethiopia. *Ecological Processes, 8*(1), 50.

Di Falco, S., Bezabih, M., & Yesuf, M. (2010). Seeds for livelihood: Crop biodiversity and food production in Ethiopia. *Ecological Economics, 69*(8), 1695–1702.

Ding, Y., Schoengold, K., & Tadesse, T. (2009). The impact of weather extremes on agricultural production methods: Does drought increase adoption of conservation tillage practices? *Journal of Agricultural and Resource Economics*, 395–411.

Duru, M. J. (2016). Too certain to invest? Public safety nets and insurance markets in Ethiopia. *World Development, 78*, 37–51.

Elagib, N. A. (2015). Drought risk during the early growing season in Sahelian Sudan. *Natural Hazards, 79*(3), 1549–1566.

Ellis, F. (2000). *Rural livelihoods and diversity in developing countries*. Oxford University Press.

Erenstein, O., & Kassie, G. T. (2018). Seeding eastern Africa's maize revolution in the post-structural adjustment era: A review and comparative analysis of the formal maize seed sector. *International Food and Agribusiness Management Review, 21*(1), 39–52.

Fafchamps, M. (1992). Cash crop production, food price volatility, and rural market integration in the third world. *American Journal of Agricultural Economics, 74*(1), 90–99.

FAO. (2018). *Reviewing and enhancing the existing seed security conceptual framework*. Available at: http://www.fao.org/in-action/food-security-capacity-building/project-components/seeds/seed-security-conceptual-framework/en/.

Fick, S. E., & 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.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography, 38*(1), 115–132.

Fisher, M., Abate, T., Lunduka, R. W., Asnake, W., Alemayehu, Y., & Madulu, R. B. (2015). Drought tolerant maize for farmer adaptation to drought in sub-Saharan Africa: Determinants of adoption in eastern and southern Africa. *Climatic Change, 133*(2), 283–299.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin, 118*(3), 392.

Gebregziabher, G., & Holden, S. T. (2011). Distress rentals and the land rental market as a safety net: Contract choice evidence from Tigray, Ethiopia. *Agricultural Economics, 42*, 45–60.

Gebremariam, G., & Tesfaye, W. (2018). The heterogeneous effect of shocks on agricultural innovations adoption: Microeconometric evidence from rural Ethiopia. *Food Policy, 74*, 154–161.

Gebru, M., Holden, S. T., & Alfnes, F. (2021). Adoption analysis of agricultural technologies in the semiarid northern Ethiopia: A panel data analysis. *Agricultural and Food Economics, 9*(1), 1–16.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate, 30*(14), 5419–5454.

Holden, S. T., & Quiggin, J. (2017). Climate risk and state-contingent technology adoption: Shocks, drought tolerance and preferences. *European Review of Agricultural Economics, 44*(2), 285–308.

Holden, S., & Shiferaw, B. (2004). Land degradation, drought and food security in a less-favoured area in the Ethiopian highlands: A bio-economic model with market imperfections. *Agricultural Economics, 30*(1), 31–49.

Holden, S. T., & Tilahun, M. (2020). Farm size and gender distribution of land: Evidence from Ethiopian land registry data. *World Development, 130* 104926.

Howden, S. M., Soussana, J.-F., Tubiello, F. N., Chhetri, N., Dunlop, M. & Meinke, H. (2007). Adapting agriculture to climate change. 104 (50): 19691-19696.

IPCC. (2014). Climate Change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group 2 to the fifth assessment report of the intergovernmental panel on climate change. *Intergovernmental Panel on Climate Change (IPCC)*. New York: Cambridge University Press. 1132 pp.

IPCC. (2022). Summary for Policymakers. In H.-O. Pörtner, D.C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S., Langsdorf, S. Löschke, V. Möller, et al. (eds) *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* In Press: Cambridge University Press.

Katengeza, S. P., & Holden, S. T. (2021). Productivity impact of drought tolerant maize varieties under rainfall stress in Malawi: A continuous treatment approach. *Agricultural Economics, 52*(1), 157–171.

Katengeza, S. P., Holden, S. T., & Fisher, M. (2019). Use of Integrated Soil Fertility Management Technologies in Malawi: Impact of Dry Spells Exposure. *Ecological Economics, 156*, 134–152.

Katengeza, S. P., Holden, S. T., & Lunduka, R. W. (2019). Adoption of drought tolerant maize varieties under rainfall stress in Malawi. *Journal of Agricultural Economics, 70*(1), 198–214.

Kurosaki, T., & Fafchamps, M. (2002). Insurance market efficiency and crop choices in Pakistan. *Journal of Development Economics, 67*(2), 419–453.

Letta, M., Montalbano, P., & Tol, R. S. (2018). Temperature shocks, short-term growth and poverty thresholds: Evidence from rural Tanzania. *World Development, 112*, 13–32.

Louwaars, N. P., & de Boef, W. S. (2012). Integrated Seed Sector Development in Africa: A Conceptual Framework for Creating Coherence Between Practices, Programs, and Policies. *Journal of Crop Improvement, 26*(1), 39–59.

Makate, C., Wang, R., Makate, M., & Mango, N. (2016). Crop diversification and livelihoods of smallholder farmers in Zimbabwe: Adaptive management for environmental change. *SpringerPlus, 5*(1), 1–18.

Masarie, K. A., & Tans, P. P. (1995). Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record. *Journal of Geophysical Research: Atmospheres, 100*(D6), 11593–11610.

Masters, W. A., Djurfeldt, A. A., De Haan, C., Hazell, P., Jayne, T., Jirström, M., et al. (2013). Urbanization and farm size in Asia and Africa: Implications for food security and agricultural research. *Global Food Security, 2*(3), 156–165.

McGuire, S., & Sperling, L. (2013). Making seed systems more resilient to stress. *Global Environmental Change, 23*, 644–653.

Mekbib, F. (2007). Farmers' seed system of sorghum (Sorghum bicolor (L.) Moench) in the center of diversity: I. Seed sources, distribution, and networking. *Journal of New Seeds, 8*(3), 63–86.

Michler, J. D., Baylis, K., Arends-Kuenning, M., & Mazvimavi, K. (2019). Conservation agriculture and climate resilience. *Journal of Environmental Economics and Management, 93*, 148–169.

MoA and ATA. (2017). *Seed System Development Strategy: Vision, systematic challenges, and prioritized interventions. Working Strategy Document.* Addis Ababa, Ethiopia: Federal Democratic Republic of Ethiopia, Ministry of Agriculture (MoA) and Agricultural Transformation Agency (ATA).

MoA. (2019). *Draft National Seed Industry Policy (final draft in Amharic). Unpublished.* Addis Ababa, Ethiopia: Ministry of Agriculture (MoA). 24 pp.

Mortimore, M. J., & Adams, W. M. (2001). Farmer adaptation, change and 'crisis' in the Sahel. *Global Environmental Change, 11*(1), 49–57.

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. 104 (50): 19680-19685.

Mulesa, T. H., Dalle, S. P., Makate, C., Haug, R., & Westengen, O. T. (2021). Pluralistic Seed System Development: A Path to Seed Security? *Agronomy, 11*(2), 372.

Mulwa, C. K., & Visser, M. (2020). Farm diversification as an adaptation strategy to climatic shocks and implications for food security in northern Namibia. *World Development, 129* 104906.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69–85.

Nguyen, T.-T., Nguyen, T. T., & Grote, U. (2020). Multiple shocks and households' choice of coping strategies in rural Cambodia. *Ecological Economics, 167* 106442.

Papke, L. E., & Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics, 145*(1–2), 121–133.

Porter, C. (2012). Shocks, consumption and income diversification in rural Ethiopia. *Journal of Development Studies, 48*(9), 1209–1222.

Pradhan, K. C., & Mukherjee, S. (2018). Covariate and idiosyncratic shocks and coping strategies for poor and non-poor rural households in India. *Journal of Quantitative Economics, 16*(1), 101–127.

Rashid, S. & Minot, N. (2010). *Are staple food markets in Africa efficient? Spatial price analyses and beyond.* Food Security Collaborative Working Papers 58562. Michigan State University: Department of Agricultural, Food, and Resource Economics.

Renard, D., & Tilman, D. (2019). National food production stabilized by crop diversity. *Nature, 571*(7764), 257–260.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *Stata Journal, 11*, 159–206.

Rose, E. (2001). Ex ante and ex post labor supply response to risk in a low-income area. *Journal of Development Economics, 64*(2), 371–388.

Sheahan, M., & Barrett, C. B. (2017). Ten striking facts about agricultural input use in Sub-Saharan Africa. *Food Policy, 67*, 12–25.

Sisay, D. T., Verhees, F. J., & van Trijp, H. C. (2017). Seed producer cooperatives in the Ethiopian seed sector and their role in seed supply improvement: A review. *Journal of Crop Improvement*, 1–33.

Sperling, L. (2020). Seed security response during COVID-19: Building on evidence and orienting to the future. *Food Security*, 1–5.

Sperling, L., Cooper, H. D., & Remington, T. (2008). Moving towards more effective seed aid. *The Journal of Development Studies, 44*(4), 586–612.

Taffesse, A. S., Dorosh, P., & Gemessa, S. A. (2012). Crop production in Ethiopia: Regional patterns and trends. *Food and Agriculture in Ethiopia: Progress and Policy Challenges*, 53–83.

Takahashi, K., Muraoka, R., & Otsuka, K. (2020). Technology adoption, impact, and extension in developing countries' agriculture: A review of the recent literature. *Agricultural Economics, 51*(1), 31–45.
Takasaki, Y. (2011). Do the commons help augment mutual insurance among the poor? *World Development, 39*(3), 429–438.
Tesfaye, W., & Tirivayi, N. (2020). Crop diversity, household welfare and consumption smoothing under risk: Evidence from rural Uganda. *World Development, 125* 104686.
Thijssen, M. H., Bishaw, Z., Beshir, A., De Boef, W. S., & (eds). (2008). *Farmers, seeds and varieties: Supporting informal seed supply in Ethiopia* (p. 348). Wageningen, The Netherlands: Wageningen International.
Thornton, P. K., & Herrero, M. (2015). Adapting to climate change in the mixed crop and livestock farming systems in sub-Saharan Africa. *Nature Climate Change, 5*(9), 830–836.
Ward, P. S., & Shively, G. E. (2015). Migration and Land Rental as Responses to Income Shocks in Rural China. *Pacific Economic Review, 20*(4), 511–543.
Westengen, O. T., Jeppson, S., & Guarino, L. (2013). Global ex-situ crop diversity conservation and the Svalbard Global Seed Vault: Assessing the current status. *PloS one, 8*(5), e64146.
Westengen, O. T., Ring, K. H., Berg, P. R., & Brysting, A. K. (2014). Modern maize varieties going local in the semi-arid zone in Tanzania. *BMC Evolutionary Biology, 14*(1), 1.
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.
Wooldridge, J. M. (2011). Fractional response models with endogeneous explanatory variables and heterogeneity. *CHI11 Stata Conference: Stata Users Group*.
Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics, 211*(1), 137–150.
Yirga, C., Mohammad, A., Kassie, M., Groote, H. d., Mebratu, T., Jaleta, M. & Shiferaw, B. (2013). *Analysisof Adoption and Diffusion of Improved Wheat Technologies inEthiopia*: Ethiopian Institute of Agricultural Research (EIAR) and CIMMYT.
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57*(298), 348–368.

Appendix: *Crops in crises: shocks shape smallholders' diversification in rural Ethiopia*

Clifton Makate* , Arild Angelsen, Stein Terje Holden, & Ola Tveitereid Westengen

- ***Distribution of households in panel***

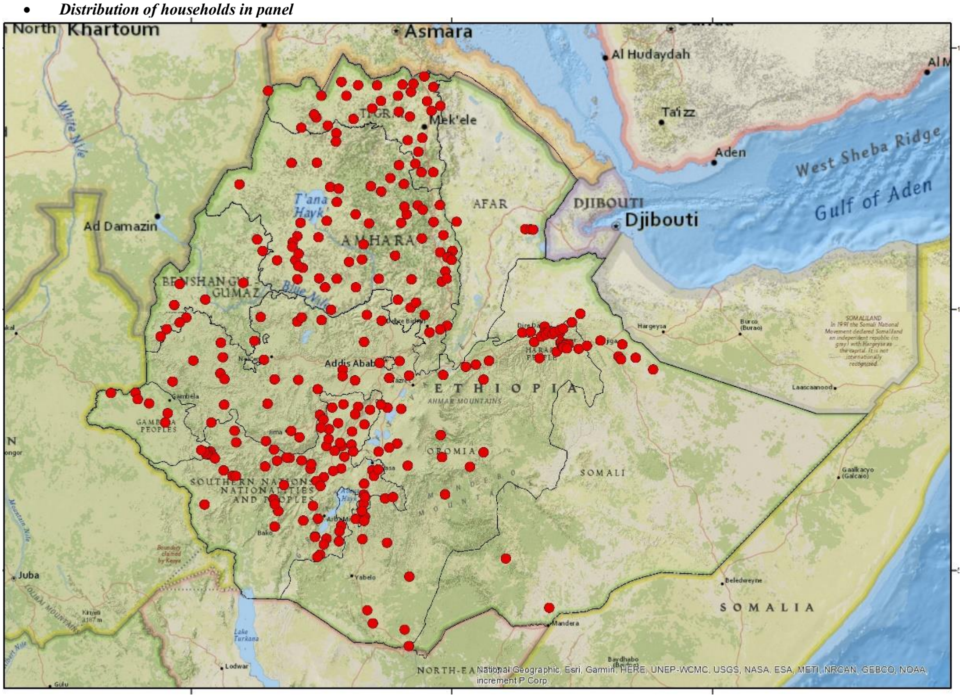

Figure i: Distribution of rural households included in the panel. Red dots represent clusters from which households were sampled

- ***LSMS-ISA Data description***

The LSMS-ISA survey contained various modules on, for example, household characteristics, agricultural activities, asset and livestock ownership, and geographical and spatial variables. This study uses data from the agricultural crop module, seed use module, household shock exposure module, and livestock modules. The seed module collects data on seed use and quantities for different crops, and information from that module is used to identify seed use decisions. Also, the agricultural crop module collects information on various crops grown by the household in the primary rainy season (*Meher* season). That information is used to construct crop diversification indices. More information about the household and its related characteristics is used to define other control variables used in the analysis. Although the bulk of agricultural data is collected at the plot level, the agricultural seed module is administered at the household level; hence all data used in this study was collapsed and analyzed at the household level.

- ***Historical rainfall data processing & descriptions***

In addition to the LSMS-ISA data for Ethiopia, we also use historical monthly weather data from 1980 to 2017 taken from WorldClim (Fick and Hijmans 2017; Masarie and Tans 1995) to define historical climate variables (precipitation and maximum temperature) and lagged climate shocks. The rainfall climate data from WorldClim are at a high spatial resolution of 2.5 minutes (which is approximately ~21 km2). These data can be easily accessed online on the following website https://www.worldclim.org/data/monthlywth.html. We take georeferenced data for households, which is available with LSMS-ISA. The georeferenced data include the longitude and latitude of interviewed households, enabling us to merge survey data with historical climate data. In addition, we also obtained climate data from a different source, NASA's Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al. 2017; Rienecker et al. 2011), and used to check the sensitivity of our results to an alternative source of climate data. MERRA-2 data can be accessed through the following NASA data access viewer tool: https://power.larc.nasa.gov/data-access-viewer/. We process MERRA-2 data the same way we did with WorldClim data. Figure ii and Figure iii show distribution of rainfall in the the Meher season by weather data source.

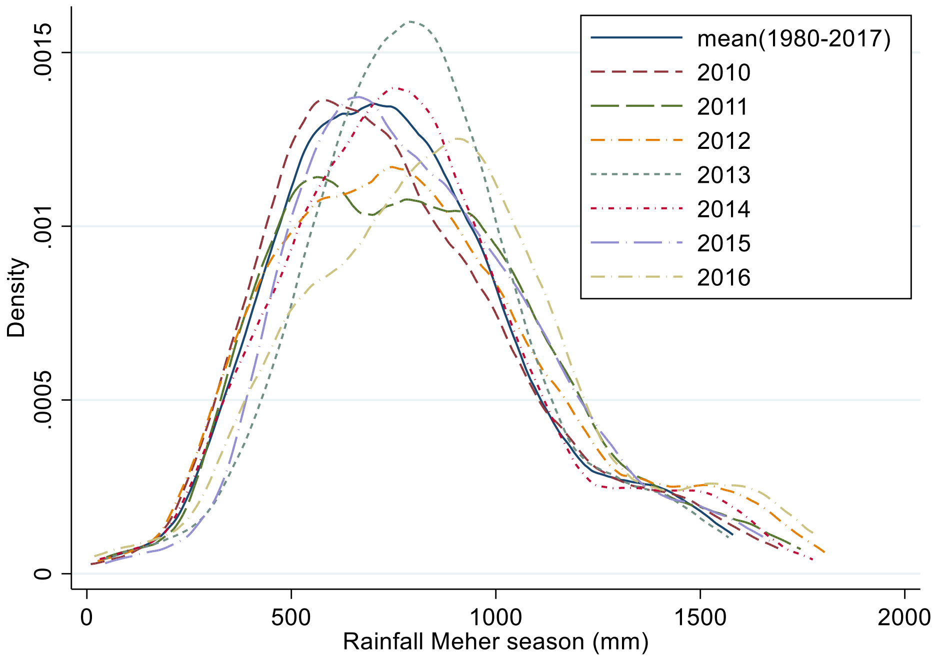

Figure ii: Rainfall distribution for growing seasons from 2010 to 2016, and the historical average for growing season (1980-2017). The plots are based on data for households included in sample and it's based on the WorldClim weather data.

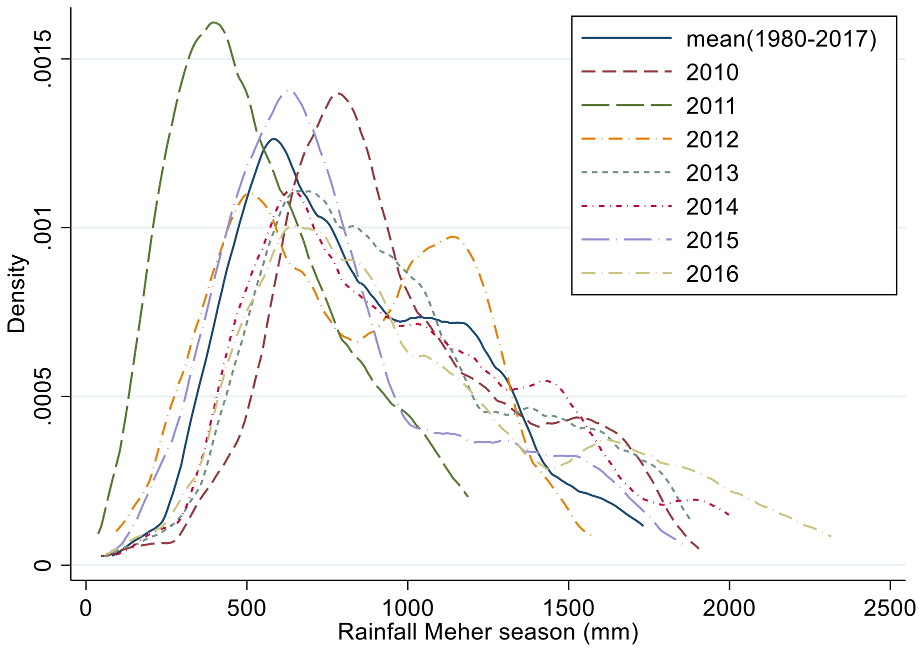

Figure iii: Rainfall distribution for growing seasons from 2010 to 2016, and the historical average for growing season (1980-2017). The plots are based on data for households included in sample and it's based on the MERRA 2 weather data.

Estimating crop diversification indices: We use two measures of crop diversification, the crop count index and Simpson index of diversity.

i. *Crop count (number of crops grown)*-Crop count is measured as the total number of crops grown by the household and it reflect crop species richness.

ii. *Simpson index of diversity*-The Simpson index of crop diversification considers both the number of crops grown and the land share of the different crops in the cropping portfolio. In other words the Simpson index of crop diversification reflect both species richness and evenness. Following Budescu and Budescu (2012), we estimate the Simpson index of crop diversification as follows: $Simpson_{index} = \sum_{i=1}^{H} P_i\,(1 - P_i) = 1 - \sum_{1=1}^{H} P_i^2$, where P_i is the share(land area) in the cropping portfolio of the crop category i, and $i = 1, \ldots., H$, and H is the number of mutually exclusive crop categories. $P_i \geq 0$ *and* $\sum_{i=1}^{H} P_i = 1$, implying that the index ranges from 0 (no diversification, i.e. only one crop) to 1 (highest degree of diversity, i.e., maximum number of crops with all having the same share).

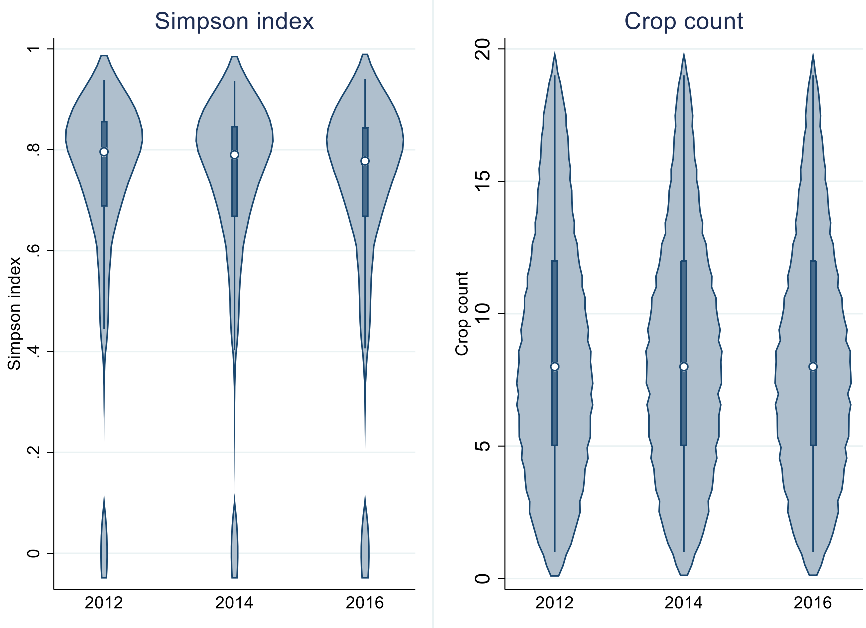

Figure iv: Crop diversification indices by survey round

Descriptive statistics for other control variables considered

Table A: Descriptive statistics for control variables used in the analysis

Variable	Variable definitions	2012 mean	2014 mean	2016 mean
farm size	Farm size owned (hectares)	1.197	1.379	1.361
Household asset index	Household asset index[1]	-0.003	-0.451	-0.194
Total Livestock Units	Total livestock units (TLU)	3.023	3.535	3.971
Literate household members (#)	Literate household members (#)	5.357	5.262	5.251
Mean Slope on the field (% and degrees)	Mean Slope on the field (% and degrees) on the field (% and degrees)	12.97	-	-
Hired male labor (#)	Hired male labor (#)	2.156	2.111	1.849
Elevation (m)	Elevation (m)	1728.4	-	-
Distance to the nearest market in km	Distance to the nearest market in km	66.464	-	-
Distance to the nearest paved road in km	Distance to the nearest paved road in km	15.346	-	-
Household dependency ratio (%)	Household dependency ratio[2] (%)	70.256	108.561	107.204
Sex of household head (1=female; 0 otherwise)	Sex of household head (1=female; 0 otherwise)	0.177	0.193	0.202
Age of household head (years)	Age of household head (years)	44.538	46.175	48.119
Age of household head squared	Age of household head squared	2204.081	2342.197	2529.121
The household head is single	The household head is single (divorced, separated, or widowed) †	0.148	0.176	0.201
Received safety nets (any)	Household received social safety nets of any form†	0.153	0.122	0.202
Received assistance of free food†	Received assistance of free food†	0.064	0.043	0.134
Recipient of the Productive Safety Net(PNSP)	Recipient of the Productive Safety Net (PNSP) program support†	0.036	0.043	0.055
-	Asset poor† (*Households in the two lowest quintiles of asset-wealth*)	0.392	0.428	0.445
-	Less than average farm size†	0.427	0.401	0.372
Observations		2398	2398	2398

Notes: summary statistics are not weighted, † denotes dummy variable

[1] The index is generated from principal components analysis. The index considers; (i) ownership of household durable goods(household assets and equipment), (ii) ownership of agricultural equipment (agricultural assets), and (iii) housing conditions (characteristics of the homestead and access to essential services (e.g. clean water, electricity and ablution services).

[2] The household dependency ratio is estimated as a ratio of household dependency (members aged below 15 years and above 65 years) to the economic active population (aged between 15 and 65) expressed as a percentage.

Full Tables showing all variables controlled and Mundlak and Chamberlain controls

- Tables B, C and D show the full tables for the main results reported in Manuscript on the influence of shocks on seed use decisions and diversification with additional controls.

Table B: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Rainfall shortage 1-year lag	2.740***	1.723***	0.448	0.265***
	(0.8786)	(0.5491)	(0.2594)	(0.0929)
Rainfall shortage 2-year lag	0.418***	-1.414***	3.790***	0.014
	(0.0659)	(0.2722)	(1.8145)	(0.0414)
Temperature shock 1-year lag	1.097	0.221	0.471***	-0.196***
	(0.1064)	(0.1701)	(0.0928)	(0.0269)
Temperature shock 2-year lag	1.662***	0.858***	4.432***	0.104***
	(0.2521)	(0.2643)	(0.8464)	(0.0374)
Historical mean temperature (1980-2017)	0.959*	-0.124***	0.937	-0.104***
	(0.0239)	(0.0449)	(0.0381)	(0.0078)
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.0001**
	(0.0003)	(0.0005)	(0.0004)	(0.0001)
Livestock loss†	0.962	0.041	1.324	-0.037
	(0.1928)	(0.3428)	(0.6348)	(0.0557)
Job loss†	0.875	-0.063	0.972	0.072
	(0.4314)	(0.8468)	(0.9032)	(0.1532)
Log of farm size(ha)	1.104	0.392***	1.619***	0.347***
	(0.0882)	(0.1395)	(0.2150)	(0.0214)
Log of Tropical Livestock Units (TLU)	0.995	-0.019	0.758	-0.019
	(0.1157)	(0.1964)	(0.2072)	(0.0319)
Log number of literate household members	1.074	0.164	0.777	0.060
	(0.2358)	(0.3708)	(0.3779)	(0.0626)
Slope on field (% and degrees)	0.982	-0.021	0.997	0.007*
	(0.0145)	(0.0247)	(0.0301)	(0.0034)
Log elevation(meters)	1.037	0.073	1.352**	0.088***
	(0.0663)	(0.1080)	(0.1652)	(0.0168)
Log distance to market (km)	0.025*	-7.036*	1.244	0.367
	(0.0527)	(3.6198)	(6.6496)	(0.4834)
Log distance to paved road(km)	1.029	0.611	0.204	-0.203
	(0.5087)	(0.8797)	(0.2469)	(0.1545)
Household dependency ratio (%)	1.001	0.001	1.001	-0.000
	(0.0008)	(0.0014)	(0.0020)	(0.0002)
Female household head(1=yes)	1.104	-0.021	0.905	-0.164
	(0.4751)	(0.7356)	(0.8327)	(0.1229)
Age of household head(years)	0.929	-0.116	1.070	-0.006
	(0.0467)	(0.0852)	(0.1138)	(0.0139)
Square of age of household head(years)	1.001*	0.001*	1.000	-0.000
	(0.0005)	(0.0008)	(0.0010)	(0.0001)
Household head is single(1=yes)	1.331	0.728	1.176	0.043
	(0.4224)	(0.5400)	(0.7306)	(0.0872)
Received assistance of free food(1=yes)	0.791	-0.258	1.382	0.095*
	(0.1773)	(0.3908)	(0.7003)	(0.0563)
Recipient of the Productive Safety Net (PNSP)(1=yes)	1.305	0.917*	0.437	-0.062
	(0.3822)	(0.4968)	(0.2761)	(0.0775)
Mundlak and Chamberlain controls				
mean_log_farmsize	1.105	0.285	1.270	0.532***
	(0.1258)	(0.2029)	(0.2215)	(0.0333)
mean_logTLU_0	1.310	0.584*	2.346**	0.438***
	(0.2286)	(0.3051)	(0.7919)	(0.0521)
mean_logliteracy_count	1.278	0.264	1.246	-0.295***
	(0.4022)	(0.5491)	(0.7400)	(0.0948)
mean_slope	0.953***	-0.095***	1.041	-0.007*
	(0.0160)	(0.0286)	(0.0345)	(0.0042)
mean_logelevation	1.415**	0.646***	0.751	0.054
	(0.2005)	(0.2506)	(0.1853)	(0.0426)
mean_log_distance_mkt	31.972	6.537*	0.862	-0.380
	(6.8443)	(3.625)	(0.4605)	(0.4842)
mean_log_distance_road	0.820	-0.865	4.796	0.173
	(0.4088)	(0.8871)	(5.7877)	(0.1560)
mean_hh_depend_ratio	1.000	0.000	1.000	0.000
	(0.0015)	(0.0026)	(0.0026)	(0.0005)
mean_sex_hhh_female	0.911	0.078	0.509	0.168
	(0.4663)	(0.8907)	(0.5085)	(0.1522)
mean_age_hhh	1.130**	0.190*	0.904	0.012
	(0.0680)	(0.1039)	(0.1070)	(0.0171)
mean_Age_sqrd	0.999**	-0.002**	1.001	-0.000
	(0.0006)	(0.0010)	(0.0011)	(0.0002)
mean_single_dsw	0.690	-1.042	1.547	-0.096
	(0.3136)	(0.7963)	(1.1918)	(0.1352)
mean_assistance_freefood	0.527	-1.166	2.453	0.039
	(0.2396)	(0.8090)	(1.8192)	(0.1337)
mean_assistance_PSNP	1.375	0.299	4.195	0.676***
	(0.7856)	(1.0166)	(4.1444)	(0.1646)

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
year_14	2.140***	1.331***	7.394***	0.443***
	(0.5749)	(0.4642)	(2.4712)	(0.0657)
year_16	1.792**	0.873**	17.004***	0.406***
	(0.4511)	(0.4344)	(6.4818)	(0.0630)
lnsig2u	4.147***		1.156	
	(0.3953)		(0.5568)	
sigma_u		3.754***		0.809***
		(0.1448)		(0.0182)
sigma_e		3.453***		0.938***
		(0.0886)		(0.0102)
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors at EA level in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table C: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Rainfall shortage 1-year lag	0.987	-0.078	0.640	-0.010	5.603**	5.510**	0.911	-0.200
	(0.4659)	(0.6032)	(0.3057)	(0.1637)	(3.7960)	(2.2085)	(0.7994)	(0.2737)
Rainfall shortage 2-year lag	0.226***	-1.935***	6.049***	0.433***	1.270	0.920	0.845	-0.055
	(0.0540)	(0.3047)	(1.6576)	(0.0723)	(0.5297)	(1.3455)	(0.4168)	(0.1733)
Temperature shock 1-year lag	1.227	0.337	1.157	0.035	1.312	0.893	0.808	-0.125
	(0.1961)	(0.2071)	(0.2004)	(0.0539)	(0.2933)	(0.7263)	(0.2177)	(0.0916)
Temperature shock 2-year lag	1.195	0.289	1.034	-0.002	2.799**	3.042**	0.634	-0.201*
	(0.2718)	(0.3044)	(0.2311)	(0.0730)	(1.2516)	(1.3378)	(0.2759)	(0.1210)
Historical mean temperature (1980-2017)	0.770***	-0.407***	1.224***	0.002	1.044	0.104	0.904*	-0.100***
	(0.0360)	(0.0626)	(0.0526)	(0.0158)	(0.0515)	(0.1578)	(0.0541)	(0.0201)
Historical mean rainfall (1980-2017)	1.004***	0.005***	0.998***	-0.000***	1.000	-0.002	1.000	-0.001***
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0005)	(0.0002)
Livestock loss†	1.318	0.479	1.101	0.146	0.760	-0.677	0.902	-0.084
	(0.3770)	(0.3639)	(0.3529)	(0.1050)	(0.3495)	(1.4601)	(0.4557)	(0.1690)
Job loss†	0.846	-0.142	3.020	0.536*	1.307	0.666	0.653	-0.129
	(0.5792)	(0.8580)	(2.4233)	(0.2831)	(1.5805)	(4.0002)	(0.8213)	(0.5173)
Log of farm size(ha)	0.979	0.270*	1.274**	0.288***	1.131	0.515	1.020	0.309***
	(0.1110)	(0.1460)	(0.1525)	(0.0412)	(0.2412)	(0.6972)	(0.2396)	(0.0887)
Log of Tropical Livestock Units (TLU)	1.243	0.291	0.738*	-0.102*	0.838	-0.830	1.286	0.141
	(0.2065)	(0.2070)	(0.1324)	(0.0576)	(0.2016)	(0.7677)	(0.3722)	(0.1055)
Log number of literate household members	1.059	-0.020	1.804*	0.305**	1.208	0.690	0.945	0.011
	(0.3382)	(0.3952)	(0.6254)	(0.1223)	(0.5773)	(1.5336)	(0.5183)	(0.1948)
Slope on field (% and degrees)	0.992	-0.004	1.054**	0.014*	0.962	-0.128	1.045	0.002
	(0.0220)	(0.0282)	(0.0268)	(0.0070)	(0.0311)	(0.1028)	(0.0393)	(0.0100)
Log elevation(meters)	1.019	0.004	0.949	-0.002	1.106	0.328	0.844	-0.055
	(0.0917)	(0.1104)	(0.0881)	(0.0356)	(0.1417)	(0.4195)	(0.1184)	(0.0520)
Log distance to market (km)	0.002**	-8.456**	72.294	2.001**	0.010	-15.972	10.910	0.752
	(0.0063)	(3.3866)	(235.7965)	(0.8057)	(0.0535)	(18.4266)	(70.1965)	(2.4270)
Log distance to paved road(km)	2.862	1.858*	1.022	-0.068	3.963	4.849	0.277	-0.429
	(2.1063)	(0.9521)	(0.8419)	(0.2810)	(4.3221)	(3.7466)	(0.3203)	(0.4808)
Household dependency ratio (%)	1.001	0.001	1.002	-0.000	1.000	0.002	0.999	-0.001
	(0.0012)	(0.0014)	(0.0012)	(0.0005)	(0.0017)	(0.0054)	(0.0019)	(0.0007)
Female household head(1=yes)	2.197	1.052	1.267	0.108	0.719	-2.141	0.948	-0.515
	(1.5945)	(0.8905)	(0.9935)	(0.2612)	(0.6006)	(2.7958)	(0.9223)	(0.3553)
Age of household head(years)	0.939	-0.057	1.147*	0.047*	0.845*	-0.507	1.308**	0.079*
	(0.0664)	(0.0877)	(0.0877)	(0.0265)	(0.0825)	(0.3176)	(0.1545)	(0.0414)
Square of age of household head(years)	1.001	0.001	0.999*	-0.001**	1.002*	0.005*	0.997**	-0.001**
	(0.0007)	(0.0008)	(0.0007)	(0.0002)	(0.0009)	(0.0030)	(0.0011)	(0.0004)
Household head is single(1=yes)	1.479	0.531	1.047	-0.009	0.547	-1.269	1.924	0.052
	(0.7121)	(0.5798)	(0.5241)	(0.1819)	(0.4058)	(2.4743)	(1.6183)	(0.2864)
Received assistance of free food(1=yes)	0.894	-0.103	1.171	0.199*	0.821	-0.315	1.164	0.091
	(0.3509)	(0.5266)	(0.5146)	(0.1163)	(0.3675)	(1.4317)	(0.6345)	(0.1741)
Recipient of the Productive Safety Net (PNSP)(1=yes)	1.737	1.147*	0.451	-0.355**	6.361***	5.925***	0.229***	-0.666***
	(0.8720)	(0.6121)	(0.2286)	(0.1705)	(2.8177)	(1.4046)	(0.1150)	(0.2003)
Mundlak and Chamberlain controls								
mean_log_farmsize	1.140	0.189	1.058	0.496***	0.682	-1.042	1.988**	0.657***
	(0.1959)	(0.2288)	(0.1797)	(0.0618)	(0.1740)	(0.8343)	(0.5817)	(0.1079)
mean_logTLU_0	1.095	0.265	0.895	0.141	1.351	1.095	0.572	-0.173
	(0.2811)	(0.3356)	(0.2287)	(0.0911)	(0.4459)	(1.0587)	(0.2253)	(0.1435)
mean_logliteracy_count	0.961	-0.109	0.575	-0.302*	1.144	0.622	1.231	-0.076
	(0.4527)	(0.6121)	(0.2766)	(0.1731)	(0.6568)	(1.8533)	(0.8105)	(0.2429)
mean_slope	0.939**	-0.089***	1.018	0.005	0.979	-0.066	1.016	0.005
	(0.0238)	(0.0328)	(0.0278)	(0.0082)	(0.0339)	(0.1101)	(0.0410)	(0.0112)
mean_logelevation	0.978	-0.071	1.309	0.245***	0.880	-0.240	1.430*	0.178**

	(0.2200)	(0.2973)	(0.2664)	(0.0834)	(0.1753)	(0.6544)	(0.3048)	(0.0896)
mean_log_distance_mkt	34.9406*	7.944**	0.014	-	76.532	14.991	0.114	-0.649
	(10.5799)	(3.3920)	(0.0461)	2.082***	(40.1041)	(1.8426)	(0.7327)	(2.4276)
				(0.8066)				
mean_log_distance_road	0.239*	-2.341**	1.410	0.170	0.228	-5.111	4.279	0.472
	(0.1773)	(0.9600)	(1.1721)	(0.2832)	(0.2481)	(3.7415)	(4.9538)	(0.4825)
mean_hh_depend_ratio	1.003	0.003	0.996**	-0.001	1.002	0.005	1.000	-0.001
	(0.0022)	(0.0029)	(0.0021)	(0.0008)	(0.0025)	(0.0080)	(0.0029)	(0.0011)
mean_sex_hhh_female	0.330	-1.389	1.208	-0.035	1.947	2.608	0.430	0.211
	(0.2836)	(1.0834)	(1.0786)	(0.3092)	(1.8034)	(3.0775)	(0.4579)	(0.4038)
mean_age_hhh	1.136	0.124	0.797**	-0.056*	1.141	0.341	0.757**	-0.068
	(0.0983)	(0.1107)	(0.0725)	(0.0317)	(0.1305)	(0.3716)	(0.1062)	(0.0471)
mean_Age_sqrd	0.999*	-0.002	1.002***	0.001**	0.998	-0.004	1.003**	0.001
	(0.0008)	(0.0011)	(0.0009)	(0.0003)	(0.0011)	(0.0036)	(0.0014)	(0.0004)
mean_single_dsw	0.522	-1.039	1.031	-0.065	1.020	0.080	2.143	0.201
	(0.3697)	(0.9052)	(0.7157)	(0.2603)	(0.9087)	(2.9406)	(2.1716)	(0.3607)
mean_assistance_freefood	0.147**	-2.389**	7.698**	0.283	0.572	-2.178	2.800	0.041
	(0.1217)	(1.1073)	(6.3776)	(0.2472)	(0.4027)	(2.2576)	(2.3937)	(0.2989)
mean_assistance_PSNP	1.466	0.224	0.949	0.461	0.288	-3.850	3.432	0.757**
	(1.4437)	(1.3091)	(0.8905)	(0.3181)	(0.2489)	(2.7501)	(3.4174)	(0.3553)
year_14	1.515	0.570	1.009	0.185	8.087***	6.254***	0.417	-0.192
	(0.6233)	(0.5442)	(0.4021)	(0.1275)	(5.9219)	(2.1856)	(0.2922)	(0.2065)
year_16	1.581	0.454	0.673	-0.092	2.649	2.763	1.034	0.161
	(0.6036)	(0.5002)	(0.2469)	(0.1230)	(1.7385)	(1.9822)	(0.6765)	(0.1875)
lnsig2u	7.007***		4.637***		1.130		1.402	
	(0.8798)		(0.7003)		(0.4301)		(0.5960)	
sigma_u		3.656***		1.133***		3.217***		0.651***
		(0.1625)		(0.0345)		(0.6041)		(0.0581)
sigma_e		2.774***		1.371***		6.497***		1.414***
		(0.0852)		(0.0209)		(0.4519)		(0.0321)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors at EA level; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table D: Impact of shocks on crop diversification decisions in rural Ethiopia

	Crop Count	Simpson Index
	(APE)	(APE)
Rainfall shortage 1-year lag	-0.0613*	-0.0013
	(0.032068)	(0.013232)
Rainfall shortage 2-year lag	0.0116	-0.0052
	(0.014680)	(0.005873)
Temperature shock 1-year lag	-0.0062	-0.0041
	(0.009124)	(0.003717)
Temperature shock 2-year lag	-0.0210	0.0157***
	(0.013832)	(0.005276)
Historical mean temperature (1980-2017)	-0.0182***	-0.0101***
	(0.004395)	(0.001243)
Historical mean rainfall (1980-2017)	0.0004***	0.0002***
	(0.000045)	(0.000013)
Livestock loss†	0.0014	-0.0205**
	(0.019120)	(0.007966)
Job loss†	0.0783	-0.0050
	(0.052668)	(0.021580)
Log of farm size(ha)	0.1747***	0.0359***
	(0.007833)	(0.002925)
Log of Tropical Livestock Units (TLU)	0.0665***	0.0044
	(0.010583)	(0.004471)
Log number of literate household members	0.0195	0.0018
	(0.020455)	(0.008707)
Slope on field (% and degrees)	-0.0046***	-0.0001
	(0.001036)	(0.000473)
Log elevation(meters)	0.0268***	0.0107***
	(0.005960)	(0.002326)
Log distance to market (km)	-0.0075	0.1416**
	(0.199906)	(0.068860)
Log distance to paved road(km)	-0.0667	-0.0029
	(0.049272)	(0.021392)
Household dependency ratio (%)	-0.0000	0.0000
	(0.000077)	(0.000033)
Female household head(1=yes)	-0.0544	-0.0095
	(0.041574)	(0.017002)
Age of household head(years)	0.0191***	0.0026
	(0.004655)	(0.001912)
Square of age of household head(years)	-0.0002***	-0.0000
	(0.000043)	(0.000018)
Household head is single(1=yes)	0.0291	-0.0085
	(0.029843)	(0.012137)
Received assistance of free food(1=yes)	-0.0278	-0.0289***
	(0.020691)	(0.007914)
Recipient of the Productive Safety Net (PNSP)(1=yes)	-0.0179	0.0102

	Crop Count	Simpson Index
	(APE)	(APE)
	(0.027854)	(0.010861)
Mundlak and Chamberlain controls		
mean_log_farmsize	0.0447***	0.0160***
	(0.016337)	(0.005036)
mean_logTLU_0	-0.1793***	-0.0520***
	(0.025461)	(0.007985)
mean_logliteracy_count	0.1758***	0.0394***
	(0.042339)	(0.014376)
mean_slope	0.0233***	0.0035***
	(0.001663)	(0.000611)
mean_logelevation	0.1128***	0.0285***
	(0.022196)	(0.006706)
mean_log_distance_mkt	-0.0289	-0.1167*
	(0.0201)	(0.068993)
mean_log_distance_road	0.0097	-0.0030
	(0.050684)	(0.021671)
mean_hh_depend_ratio	-0.0004*	-0.0001*
	(0.000226)	(0.000072)
mean_sex_hhh_female	-0.0300	0.0156
	(0.062639)	(0.022265)
mean_age_hhh	-0.0018	0.0025
	(0.006963)	(0.002494)
mean_Age_sqrd	0.0000	-0.0000
	(0.000066)	(0.000024)
mean_single_dsw	-0.0064	-0.0180
	(0.061535)	(0.020554)
mean_assistance_freefood	-0.4089***	-0.1368***
	(0.069984)	(0.021123)
mean_assistance_PSNP	-0.0239	-0.0404
	(0.082892)	(0.025765)
year_14	-0.1206***	0.0066
	(0.024128)	(0.009180)
year_16	-0.0962***	-0.0017
	(0.023384)	(0.008835)
lnalpha	-1.4363***	
	(0.032542)	
sigma_u		0.1374***
		(0.002709)
sigma_e		0.1314***
		(0.001394)
Observations	7194	7194

In parenthesis are cluster robust standard errors at EA level; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Tables reporting effects of rainfall shocks in general (flood and drought shocks in one variable) on seed use decisions and crop diversification

Table E: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Rainfall shock 1-year lag	0.553***	-0.994***	0.806	-0.108***
	(0.0680)	(0.2159)	(0.1371)	(0.0348)
Rainfall shock 2-year lag	1.104**	0.183**	0.884	-0.010
	(0.0531)	(0.0868)	(0.1206)	(0.0145)
Temperature shock 1-year lag	1.358***	0.549***	0.438***	-0.168***
	(0.1463)	(0.1904)	(0.0992)	(0.0310)
Temperature shock 2-year lag	1.335**	0.540**	6.195***	0.131***
	(0.1919)	(0.2584)	(1.1138)	(0.0386)
Historical mean temperature (1980-2017)	0.945**	-0.139***	0.942*	-0.090***
	(0.0219)	(0.0434)	(0.0325)	(0.0100)
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.001***
	(0.0003)	(0.0005)	(0.0004)	(0.0001)
Livestock loss†	0.882	-0.113	2.108	0.023
	(0.1764)	(0.3504)	(1.0248)	(0.0598)
Job loss†	1.028	0.179	0.591	0.052
	(0.5096)	(0.8660)	(0.5287)	(0.1639)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				
Rainfall shock 1-year lag	0.560***	-0.962***	0.905	-0.084***
	(0.0684)	(0.2110)	(0.1584)	(0.0323)
Rainfall shock 2-year lag	1.030	0.045	0.886	-0.042***
	(0.0505)	(0.0868)	(0.1267)	(0.0140)
Temperature shock 1-year lag	1.446***	0.663***	0.488***	-0.163***
	(0.1552)	(0.1872)	(0.1104)	(0.0294)
Temperature shock 2-year lag	1.363**	0.530**	5.685***	0.128***
	(0.2001)	(0.2578)	(1.0158)	(0.0366)
Historical mean temperature (1980-2017)	0.945**	-0.148***	0.946	-0.103***
	(0.0232)	(0.0444)	(0.0385)	(0.0077)

	Improved seed		Local seed	
	Use	INT	Use	INT
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.000**
	(0.0003)	(0.0005)	(0.0005)	(0.0001)
Livestock loss†	0.953	0.009	1.367	-0.029
	(0.1913)	(0.3448)	(0.6592)	(0.0558)
Job loss†	0.890	-0.021	0.995	0.080
	(0.4404)	(0.8520)	(0.9150)	(0.1532)
Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors in parenthesis; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table F: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Rainfall shock 1-year lag	0.648**	-0.444*	1.792***	0.034	0.525***	-2.125***	1.349	0.274***
	(0.1255)	(0.2529)	(0.3749)	(0.0692)	(0.1221)	(0.8073)	(0.3422)	(0.1014)
Rainfall shock 2-year lag	1.365***	0.463***	0.741***	-0.098***	0.797	-0.791	1.307	0.080
	(0.0960)	(0.0963)	(0.0549)	(0.0265)	(0.1257)	(0.5507)	(0.2676)	(0.0675)
Temperature shock 1-year lag	1.466**	0.570**	0.939	-0.031	1.555**	1.538**	0.789	-0.088
	(0.2542)	(0.2313)	(0.1751)	(0.0572)	(0.3118)	(0.7213)	(0.1876)	(0.0954)
Temperature shock 2-year lag	0.751	-0.353	1.784***	0.220***	2.285**	2.836**	0.634	-0.440***
	(0.1637)	(0.3027)	(0.3848)	(0.0728)	(0.8112)	(1.1875)	(0.2248)	(0.1198)
Historical mean temperature (1980-2017)	0.752***	-0.437***	1.246***	-0.000	1.092**	0.301**	0.888**	-0.080***
	(0.0322)	(0.0596)	(0.0503)	(0.0158)	(0.0465)	(0.1509)	(0.0448)	(0.0223)
Historical mean rainfall (1980-2017)	1.004***	0.006***	0.998***	-0.000	1.000	-0.001	1.000	0.000
	(0.0004)	(0.0005)	(0.0004)	(0.0001)	(0.0004)	(0.0013)	(0.0004)	(0.0002)
Livestock loss†	1.109	0.281	1.415	0.267**	0.664	-1.255	1.030	-0.086
	(0.3150)	(0.3726)	(0.4567)	(0.1093)	(0.2801)	(1.4478)	(0.4680)	(0.1830)
Job loss†	0.866	-0.190	2.440	0.366	1.364	0.697	0.542	-0.490
	(0.5882)	(0.8767)	(1.9196)	(0.2947)	(1.6209)	(4.2667)	(0.6586)	(0.5603)
Models with additional controls								
Rainfall shock 1-year lag	0.639**	-0.503**	1.865***	0.055	0.603**	-1.515**	0.983	0.132
	(0.1224)	(0.2451)	(0.3898)	(0.0660)	(0.1433)	(0.7609)	(0.2586)	(0.0925)
Rainfall shock 2-year lag	1.261***	0.350***	0.788***	-0.106***	0.740*	-1.085*	1.325	-0.004
	(0.0915)	(0.0967)	(0.0612)	(0.0265)	(0.1296)	(0.5713)	(0.3006)	(0.0630)
Temperature shock 1-year lag	1.483**	0.557**	0.922	-0.028	1.543**	1.413**	0.804	-0.155*
	(0.2548)	(0.2251)	(0.1705)	(0.0557)	(0.3360)	(0.7169)	(0.2106)	(0.0914)
Temperature shock 2-year lag	0.758	-0.367	1.616**	0.157**	2.481**	2.687**	0.655	-0.227**
	(0.1682)	(0.2992)	(0.3529)	(0.0715)	(0.9788)	(1.1908)	(0.2642)	(0.1134)
Historical mean temperature (1980-2017)	0.743***	-0.459***	1.277***	0.015	1.047	0.108	0.911	-0.103***
	(0.0345)	(0.0629)	(0.0551)	(0.0157)	(0.0503)	(0.1545)	(0.0538)	(0.0197)
Historical mean rainfall (1980-2017)	1.004***	0.005***	0.998***	-0.000***	1.000	-0.002	1.000	-0.001***
	(0.0004)	(0.0005)	(0.0004)	(0.0001)	(0.0004)	(0.0014)	(0.0005)	(0.0002)
Livestock loss†	1.236	0.396	1.222	0.170	0.769	-0.632	0.901	-0.080
	(0.3542)	(0.3665)	(0.3948)	(0.1051)	(0.3508)	(1.4537)	(0.4554)	(0.1690)
Job loss†	0.752	-0.307	3.868*	0.546*	1.366	0.772	0.704	-0.191
	(0.5154)	(0.8647)	(3.1397)	(0.2832)	(1.6595)	(4.0378)	(0.8878)	(0.5192)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable.

Table G: Impact of shocks on crop diversification decisions in rural Ethiopia

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Rainfall shocks 1-year lag	0.001	-0.020***	0.001	-0.015***
	(0.0118)	(0.0047)	(0.0117)	(0.0046)
Rainfall shock 2-year lag	-0.014***	-0.008***	-0.007	-0.004*
	(0.0049)	(0.0020)	(0.0050)	(0.0020)
Temperature shock 1-year lag	-0.024**	0.003	-0.007	0.003
	(0.0104)	(0.0042)	(0.0105)	(0.0041)
Temperature shock 2-year lag	0.015	0.025***	-0.010	0.018***
	(0.0134)	(0.0052)	(0.0135)	(0.0051)
Historical mean temperature (1980-2017)	-0.034***	-0.016***	-0.018***	-0.010***
	(0.0047)	(0.0012)	(0.0044)	(0.0012)
Historical mean rainfall (1980-2017)	0.0001***	0.0001***	0.0001***	0.0001***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Livestock loss†	0.011	-0.019**	0.002	-0.020**
	(0.0191)	(0.0082)	(0.0191)	(0.0080)
Job loss†	0.091*	0.003	0.077	-0.005
	(0.0529)	(0.0222)	(0.0527)	(0.0216)
Year dummies	Yes	Yes	Yes	Yes

Other controls	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Tables reporting effects of flood shocks on seed use decisions and crop diversification

- In manuscript we report results showing the impact of drought shocks and temperature shocks on seed use decisions and diversification. Here we show results reporting the impact of flood shocks (more than normal rainfall received in the recent past) on seed use decisions and diversification.

Table H: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Flood shock 1-year lag	1.012	0.031	2.084***	0.003
	(0.0745)	(0.1321)	(0.3724)	(0.0228)
Flood shock 2-year lag	1.037	0.057	1.023	-0.027
	(0.0741)	(0.1272)	(0.2204)	(0.0217)
Temperature shock 1-year lag	1.084	0.186	0.359***	-0.214***
	(0.1052)	(0.1758)	(0.0687)	(0.0296)
Temperature shock 2-year lag	1.238	0.425	6.647***	0.128***
	(0.1788)	(0.2602)	(1.2233)	(0.0393)
Historical mean temperature (1980-2017)	0.932***	-0.163***	0.922**	-0.093***
	(0.0216)	(0.0435)	(0.0324)	(0.0100)
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.001***
	(0.0002)	(0.0004)	(0.0004)	(0.0001)
Livestock loss†	0.882	-0.118	2.384*	0.023
	(0.1763)	(0.3508)	(1.2017)	(0.0598)
Job loss†	1.016	0.169	0.525	0.043
	(0.5010)	(0.8628)	(0.4670)	(0.1640)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				
Rainfall shock 1-year lag	0.907	-0.165	2.175***	-0.040*
	(0.0665)	(0.1289)	(0.4149)	(0.0210)
Rainfall shock 2-year lag	0.949	-0.108	1.047	-0.082***
	(0.0680)	(0.1253)	(0.2322)	(0.0206)
Temperature shock 1-year lag	1.204*	0.381**	0.418***	-0.188***
	(0.1183)	(0.1748)	(0.0781)	(0.0279)
Temperature shock 2-year lag	1.210	0.334	6.255***	0.117***
	(0.1768)	(0.2573)	(1.1340)	(0.0372)
Historical mean temperature (1980-2017)	0.935***	-0.165***	0.938	-0.104***
	(0.0231)	(0.0446)	(0.0391)	(0.0076)
Historical mean rainfall (1980-2017)	1.002***	0.004***	1.001	0.000
	(0.0003)	(0.0004)	(0.0004)	(0.0001)
Livestock loss†	0.963	0.030	1.490	-0.026
	(0.1929)	(0.3443)	(0.7429)	(0.0558)
Job loss†	0.894	-0.014	0.840	0.070
	(0.4385)	(0.8470)	(0.7526)	(0.1532)
Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors in parenthesis; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table I: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Flood shock 1-year lag	1.057	0.115	1.259*	0.053	1.038	0.137	1.130	-0.041
	(0.1146)	(0.1489)	(0.1486)	(0.0406)	(0.1797)	(0.6198)	(0.2354)	(0.0844)
Flood shock 2-year lag	1.453***	0.556***	0.762***	-0.092**	0.696*	-1.340*	1.800**	0.093
	(0.1467)	(0.1381)	(0.0803)	(0.0395)	(0.1499)	(0.7307)	(0.5135)	(0.0899)
Temperature shock 1-year lag	1.339*	0.467**	0.902	-0.073	1.316	0.956	0.868	-0.021
	(0.2231)	(0.2235)	(0.1552)	(0.0568)	(0.2749)	(0.7482)	(0.2196)	(0.1024)
Temperature shock 2-year lag	0.680*	-0.457	2.086***	0.244***	1.894*	2.229**	0.725	-0.385***
	(0.1463)	(0.2995)	(0.4340)	(0.0739)	(0.6252)	(1.1157)	(0.2410)	(0.1177)
Historical mean temperature (1980-2017)	0.739***	-0.458***	1.273***	0.004	1.089**	0.292*	0.893**	-0.079***
	(0.0314)	(0.0594)	(0.0505)	(0.0157)	(0.0473)	(0.1519)	(0.0457)	(0.0225)
Historical mean rainfall (1980-2017)	1.004***	0.006***	0.999***	-0.000	0.999	-0.002	1.000	0.000
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0004)	(0.0002)
Livestock loss†	1.076	0.238	1.432	0.270**	0.664	-1.213	0.998	-0.100
	(0.3048)	(0.3730)	(0.4568)	(0.1094)	(0.2852)	(1.4538)	(0.4609)	(0.1829)
Job loss†	0.855	-0.191	2.317	0.364	1.014	-0.423	0.664	-0.352
	(0.5769)	(0.8757)	(1.7988)	(0.2948)	(1.2205)	(4.2878)	(0.8211)	(0.5584)
Models with additional controls								

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Rainfall shock 1-year lag	0.935	-0.040	1.382***	0.062	0.993	-0.020	1.190	-0.154**
	(0.1022)	(0.1467)	(0.1632)	(0.0396)	(0.1860)	(0.6160)	(0.2713)	(0.0777)
Rainfall shock 2-year lag	1.301**	0.418***	0.825*	-0.107***	0.640*	-1.577**	2.042**	0.001
	(0.1344)	(0.1376)	(0.0890)	(0.0392)	(0.1473)	(0.7266)	(0.6433)	(0.0830)
Temperature shock 1-year lag	1.405**	0.486**	0.852	-0.073	1.323	0.887	0.873	-0.064
	(0.2358)	(0.2206)	(0.1481)	(0.0555)	(0.3014)	(0.7472)	(0.2477)	(0.0970)
Temperature shock 2-year lag	0.652**	-0.550*	2.003***	0.192***	2.143**	2.301**	0.717	-0.205*
	(0.1419)	(0.2937)	(0.4197)	(0.0726)	(0.7965)	(1.1277)	(0.2730)	(0.1115)
Historical mean temperature (1980-2017)	0.731***	-0.478***	1.298***	0.019	1.044	0.108	0.917	-0.103***
	(0.0340)	(0.0628)	(0.0552)	(0.0156)	(0.0508)	(0.1545)	(0.0540)	(0.0196)
Historical mean rainfall (1980-2017)	1.003***	0.005***	0.999***	-0.000**	0.999	-0.002	1.000	-0.001***
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0005)	(0.0002)
Livestock loss†	1.220	0.372	1.209	0.174*	0.784	-0.579	0.877	-0.073
	(0.3481)	(0.3661)	(0.3860)	(0.1051)	(0.3622)	(1.4586)	(0.4480)	(0.1689)
Job loss†	0.761	-0.289	3.456	0.548*	1.122	0.097	0.720	-0.109
	(0.5170)	(0.8617)	(2.7524)	(0.2833)	(1.3745)	(4.0353)	(0.9085)	(0.5173)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable.

Table J: Impact of shocks on crop diversification decisions in rural Ethiopia

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Flood shock 1-year lag	0.0568***	0.0055*	0.0657***	0.0098***
	(0.007468)	(0.003066)	(0.007514)	(0.003015)
Flood shock 2-year lag	0.0111	-0.0043	0.0218***	0.0006
	(0.007195)	(0.002962)	(0.007315)	(0.002931)
Temperature shock 1-year lag	-0.0517***	-0.0088**	-0.0369***	-0.0077**
	(0.009732)	(0.003918)	(0.009768)	(0.003859)
Temperature shock 2-year lag	0.0269**	0.0235***	0.0042	0.0179***
	(0.013579)	(0.005300)	(0.013638)	(0.005215)
Historical mean temperature (1980-2017)	-0.0345***	-0.0164***	-0.0185***	-0.0103***
	(0.004607)	(0.001220)	(0.004323)	(0.001227)
Historical mean rainfall (1980-2017)	0.0005***	0.0002***	0.0004***	0.0002***
	(0.000047)	(0.000013)	(0.000044)	(0.000013)
Livestock loss†	0.0083	-0.0199**	-0.0010	-0.0206***
	(0.019127)	(0.008181)	(0.019115)	(0.007973)
Job loss†	0.0909*	0.0016	0.0781	-0.0056
	(0.052859)	(0.022220)	(0.052620)	(0.021567)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Reproducing main Tables reported in manuscript now using different Weather data source (MERRA2 Data) in measuring Climate variables and shocks

- Main effects (seed use decisions)

Table K: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Rainfall shortage 1-year lag	1.211	0.580*	1.991**	0.156***
	(0.2295)	(0.3358)	(0.5849)	(0.0574)
Rainfall shortage 2-year lag	0.810	-0.363	2.207**	-0.104***
	(0.1144)	(0.2549)	(0.7295)	(0.0403)
Temperature shock 1-year lag	1.147	0.277	0.335***	-0.222***
	(0.1077)	(0.1699)	(0.0656)	(0.0284)
Temperature shock 2-year lag	1.392**	0.636**	5.712***	0.175***
	(0.2130)	(0.2739)	(1.1296)	(0.0418)
Historical mean temperature (1980-2017)	0.929***	-0.172***	0.902***	-0.096***
	(0.0206)	(0.0413)	(0.0308)	(0.0096)
Historical mean rainfall (1980-2017)	1.003***	0.005***	0.999***	0.001***
	(0.0002)	(0.0004)	(0.0003)	(0.0001)
Livestock loss†	0.931	-0.022	1.645	0.027
	(0.1838)	(0.3477)	(0.7811)	(0.0597)
Job loss†	1.033	0.184	0.530	0.050
	(0.5043)	(0.8570)	(0.4789)	(0.1638)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				

	Improved seed		Local seed	
	Use	INT	Use	INT
Rainfall shortage 1-year lag	1.082	0.308	1.941**	0.201***
	(0.2053)	(0.3310)	(0.5793)	(0.0536)
Rainfall shortage 2-year lag	0.859	-0.276	1.973**	-0.111***
	(0.1246)	(0.2553)	(0.6535)	(0.0385)
Temperature shock 1-year lag	1.209**	0.383**	0.414***	-0.214***
	(0.1147)	(0.1689)	(0.0794)	(0.0268)
Temperature shock 2-year lag	1.328*	0.493*	5.431***	0.160***
	(0.2047)	(0.2703)	(1.0776)	(0.0393)
Historical mean temperature (1980-2017)	0.925***	-0.186***	0.911**	-0.105***
	(0.0219)	(0.0428)	(0.0371)	(0.0074)
Historical mean rainfall (1980-2017)	1.003***	0.004***	0.999*	0.000***
	(0.0002)	(0.0004)	(0.0004)	(0.0001)
Livestock loss†	0.985	0.067	1.174	-0.033
	(0.1953)	(0.3423)	(0.5535)	(0.0557)
Job loss†	0.915	0.027	0.887	0.086
	(0.4474)	(0.8450)	(0.8247)	(0.1530)
Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table L: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Rainfall shortage 1-year lag	0.805	-0.065	0.879	-0.046	2.017*	2.657**	0.372**	-0.251
	(0.2188)	(0.3630)	(0.2414)	(0.1008)	(0.7436)	(1.2901)	(0.1623)	(0.1671)
Rainfall shortage 2-year lag	0.642**	-0.558*	1.697**	0.142*	1.506	1.469	0.758	-0.033
	(0.1427)	(0.3065)	(0.4004)	(0.0755)	(0.3833)	(0.8954)	(0.2194)	(0.1169)
Temperature shock 1-year lag	1.504**	0.616***	0.869	-0.051	1.513**	1.491**	0.769	-0.081
	(0.2391)	(0.2149)	(0.1503)	(0.0551)	(0.2931)	(0.6988)	(0.1801)	(0.0945)
Temperature shock 2-year lag	0.867	-0.183	1.625**	0.129	1.685	1.856	0.709	-0.433***
	(0.1983)	(0.3177)	(0.3638)	(0.0783)	(0.5552)	(1.1334)	(0.2489)	(0.1266)
Historical mean temperature (1980-2017)	0.736***	-0.481***	1.222***	-0.010	1.114**	0.359**	0.868***	-0.094***
	(0.0299)	(0.0562)	(0.0456)	(0.0154)	(0.0469)	(0.1478)	(0.0430)	(0.0217)
Historical mean rainfall (1980-2017)	1.004***	0.006***	0.998***	-0.000***	1.000	0.000	1.000	-0.000**
	(0.0003)	(0.0004)	(0.0003)	(0.0001)	(0.0003)	(0.0009)	(0.0003)	(0.0001)
Livestock loss†	1.210	0.377	1.241	0.236**	0.683	-1.151	0.986	-0.139
	(0.3338)	(0.3680)	(0.3855)	(0.1098)	(0.2935)	(1.4641)	(0.4539)	(0.1833)
Job loss†	0.881	-0.196	2.244	0.359	1.138	-0.010	0.569	-0.423
	(0.5820)	(0.8698)	(1.7256)	(0.2951)	(1.3729)	(4.3180)	(0.7002)	(0.5578)
Models with additional controls								
Rainfall shortage 1-year lag	0.720	-0.262	0.927	-0.009	1.458	1.007	0.470	0.032
	(0.1987)	(0.3600)	(0.2568)	(0.0970)	(0.5780)	(1.2897)	(0.2237)	(0.1586)
Rainfall shortage 2-year lag	0.677*	-0.474	1.701**	0.118	1.676*	1.731*	0.673	-0.164
	(0.1568)	(0.3106)	(0.4230)	(0.0736)	(0.4676)	(0.9038)	(0.2167)	(0.1117)
Temperature shock 1-year lag	1.473**	0.557***	0.904	-0.053	1.516**	1.403**	0.766	-0.156*
	(0.2344)	(0.2113)	(0.1569)	(0.0536)	(0.3217)	(0.7032)	(0.2019)	(0.0901)
Temperature shock 2-year lag	0.890	-0.187	1.451*	0.071	1.873	1.856	0.702	-0.233*
	(0.2062)	(0.3135)	(0.3254)	(0.0762)	(0.7159)	(1.1878)	(0.2906)	(0.1197)
Historical mean temperature (1980-2017)	0.723***	-0.513***	1.264***	0.013	1.070	0.191	0.892**	-0.097***
	(0.0319)	(0.0597)	(0.0510)	(0.0152)	(0.0504)	(0.1510)	(0.0507)	(0.0187)
Historical mean rainfall (1980-2017)	1.003***	0.005***	0.998***	-0.001***	1.000	0.000	1.000	-0.001***
	(0.0003)	(0.0004)	(0.0003)	(0.0001)	(0.0003)	(0.0009)	(0.0003)	(0.0001)
Livestock loss†	1.300	0.443	1.132	0.144	0.770	-0.672	0.917	-0.119
	(0.3620)	(0.3633)	(0.3539)	(0.1056)	(0.3542)	(1.4694)	(0.4645)	(0.1688)
Job loss†	0.798	-0.257	3.414	0.534*	1.232	0.398	0.655	-0.121
	(0.5326)	(0.8608)	(2.7106)	(0.2836)	(1.5119)	(4.0881)	(0.8273)	(0.5160)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable.

Table M: Impact of shocks on crop diversification decisions in rural Ethiopia

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Rainfall shortage 1-year lag	-0.114***	-0.006	-0.108***	-0.005
	(0.0197)	(0.0078)	(0.0197)	(0.0076)
Rainfall shortage 2-year lag	0.026**	0.001	0.021	-0.002
	(0.0130)	(0.0055)	(0.0131)	(0.0054)
Temperature shock 1-year lag	-0.023**	-0.006	-0.006	-0.003
	(0.0091)	(0.0037)	(0.0091)	(0.0037)

Temperature shock 2-year lag	-0.005	0.019***	-0.022	0.014**
	(0.0144)	(0.0057)	(0.0144)	(0.0056)
Historical mean temperature (1980-2017)	-0.042***	-0.018***	-0.025***	-0.012***
	(0.0045)	(0.0012)	(0.0043)	(0.0012)
Historical mean rainfall (1980-2017)	0.0001***	0.0001***	0.0001***	0.0001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Livestock loss†	0.009	-0.020**	0.002	-0.020**
	(0.0191)	(0.0082)	(0.0191)	(0.0080)
Job loss†	0.086	0.0001	0.073	-0.006
	(0.0529)	(0.0222)	(0.0527)	(0.0216)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Reproducing main Tables in the manuscript but now applying a different estimation strategy (conditional mixed process (CMP) framework)

Table N: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Rainfall shortage 1-year lag	2.016***	3.699***	0.863	0.973***
	(0.2453)	(0.6116)	(0.1886)	(0.1279)
Rainfall shortage 2-year lag	0.592***	-2.605***	1.797***	0.233***
	(0.0341)	(0.2953)	(0.3029)	(0.0518)
Temperature shock 1-year lag	1.155***	0.699***	0.720***	-0.478***
	(0.0482)	(0.2218)	(0.0630)	(0.0484)
Temperature shock 2-year lag	1.216***	1.118***	2.072***	0.398***
	(0.0625)	(0.2832)	(0.1277)	(0.0685)
Historical mean temperature (1980-2017)	0.983***	-0.125***	0.980*	-0.103***
	(0.0055)	(0.0303)	(0.0115)	(0.0069)
Historical mean rainfall (1980-2017)	1.001***	0.006***	1.000	0.00043***
	(0.0001)	(0.0003)	(0.0001)	(0.0001)
Livestock loss†	0.948	-0.224	1.328	0.127*
	(0.0727)	(0.3920)	(0.2475)	(0.0757)
Job loss†	1.312	1.194	0.753	-0.292
	(0.2611)	(0.9636)	(0.2487)	(0.2354)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				
Rainfall shock 1-year lag	1.686***	2.451***	0.737	0.501***
	(0.2262)	(0.6333)	(0.1671)	(0.1009)
Rainfall shock 2-year lag	0.660***	-1.852***	1.692***	0.229***
	(0.0394)	(0.2858)	(0.3034)	(0.0442)
Temperature shock 1-year lag	1.217***	0.938***	0.767***	-0.324***
	(0.0514)	(0.2068)	(0.0666)	(0.0359)
Temperature shock 2-year lag	1.224***	0.995***	2.106***	0.256***
	(0.0684)	(0.2822)	(0.1522)	(0.0503)
Historical mean temperature (1980-2017)	0.981***	-0.145***	0.978	-0.111***
	(0.0065)	(0.0328)	(0.0150)	(0.0056)
Historical mean rainfall (1980-2017)	1.001***	0.005***	1.000	0.000
	(0.0001)	(0.0003)	(0.0002)	(0.0001)
Livestock loss†	0.997	0.004	1.122	-0.071
	(0.0788)	(0.3714)	(0.2217)	(0.0571)
Job loss†	1.178	0.734	0.983	0.006
	(0.2293)	(0.8720)	(0.3479)	(0.1445)
Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors in parenthesis; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable.

Table O: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Rainfall shortage 1-year lag	1.234	0.849	0.779	0.272	2.718***	7.408***	0.584*	-0.879***
	(0.1802)	(0.6500)	(0.1466)	(0.1926)	(0.7003)	(1.8981)	(0.1626)	(0.2704)
Rainfall shortage 2-year lag	0.497***	-3.211***	2.194***	0.641***	1.024	-0.070	0.971	-0.302
	(0.0386)	(0.3503)	(0.1945)	(0.0697)	(0.1884)	(1.3771)	(0.1916)	(0.1941)
Temperature shock 1-year lag	1.197***	0.847***	0.956	-0.050	1.235**	1.602**	0.894	-0.064
	(0.0669)	(0.2587)	(0.0528)	(0.0604)	(0.1116)	(0.6837)	(0.0888)	(0.0991)
Temperature shock 2-year lag	0.970	-0.072	1.152**	0.332***	1.431**	2.719**	0.808	-0.583***
	(0.0572)	(0.2751)	(0.0772)	(0.0796)	(0.2113)	(1.1327)	(0.1269)	(0.1144)
Historical mean temperature (1980-2017)	0.906***	-0.481***	1.088***	0.004	1.045***	0.325**	0.946***	-0.082***
	(0.0095)	(0.0477)	(0.0126)	(0.0123)	(0.0172)	(0.1275)	(0.0180)	(0.0177)

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Historical mean rainfall (1980-2017)	1.001***	0.006***	0.999***	-0.000**	1.000	-0.001	1.000	0.0001
	(0.0001)	(0.0004)	(0.0001)	(0.0001)	(0.0002)	(0.0012)	(0.0002)	(0.0001)
Livestock loss†	1.049	0.233	1.092	0.200*	0.828	-1.236	1.068	0.076
	(0.1015)	(0.4324)	(0.1255)	(0.1152)	(0.1573)	(1.4635)	(0.2098)	(0.1850)
Job loss†	1.424	1.287	0.949	-0.091	1.029	0.036	0.839	-0.461
	(0.3535)	(1.0344)	(0.2803)	(0.3127)	(0.5802)	(4.1448)	(0.4668)	(0.5914)
Models with additional controls								
Rainfall shortage 1-year lag	1.022	-0.098	0.795	0.150	2.378***	5.947***	0.924	-0.263
	(0.1575)	(0.6445)	(0.1764)	(0.1727)	(0.6970)	(2.0314)	(0.3049)	(0.2359)
Rainfall shortage 2-year lag	0.554***	-2.466***	2.038***	0.567***	1.192	1.199	0.898	-0.015
	(0.0453)	(0.3444)	(0.1976)	(0.0655)	(0.2420)	(1.4078)	(0.1944)	(0.1802)
Temperature shock 1-year lag	1.232***	0.890***	0.950	-0.065	1.242**	1.494**	0.900	-0.169*
	(0.0720)	(0.2521)	(0.0571)	(0.0555)	(0.1205)	(0.6784)	(0.1004)	(0.0877)
Temperature shock 2-year lag	0.989	-0.034	1.099	0.155**	1.447**	2.578**	0.871	-0.238**
	(0.0666)	(0.2941)	(0.0871)	(0.0752)	(0.2476)	(1.1892)	(0.1677)	(0.1124)
Historical mean temperature (1980-2017)	0.897***	-0.518***	1.105***	0.023*	1.023	0.123	0.957*	-0.104***
	(0.0106)	(0.0498)	(0.0146)	(0.0130)	(0.0200)	(0.1386)	(0.0228)	(0.0168)
Historical mean rainfall (1980-2017)	1.001***	0.005***	0.999***	-0.001***	1.000	-0.002	1.000	-0.001***
	(0.0001)	(0.0004)	(0.0001)	(0.0001)	(0.0002)	(0.0011)	(0.0002)	(0.0001)
Livestock loss†	1.101	0.397	1.038	0.059	0.895	-0.612	0.988	-0.006
	(0.1091)	(0.4029)	(0.1217)	(0.1019)	(0.1787)	(1.4122)	(0.2087)	(0.1680)
Job loss†	1.257	0.825	1.205	0.354*	1.044	0.370	0.934	-0.114
	(0.2886)	(0.8796)	(0.3290)	(0.2133)	(0.5927)	(3.9058)	(0.5454)	(0.4994)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable.

Table P: Impact of shocks on crop diversification decisions in rural Ethiopia

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Rainfall shortage 1-year lag	0.081	0.024	0.002	0.019
	(0.0530)	(0.0179)	(0.0482)	(0.0171)
Rainfall shortage 2-year lag	-0.070***	0.036***	-0.125***	0.003
	(0.0225)	(0.0086)	(0.0209)	(0.0081)
Temperature shock 1-year lag	0.0338**	0.0160***	0.044***	0.011*
	(0.01720)	(0.00611)	(0.0152)	(0.0059)
Temperature shock 2-year lag	0.1401***	0.0611***	0.035*	0.039***
	(0.02208)	(0.00947)	(0.0187)	(0.0082)
Historical mean temperature (1980-2017)	-0.0357***	-0.0155***	-0.018***	-0.010***
	(0.00269)	(0.00110)	(0.0025)	(0.0010)
Historical mean rainfall (1980-2017)	0.0005***	0.0002***	0.000***	0.000***
	(0.00003)	(0.00001)	(0.0000)	(0.0000)
Livestock loss†	-0.0362	-0.0358***	-0.061**	-0.040***
	(0.03210)	(0.01182)	(0.0276)	(0.0106)
Job loss†	0.0080	-0.0383	0.046	-0.037
	(0.09096)	(0.03911)	(0.0812)	(0.0389)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable.

Reproducing main Tables in the manuscript but now applying different estimation approaches:

a) *Seed use decisions- using Cragg Double Hurdle Models*
b) *Crop diversification decisions: Count index-Panel negative binomial regression; Simpson index-panel fractional probit regression*

Table Q: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia:Cragg Double Hurdle Models

	Improved seed		Local seed	
	Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2
Variables	AME	AME	AME	AME
Rainfall shortage 1-year lag	0.095***	0.206	-0.012	0.030
	(0.0295)	(0.1666)	(0.0102)	(0.0667)
Rainfall shortage 2-year lag	-0.080***	0.190**	0.021***	-0.034
	(0.0144)	(0.0815)	(0.0075)	(0.0305)
Temperature shock 1-year lag	0.007	0.073	-0.011***	-0.045**
	(0.0090)	(0.0587)	(0.0033)	(0.0208)
Temperature shock 2-year lag	0.046***	0.132	0.026***	-0.064**
	(0.0138)	(0.0878)	(0.0034)	(0.0271)
Historical mean temperature (1980-2017)	-0.004*	-0.137***	-0.001*	-0.105***
	(0.0023)	(0.0149)	(0.0007)	(0.0065)
Historical mean rainfall (1980-2017)	0.000***	-0.000	0.000	0.000*
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
Livestock loss†	-0.002	-0.177*	0.005	-0.048
	(0.0185)	(0.1063)	(0.0078)	(0.0388)
Job loss†	-0.013	-0.129	0.002	-0.063
	(0.0460)	(0.0949)	(0.0166)	(0.0848)

Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	1406	7194	6591

Notes: Cluster robust standard errors in parenthesis; * p < 0.10, ** p < 0.05, *** p < 0.01; AME=Average Marginal Effects

Table R: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia

	Improved maize		Local maize		Improved wheat		Local wheat	
	Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2	Hurdle 1	Hurdle 2
Variables	AME	AME	AME	AME	AME	AME	AME	AME
Rainfall shortage 1-year lag	0.001	-0.011	-0.042	0.018	0.141**	0.960**	0.000	-0.167
	(0.0381)	(0.2285)	(0.0308)	(0.0795)	(0.0554)	(0.3804)	(0.0520)	(0.1186)
Rainfall shortage 2-year lag	-0.120***	-0.085	0.118***	-0.020	0.022	0.481**	-0.011	-0.002
	(0.0187)	(0.1114)	(0.0177)	(0.0347)	(0.0339)	(0.1906)	(0.0294)	(0.0830)
Temperature shock 1-year lag	0.014	0.035	0.012	-0.037	0.022	0.219	-0.013	-0.052*
	(0.0128)	(0.0850)	(0.0116)	(0.0267)	(0.0182)	(0.1906)	(0.0160)	(0.0310)
Temperature shock 2-year lag	0.014	0.080	0.003	-0.048	0.077**	0.313	-0.020	-0.085*
	(0.0182)	(0.1238)	(0.0148)	(0.0360)	(0.0338)	(0.2613)	(0.0232)	(0.0453)
Historical mean temperature (1980-2017)	-0.021***	-0.117***	0.013***	-0.072***	0.003	-0.051	-0.006*	-0.088***
	(0.0037)	(0.0203)	(0.0029)	(0.0092)	(0.0040)	(0.0346)	(0.0035)	(0.0111)
Historical mean rainfall (1980-2017)	0.000***	0.000	-0.000***	0.000***	-0.000	-0.000	0.000	-0.001***
	(0.0000)	(0.0002)	(0.0000)	(0.0001)	(0.0000)	(0.0003)	(0.0000)	(0.0001)
Livestock loss†	0.023	-0.132	0.006	0.046	-0.023	0.013	-0.006	-0.028
	(0.0233)	(0.1022)	(0.0215)	(0.0432)	(0.0374)	(0.2987)	(0.0303)	(0.0664)
Job loss†	-0.016	-0.004	0.079	-0.111	0.018	0.087	-0.025	-0.046
	(0.0559)	(0.1285)	(0.0545)	(0.1204)	(0.0998)	(0.3040)	(0.0770)	(0.1068)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4479	1003	4479	3772	1937	223	1937	1754

Notes: Cluster robust standard errors in parenthesis; * p < 0.10, ** p < 0.05, *** p < 0.01; AME=Average Marginal Effects

Table S: Impact of shocks on crop diversification decisions in rural Ethiopia

	Crop count	Simpson index
Variables	negative binomial regression	Fractional probit regression
Rainfall shortage 1-year lag	-0.0676*	0.0178
	(0.038813)	(0.016015)
Rainfall shortage 2-year lag	-0.0052	0.0017
	(0.017339)	(0.007904)
Temperature shock 1-year lag	-0.0080	0.0023
	(0.010973)	(0.004548)
Temperature shock 2-year lag	-0.0026	0.0245***
	(0.016349)	(0.007555)
Historical mean temperature (1980-2017)	-0.0205***	-0.0080***
	(0.004210)	(0.001307)
Historical mean rainfall (1980-2017)	0.0004***	0.0002***
	(0.000044)	(0.000011)
Livestock loss†	-0.0034	-0.0337***
	(0.022830)	(0.009550)
Job loss†	0.0786	-0.0324
	(0.062988)	(0.031535)
Year dummies	Yes	Yes
Other controls	Yes	Yes
Observations	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; Reported are Average partial effects

Weighted statistics for main Outcome variables

Table T: Descriptive statistics of selected outcome variables used in the analysis-weighted using sampling weights given with Ethiopia Socioeconomic Survey(ESS)

Variable definitions	2012 mean	sd	2014 mean	sd	2016 mean	sd
All crops (N=2398)						
Improved seed use (1=yes; 0=otherwise)	0.21	0.41	0.26	0.44	0.26	0.44
Quantity of improved seeds used per household	20.84	75.32	21.26	61.31	24.36	81.49
Local seed use (1=yes; 0=otherwise)	0.98	0.14	0.97	0.17	0.99	0.12
Quantity of local seeds used per household	75.65	122.80	90.31	138.73	83.45	126.90
Grow maize (1=yes; 0=otherwise)	0.65	0.48	0.63	0.48	0.66	0.47
Grow wheat (1=yes; 0=otherwise)	0.32	0.47	0.33	0.47	0.33	0.47
Number of crops grown per household	9.20	4.68	8.91	4.49	8.75	4.61
Simpson index of crop diversity	0.76	0.17	0.74	0.18	0.73	0.19
Maize growers (N=1539)						
Improved Maize seed (1=yes; 0=otherwise)	0.24	0.43	0.32	0.47	0.33	0.47
Quantity of improved Maize seed used per household	23.34	80.54	26.06	65.13	28.27	83.44
Local Maize seed use (1=yes; 0=otherwise)	0.88	0.33	0.82	0.38	0.80	0.40
Quantity of local Maize seeds used per household	64.14	107.02	73.46	128.20	65.12	113.07
Wheat growers (N=628)						
Improved Wheat seed (1=yes; 0=otherwise)	0.11	0.31	0.11	0.32	0.09	0.28
Quantity of improved Wheat seed used per household	19.52	87.25	11.80	47.11	19.63	93.21
Local Wheat seed use (1=yes; 0=otherwise)	0.91	0.28	0.93	0.26	0.95	0.22
Quantity of local Wheat seeds used per household	138.89	162.61	165.19	193.51	165.47	165.32

Notes: Summary statistics are weighted, standard deviations (s.d) in parentheses.

Attrition and Attrition adjusted estimates

- *Here we provide results where we test and control for systematic attrition*
- *Our results show that systematic was systematic (*Table U*) but when we adjust for attrition bias by incuding inverse mills ratio from the probit model of atrition in all our seed use and diversification equations we do not alter our conclusins indicating that our estimates are robust to attrition bias(Table V to Table X).*

Table U:Probit estimation of attrition bias in Ethiopia.

	Attrition probit model
VARIABLES	*Drop out in 2014(1=yes)*
Female household head(1=yes)	0.2063*
	(0.1037)
Age of household head(years)	-0.0029
	(0.0027)
Household size (count)	-0.1573**
	(0.0607)
Household labor units	0.0559
	(0.0892)
Farm size(ha)	-0.0657
	(0.0486)
Household wealth index (PCA)	-0.0007
	(0.0120)
Distance to nearest market (Km)	0.0004
	(0.0008)
Number of plots	-0.0820***
	(0.0241)
Constant	-0.6989***
	(0.1849)
LR chi2(8)	89.06
Prob > chi2	0.0000
Observations	3466

Normal standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; The dependent variable is a dummy for household dropping out in follow up sample from baseline sample which is 2012 Survey.

Table V: Impact of shocks on household seed use decisions (all crop model) in rural Ethiopia-with Inverse mills ratio from attrition probit model

	Improved seed		Local seed	
	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)
Models without additional controls				
Rainfall shortage 1-year lag	3.718***	2.285***	0.618	0.383***
	(1.1755)	(0.5539)	(0.3522)	(0.0992)
Rainfall shortage 2-year lag	0.362***	-1.683***	4.891***	-0.013
	(0.0561)	(0.2721)	(2.3161)	(0.0429)
Temperature shock 1-year lag	1.024	0.100	0.399***	-0.216***
	(0.0989)	(0.1711)	(0.0780)	(0.0281)
Temperature shock 2-year lag	1.785***	1.034***	4.467***	0.125***
	(0.2708)	(0.2683)	(0.8302)	(0.0393)
Historical mean temperature (1980-2017)	0.963	-0.105**	0.936**	-0.083***
	(0.0223)	(0.0429)	(0.0304)	(0.0092)
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.001***
	(0.0002)	(0.0004)	(0.0004)	(0.0001)
Livestock loss†	0.882	-0.123	1.721	0.002
	(0.1761)	(0.3475)	(0.8048)	(0.0592)
Job loss†	1.022	0.142	0.630	0.043
	(0.5030)	(0.8581)	(0.5685)	(0.1626)
Inverse mills ratio (IMR)	1.947***	1.646***	8.844***	1.432***
	(0.3119)	(0.2934)	(2.8897)	(0.0683)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	No	No
Observations	7194	7194	7194	7194
Models with additional controls				
Rainfall shock 1-year lag	2.664***	1.681***	0.443	0.258***
	(0.8561)	(0.5501)	(0.2568)	(0.0931)
Rainfall shock 2-year lag	0.415***	-1.427***	3.872***	0.012
	(0.0655)	(0.2723)	(1.8486)	(0.0415)
Temperature shock 1-year lag	1.098	0.221	0.461***	-0.196***
	(0.1065)	(0.1701)	(0.0907)	(0.0269)
Temperature shock 2-year lag	1.673***	0.867***	4.405***	0.105***
	(0.2541)	(0.2647)	(0.8412)	(0.0374)
Historical mean temperature (1980-2017)	0.957*	-0.127***	0.937	-0.105***
	(0.0239)	(0.0449)	(0.0377)	(0.0078)
Historical mean rainfall (1980-2017)	1.003***	0.005***	1.000	0.000**
	(0.0003)	(0.0005)	(0.0004)	(0.0001)
Livestock loss†	0.964	0.048	1.359	-0.037
	(0.1929)	(0.3427)	(0.6480)	(0.0557)
Job loss†	0.874	-0.063	0.968	0.071
	(0.4308)	(0.8466)	(0.8901)	(0.1532)
Inverse mills ratio (IMR)	0.721	-0.564	3.405**	-0.093
	(0.1694)	(0.4181)	(1.9276)	(0.0772)
Year dummies	Yes	Yes	Yes	Yes
Other controls	Yes	Yes	Yes	Yes
Observations	7194	7194	7194	7194

Notes: Cluster robust standard errors in parenthesis; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties are first measured as dummy variables for use and then secondly as continuous variables indicating the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denotes dummy variable. Inverse mills ratio (IMR) is from the attrition probit model

Table W: Impact of shocks on Maize and Wheat seed use decisions in rural Ethiopia- with Inverse mills ratio from attrition probit model

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Models without additional controls								
Rainfall shortage 1-year lag	1.449	0.370	0.555	0.134	6.778***	6.833***	0.415	-0.377
	(0.6753)	(0.6095)	(0.2609)	(0.1683)	(4.1248)	(2.1975)	(0.3054)	(0.2912)
Rainfall shortage 2-year lag	0.196***	-2.172***	7.258***	0.477***	0.967	-0.232	1.017	-0.172
	(0.0457)	(0.3074)	(1.9488)	(0.0726)	(0.3726)	(1.3393)	(0.4507)	(0.1797)
Temperature shock 1-year lag	1.195	0.344	1.210	0.025	1.302	0.887	0.843	-0.069
	(0.1913)	(0.2110)	(0.2115)	(0.0547)	(0.2719)	(0.7347)	(0.2100)	(0.0948)
Temperature shock 2-year lag	1.354	0.496	1.001	0.033	2.797**	3.550***	0.580	-0.364***
	(0.3102)	(0.3129)	(0.2252)	(0.0739)	(1.1596)	(1.3611)	(0.2318)	(0.1254)
Historical mean temperature (1980-2017)	0.784***	-0.372***	1.191***	-0.002	1.089*	0.293*	0.881**	-0.081***
	(0.0334)	(0.0587)	(0.0472)	(0.0152)	(0.0478)	(0.1538)	(0.0454)	(0.0216)
Historical mean rainfall (1980-2017)	1.004***	0.006***	0.998***	-0.000	1.000	-0.001	1.000	-0.000
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0004)	(0.0002)
Livestock loss†	1.174	0.343	1.237	0.207*	0.677	-1.191	0.984	-0.101
	(0.3330)	(0.3696)	(0.3948)	(0.1081)	(0.2884)	(1.4528)	(0.4482)	(0.1790)
Job loss†	0.973	-0.032	2.093	0.393	1.112	0.086	0.637	-0.221
	(0.6591)	(0.8696)	(1.6447)	(0.2918)	(1.3236)	(4.2415)	(0.7698)	(0.5473)
Inverse mills ratio (IMR)	1.535*	1.016***	1.469	1.219***	0.710	-0.697	2.170***	1.146***
	(0.3717)	(0.3346)	(0.3454)	(0.0965)	(0.1620)	(0.7797)	(0.5952)	(0.1164)
Models with additional controls								
Rainfall shortage 1-year lag	0.932	-0.141	0.744	0.014	5.199**	5.280**	0.942	-0.197
	(0.4410)	(0.6045)	(0.3572)	(0.1641)	(3.5442)	(2.2128)	(0.8297)	(0.2746)
Rainfall shortage 2-year lag	0.224***	-1.949***	6.195***	0.439***	1.241	0.843	0.848	-0.054
	(0.0534)	(0.3048)	(1.6990)	(0.0724)	(0.5180)	(1.3465)	(0.4182)	(0.1734)

	Improved maize		Local maize		Improved wheat		Local wheat	
	Use	INT	Use	INT	Use	INT	Use	INT
	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)	(OR)	(APE)
Temperature shock 1-year lag	1.227	0.335	1.155	0.036	1.338	0.958	0.801	-0.126
	(0.1961)	(0.2071)	(0.1999)	(0.0539)	(0.2999)	(0.7274)	(0.2161)	(0.0917)
Temperature shock 2-year lag	1.210	0.302	1.007	-0.007	2.920**	3.148**	0.629	-0.201*
	(0.2756)	(0.3049)	(0.2264)	(0.0730)	(1.3299)	(1.3559)	(0.2754)	(0.1211)
Historical mean temperature (1980-2017)	0.768***	-0.410***	1.227***	0.004	1.034	0.076	0.908	-0.100***
	(0.0359)	(0.0626)	(0.0525)	(0.0158)	(0.0513)	(0.1583)	(0.0546)	(0.0202)
Historical mean rainfall (1980-2017)	1.004***	0.005***	0.998***	-0.000***	1.000	-0.002	1.000	-0.001***
	(0.0004)	(0.0005)	(0.0003)	(0.0001)	(0.0004)	(0.0013)	(0.0005)	(0.0002)
Livestock loss†	1.324	0.488	1.091	0.145	0.771	-0.577	0.898	-0.084
	(0.3783)	(0.3637)	(0.3494)	(0.1050)	(0.3547)	(1.4533)	(0.4536)	(0.1690)
Job loss†	0.843	-0.147	2.995	0.541*	1.229	0.455	0.667	-0.127
	(0.5758)	(0.8574)	(2.3810)	(0.2830)	(1.4874)	(3.9987)	(0.8386)	(0.5175)
Inverse mills ratio (IMR)	0.535*	-0.795*	3.960***	0.280**	0.462**	-2.340*	1.364	0.020
	(0.1910)	(0.4774)	(1.4639)	(0.1315)	(0.1801)	(1.2123)	(0.6209)	(0.1494)
Observations	4479	4479	4479	4479	1937	1937	1937	1937

Notes: In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; INT=intensity. APE=Average partial effects, OR=odds ratios. Improved and local varieties for both maize and wheat are first measured as dummy variables for use. Secondly, continuous variables indicate the intensity of use (kgs of seed used). We model use and intensity (INT) of use equations using Correlated Random Effects logit and Tobit, respectively, †denoted dummy variable. Inverse mills ratio (IMR) is from attrition probit model

Table X: Impact of shocks on crop diversification decisions in rural Ethiopia- with Inverse mills ratio from attrition probit model

	No additional covariates		With additional covariates	
	Crop Count	Simpson Index	Crop Count	Simpson Index
	(APE)	(APE)	(APE)	(APE)
Rainfall shortage 1-year lag	-0.0259	0.0065	-0.0560*	-0.0023
	(0.031980)	(0.013551)	(0.032079)	(0.013250)
Rainfall shortage 2-year lag	0.0413***	0.0079	0.0134	-0.0055
	(0.014401)	(0.005902)	(0.014682)	(0.005876)
Temperature shock 1-year lag	-0.0197**	-0.0046	-0.0062	-0.0041
	(0.009080)	(0.003782)	(0.009121)	(0.003716)
Temperature shock 2-year lag	-0.0103	0.0171***	-0.0220	0.0159***
	(0.013714)	(0.005365)	(0.013827)	(0.005277)
Historical mean temperature (1980-2017)	-0.0325***	-0.0161***	-0.0170***	-0.0102***
	(0.004441)	(0.001217)	(0.004385)	(0.001245)
Historical mean rainfall (1980-2017)	0.0005***	0.0002***	0.0004***	0.0002***
	(0.000046)	(0.000013)	(0.000045)	(0.000013)
Livestock loss†	0.0039	-0.0222***	0.0010	-0.0205**
	(0.019086)	(0.008165)	(0.019117)	(0.007965)
Job loss†	0.0942*	0.0032	0.0791	-0.0051
	(0.052757)	(0.022210)	(0.052657)	(0.021579)
Inverse mills ratio (IMR)	0.4799***	0.0718***	0.1733***	-0.0165
	(0.030109)	(0.008976)	(0.041010)	(0.012404)
Year dummies	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes
Observations	7194	7194	7194	7194

In parenthesis are cluster robust standard errors; * p < 0.10, ** p < 0.05, *** p < 0.01; APE=Average partial effects, we model crop count and Simpson diversity equations using Correlated Random Effects Poisson and Tobit, respectively, †denotes dummy variable. Inverse mills ratio (IMR) is from the attrition probit model

Quoted references

Budescu DV, Budescu M (2012) How to measure diversity when you must Psychological Methods 17:215

Gelaro R et al. (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2) Journal of Climate 30:5419-5454

Hirshman A (1964) The paternity of an index American Economic Review 54:761

Letta M, Montalbano P, Tol RS (2018) Temperature shocks, short-term growth and poverty thresholds: evidence from rural Tanzania World Development 112:13-32

Michler JD, Baylis K, Arends-Kuenning M, Mazvimavi K (2019) Conservation agriculture and climate resilience Journal of environmental economics and management 93:148-169

Rienecker MM et al. (2011) MERRA: NASA's modern-era retrospective analysis for research and applications Journal of climate 24:3624-3648

Simpson EH (1949) Measurement of diversity nature 163:688-688

Ward PS, Shively GE (2015) Migration and Land Rental as Responses to Income Shocks in Rural C hina Pacific Economic Review 20:511-543

Paper II

Rainfall shocks and inequality have heterogeneous effects on farmers' seed purchase decisions in East Africa

Clifton Makate[1*] , Arild Angelsen[1], Stein Terje Holden[1], & Ola Tveitereid Westengen[2]

[1*]School of Economics and Business, Norwegian University of Life Sciences, P.O.Box 5003, 1432 Ås, Norway; clifton.makate@nmbu.no (C.M); arild.angelsen@nmbu.no (A.A); stein.holden@nmbu.no (S.T.H)
[2]Department of International Environment and Development Studies, Noragric, Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway; ola.westengen@nmbu.no (O.T.W.)
* Corresponding author: clifton.makate@nmbu.no (C.M)

Abstract

Climate shocks and poverty worsen seed and food insecurity in smallholder farming. Here we use rich and representative household-level data for Ethiopia, Malawi, and Tanzania, combined with historical monthly weather data to examine the effects of drought exposure and wealth (asset endowment) inequality on seed purchasing. We find that between a third and half of the farmers purchase seed and among seed purchasers, more than half of the total seed volume is purchased. We model seed purchasing decisions using double-hurdle (Cragg) models and find that drought shocks experienced in the past season encourage seed purchasing in the following season. On average, drought shock exposure increases seed purchasing for both improved and local seeds in Malawi and Tanzania while encouraging (discouraging) local (improved) seed purchases in Ethiopia. In all three countries, farmers better endowed with household assets increase seed purchasing, particularly for improved seeds, after a drought shock exposure. In addition, smaller farm sizes and low asset wealth endowments in all study countries are significant deterrents for buying seeds in the market, particularly improved seeds. Policies need to support both formal and informal seed systems and address inequalities in access to seed from formal seed channels to achieve seed and food security under elevated climate risk.

Keywords: Adverse rainfall shocks; seed security; household socioeconomic inequality; sub-Sahara Africa

1 Introduction

Smallholder farmers in developing regions obtain crop seeds from both formal and informal seed sources (FAO 1998; Sperling et al. 2008). Farmers' own harvest, so called farm-saved seeds, is the main source, but farmers' seed systems are typically also open to new seeds from outside the farm (Bellon et al. 2006; Coomes et al. 2015). New seeds can harbor genetic traits and adaptations not available in the local varieties and therefore represent a major potential for adaptation to climate change, yield increase, and satisfying other agronomic and consumption preferences (Acevedo et al. 2020; Challinor et al. 2016; McGuire and Sperling 2013). Farmers can sometimes access new seeds free of charge from formal seed sources such as development programs and emergency aid and informal seed sources such as social networks. However, a substantial share of new seeds is paid for in cash, either at local markets or from formal agro-dealers (Sperling 2020). Local markets can supply seeds of old and new local varieties as well as new seeds from formal breeding programs, so-called improved varieties, while agro-dealers supply certified seed of improved varieties from the formal system.

Most seed system development efforts focus on strengthening the formal seed supply system. After decades of the limited effect of this formalization strategy in countries in sub-Saharan Africa, scholars and practitioners are increasingly calling for *integrated* approaches to seed system development that recognize and build on the complementary role of informal and formal elements of the seed systems farmers use (Louwaars and de Boef 2012; Sperling 2020). This paper addresses questions about seed purchasing decisions and how climate shocks, farm size, and asset wealth variation among smallholder households affect purchasing of local and improved varieties of key crops in Ethiopia, Malawi, and Tanzania.

When farmers purchase seeds, it affects their seed security. Seed security is often defined by the three parameters of *availability* (seed being available in space and time), *access* (physical and economic access), and *utilization* (seed quality meet user's needs and preferences) (McGuire and Sperling 2011; Sperling et al. 2008). Seed security is a condition which exists when rural households, particularly farmers and farming communities, have ready access to adequate quantities of quality seed and planting materials of preferred crop varieties, adapted to their agro-ecological conditions and socioeconomic needs, at planting time, under normal and abnormal weather conditions (FAO 1998; FAO and ECHA 2015).

Seed purchasing enables the farmer to respond to negative factors that result in chronic and temporary seed insecurity (e.g., varietal deterioration with time (quality) and depleted farmer saved seed stocks through droughts (availability)), and also enables them to exploit opportunities associated with accessing new seed (e.g., growing new crops and accessing drought-tolerant crop varieties (access)) (Almekinders et al. 1994; Almekinders et al. 2007; Nordhagen and Pascual 2013). In adapting their farming systems to production shocks, farmers may adopt new crop cultivars, diversify production, and switch to more shock-tolerant varieties (Holden and Quiggin 2017; Howden et al. 2007), which may increase their propensity to purchase seed. The covariate nature of rainfall shocks may, however, cause crop failure or low yields in a larger area, thereby limiting local seed supply available for farmers in the following year. Furthermore, intermittent erratic rains, which lead to crop failure, may disrupt farmer stocks of their own saved seed or make it hard to set aside seed from harvest due to urgent consumption needs, forcing farmers to source seed from elsewhere through trade (Bellon et al. 2011; Nordhagen and Pascual 2013).

This study, therefore, focuses on seed purchasing, which is vital for seed access, utilization, and hence seed security. In this study, we test the hypothesis that exposure to lagged drought shocks in prior seasons increases the likelihood and extent of purchasing seeds in the following season. In order to understand the interactions between climatic and farm size and other wealth factors influence seed purchasing, we furthermore address the questions: (a) How do household farm size, non-land asset wealth, and gender influence seed purchasing decisions? (b) Does the impact of drought shock exposure on seed purchasing decisions vary with the households' wealth? We use large and representative Living Standard Measurement Study (LSMS) data sets, available from the World Bank[1], and historical monthly weather data (rainfall and temperature) from WorldClim (Fick and Hijmans 2017; Masarie and Tans 1995).

We draw on the seed system and seed security literature coupled with economic theory on behavior under risk and uncertainty (e.g., the state-contingent theory of adaptation by Chambers and Quiggin (2000)) to study smallholder farmers' seed purchasing decisions under climate risk. For the empirical investigation, we employ Cragg's double-hurdle models (Cragg 1971) to assess the factors that drive seed purchasing decisions in smallholder farming. We focus on the potential influence of climate variables (long-term averages for rainfall and temperature), drought shocks (1-year lag), gender of primary decision-maker on input acquisition within the household, and household asset wealth endowments (household asset wealth index and farm size) on the probability and extent of purchasing (per unit farm size) the different seed types. We also assess whether drought shocks affect (i) relatively resource-poor versus resource-richer households (based on household asset wealth), (ii) or households with female versus with male prime decision-makers on input acquisition within the household differently with respect to seed purchasing decisions. We do this by estimating the effects of drought shocks on seed purchasing separately in poor vs. rich and male vs. female sub-samples.

The rest of this article is organized as follows: the next section gives a brief overview of seed systems in studied countries. Section 3 outlines the empirical approach, while section 4 presents the results. Section 5 discusses the results, while section 6 concludes the article.

2 Defining and describing the seed systems farmers use

2.1 Defining seed systems

When farmers purchase new seeds, they do so from a *seed system*. By the term seed system, we refer to the organizations, institutions, actors, and activities involved in the process of developing, producing, disseminating, and using seed (Almekinders et al. 1994; Louwaars and de Boef 2012; Tripp 1997). If the seeds purchased are officially *certified,* they are from *formal seed systems,* while uncertified seeds are from so-called *informal seed systems*. The difference between the two systems is not clear cut. It is not only seeds of traditional varieties that circulate in informal seed systems, but also a considerable share of seeds of varieties bred in formal breeding programs from which seeds are saved from own harvest, exchanged in social networks, or purchased in local markets (Coomes et al. 2015; McGuire and Sperling 2016; Sperling et al. 2020). The two key seed system functions upstream of seed distribution/sale are *breeding* of new varieties and *seed production* (the multiplication of true-to-type seeds). In the formal seed system, these functions are governed through a variety release procedure, intellectual property rights for plant varieties, and laws governing certification and sale of seeds. The legal regime embodied in such laws and regulations is complex, and the FAO has developed a *Voluntary guide for national seed policy formulation*, defining a national seed policy as "a statement of principles that guides government action and explains the roles of relevant stakeholders in the coordination, structure, functioning, and development of the seed system comprising both formal and informal sectors." (FAO 2015).

2.2 The main characteristics of seed systems of Tanzania, Malawi, and Ethiopia

We summarize the key characteristics of seed systems in the three study countries in Table 1. In all three countries, seed policies and regulations have recently undergone revisions and amendments aimed at facilitating the growth of the formal system. In Ethiopia, this is part of an explicit *pluralistic* seed policy which aims to also strengthen the informal and an *intermediate* seed sector. The latter involves farmer cooperatives and other entities producing and distributing seeds under less stringent quality control than fully certified seeds (Mulesa et al. 2021). The Tanzanian government has also recently opened for wider use of such an intermediate system, while Malawi so far has focused regulative efforts on the formal system only.

A key dimension of the political economy of seeds in which the three countries differ is the role of private companies vs. public entities in the formal seed system (Erenstein and Kassie 2018; Langyintuo et al. 2010; Westengen et al. 2019). Ethiopia's formal system is dominated by public entities in all functions, except for hybrid maize and vegetable seed, for which private companies have a substantial market share. In Malawi's formal seed system, private companies dominate throughout the value chain, while Tanzania is in an intermediary position with a strong public institutional presence in breeding and seed production as well as a liberalized seed market with many private companies. The differences in structure and functions of the seed systems are clearly linked to the cropping systems in the three countries. Maize is the dominant staple in Malawi and Tanzania. While maize is second in terms of area harvested in Ethiopia, the crop diversity is considerably higher. Maize is among the most important crops for the seed industry globally, and this is reflected both in the make-up of the formal seed systems in the three countries and in the adoption figures in the literature.

Table 1: Characteristics of formal and intermediate seed supply systems in Ethiopia, Tanzania, and Malawi.

Major crops (share of crop land)[2]	**Formal seed system distribution and sale**[3]	**Major policies and regulations**	**Intermediate seed system distribution and sale**	**Input subsidy or other relevant social protection program**	**Improved variety adoption rate estimates (share of crop land to crop)**[4]
Ethiopia					
Teff (20%) Maize (15%) Sorghum (15%) Wheat (11%) Barley (6%) Coffee (5%)	Strong public control of entire seed value chain for most crops. Private companies play an increasing role, especially in maize.	National Seed Policy (2020) Pluralistic Seed System Development Strategy (2017) Seed Proclamation (2013) -regulating formal system Co-operative Societies Proclamation (2004) -regulating intermediate system actor Ministerial QDS directive (2015) -regulating intermediate system seed quality	Seed Producer Cooperatives (SPC) producing and marketing local and improved varieties, sometimes quality declared seeds (QDS)[5,6]	Productive Safety Net Program (PSNP) providing input support[7]	Maize 27.9%; Wheat 22.2%; Barley 33.8%
Malawi					
Maize (37%) Groundnut (9%) Sweet potato (8%) Common beans (8%) Pigeon peas (6%) Cassava (5%)	Strong commercial maize seed system. Farmers buy seed directly from agro-dealer or local trader	National Seed Policy 2018 National Seed Bill/National Seed Commission Bill/ National Seed act drafted (2018), pending parliamentary endorsement (2022)	Community Seed Banks with "pass-on-system"[8] Seed Act of 2013 does not provide for QDS seed production[9]	Farming Input Subsidy Program (FISP) directly subsidizing improved seeds of maize and legumes 2005-currently	Maize 43%; Cassava 61%; Groundnut 58%
Tanzania					
Maize (22%) Rice (7%) Sunflower seed (6%) Cassava (6%) Groundnuts (6%) Cashew nuts (6%)	Farmers buy seeds directly from agro-dealers or local traders[10] Public seed enterprise (ASA) markets and distribute 'certified seed of crops and varieties for which private sector and other multiplies interest or ability is not strong enough to mee farmers demand' [11]	No overall National Seed Policy. Seed Act (2003) regulating formal system. QDS regulations under Seed Act (2020) regulating intermediate system	Local Seed Businesses with QDS production and sale.	National Agriculture Input Voucher (NAIVS) 2009-2016	Maize 35.4%; Groundnut 32.1%; Sweet potato 0%; Bean 45.8%; Pigeon Pea 49.8%

3 Methods and data

3.1 Modeling seed purchasing decisions

3.1.1 Theoretical framework

Farmers' seed purchasing decisions under climate risk and uncertainty can be analyzed within the state-contingent framework of Chambers and Quiggin (2000). Within this framework, smallholder farmers make input decisions before weather conditions are revealed (e.g., rainfall pattern in the season). Given the alternative outcome distributions under different states of nature (that have to be anticipated by the farmer, based on their past experiences and perceptions), production decisions are made to maximize the anticipated utility of the returns in different states of nature (Holden and Quiggin 2017). Climate risk and shocks, farmers' *perceptions* of those risks based on past experiences of shocks, as well as their risk preferences,[12] influence farmers' decisions to use purchased seeds. Often a farmer's decision to use purchased seeds may depend on her/his perceptions of risk associated with that choice relative to alternative seed use options (e.g., use of farm-saved seeds) and the different states of nature (good versus bad seasons) that are revealed after adoption decisions are made. In other instances, climate risk exposure or shocks may jeopardize livelihoods and intensify poverty (Dercon 2005; Dercon and Christiaensen 2011; Enfors and Gordon 2008) which can lead to desperate seed use practices (i.e., use of poor quality seed).

Utilizing drought shocks (local negative deviations from normal rainfall as a natural experiment, we test the hypothesis that exposure to drought shocks in prior seasons increases the likelihood and extent of purchasing seeds in the following season. The covariate nature of rainfall shocks may cause low yields or crop failure in a larger geographic area, thereby limiting local seed supply in the following year and enhancing the need to import seeds to such areas. Similarly, while farmers in more deprived regions depend heavily on farmer saved seeds, exposure to climate stress can change their seed sourcing decisions (Nordhagen and Pascual 2013). Climate shocks (e.g., lagged drought shocks) may increase the need for purchased seed due to depleted on-farm seed stocks (McGuire 2008; Nordhagen and Pascual 2013), or poor performance of farmer saved seed with continued rainfall shock exposure (Howden et al. 2007; Mortimore and Adams 2001). A good example is the adoption of drought-tolerant maize varieties in Malawi where exposure to drought shocks in combination with availability of subsidized drought-tolerant maize seeds stimulated a rapid adoption process (Holden and Quiggin 2017). On the contrary, exposure to more serious climate shocks might degrade farming outcomes and intensify poverty, limiting farmers' ability to source any seed off-farm through purchase. We, therefore, test the influence of a one-year lag drought shock and long-term climate (rainfall and temperature) on seed purchasing decisions.

In addition to climate variables (drought shock 1 year lag, long-term average rainfall and temperature), we also assess the possible influence of household asset wealth endowments and gender of prime decision-maker on input acquisition within the household on seed purchasing.

Income poverty or low asset endowments have always been pointed out among significant impediments to innovative practices adoption in smallholder farming in Africa (Crawford et al. 2003; Croppenstedt et al. 2003). We therefore test the hypothesis that asset wealth endowments positively relate to seed purchasing decisions. We also perform heterogeneity analysis by assessing whether drought shocks affect resource-poor versus resource-richer households differently concerning their seed purchasing decisions. Asset endowments are an important cushion against adverse shock exposure (Dercon 2005; Speranza et al. 2014), and we expect households better endowed with assets to be more able to and likely to purchase seeds when exposed to a drought shock. For example, in Ethiopia, it is common to sell livestock in response to shocks to mobilize cash to purchase food and farm inputs, while poor households without livestock may be forced to rent out their land instead (Gebregziabher and Holden 2011; Holden and Shiferaw 2004).

Major gender disparities in agricultural outcomes in SSA are reported in the literature (Aguilar et al. 2015; Peterman et al. 2014; Slavchevska 2015). We therefore, test the hypothesis that households with male decision-makers are more likely to use purchased seed compared to their counterparts (female decision-makers). This hypothesis is motivated by studies suggesting that women farmers or decision-makers are particularly at risk of increased marginalization when there is climate change-induced competition for resources (Eastin 2018).

3.1.2 Empirical framework

Smallholder farmers participate in the market for seed as buyers in a two-step process: first, they decide on whether to purchase seed or not, and second, they decide on the amount of purchase. As some farmers decide not to or are unable to purchase, it is important to use appropriate econometric approaches that deal with zeros (censoring) to obtain unbiased and consistent results (Cameron and Trivedi 2005; Humphreys 2013). We can model such farming households' seed purchasing decisions and identify the factors that explain them using various econometric methods as proposed in the latent variable models' literature (Cameron and Trivedi 2005; Heckman 1979; Wooldridge 2010). We, hence, model farmers' seed purchasing decisions using the latent variables approach.

We use Cragg's double-hurdle models (Cragg 1971), which allow variables to have different effects on the probability of purchasing and the intensity of purchase decisions. The first hurdle (of the double hurdle model) involves estimating a probit model that determines the probability that the farming household purchased seed, while the second hurdle involves estimating a truncated regression model to determine the intensity of purchase. This article estimates the double hurdle models for seed purchasing of local and improved seed varieties (for the household cropping portfolio in general) and for specific crops including maize in all the studied countries and then Sorghum, Pigeon pea, and Common bean for Ethiopia, Malawi, and Tanzania respectively.

We specify the first and second hurdles for seed purchasing as a function of our key variables of interest and other household control variables (explained below). Our key variables of interest include a 1-year lag of normalized negative rainfall deviations (drought shock), long-term averages for growing season rainfall and temperature, a dummy variable for households with a female primary decision-maker on input acquisition, and household asset wealth endowments (household asset wealth index (elaborated below), and farm size). We combine information on household ownership of durable non-land assets (e.g., agricultural equipment and machinery) and household dwelling characteristics common in each country to create the household asset wealth index, using Principal Components Analysis(PCA) (Filmer and Pritchett 2001). The household asset index is precisely defined as a weighted sum of given asset indicators, and it gives more weight to assets that are more unequally distributed across households and less weight to more common assets (e.g., those owned by all households). The first principal component from PCA is retained and taken as a proxy for household asset wealth. The resulting asset wealth index (or score) can take both negative and positive values and the increasing(decreasing) value of the index show higher(lower) relative households asset wealth endowments. For more details on the technical explanation of such asset-based household wealth indicators, readers can refer to McKenzie (2005). Other household control variables we consider include characteristics of the household head (age, marital status, education), household characteristics (household size, family labor, tropical livestock units, access to relief or subsidized inputs), survey year dummies, and regional dummies. We refer to the vector of climate variables and drought shocks as vector (S), household wealth variables as vector (W), female primary decision-maker on input acquisition within the household (F), proximity to main agricultural markets (D), and other control variables as vector(C). We hence specify our two hurdles of the Craggit model of seed purchasing as follows:

Probability of purchasing seed, binary probit model (First Hurdle):

$$\Pr(Q_i = 1) = \theta_0 + \theta_1 S_v + \theta_2 W_i + \theta_3 F_i + \theta_4 D_i + \theta_5 C_i + \varepsilon_i \qquad \text{(Eq1)}$$

The intensity of seed purchasing, truncated regression model (Second Hurdle):

$$W_i = \beta_0 + \beta_1 S_v + \beta_2 W_i + \beta_3 F_i + \beta_4 D_i + \beta_5 C_i + \epsilon_i \text{ if } Q_i = 1 \text{ , and 0 otherwise} \quad \text{(Eq2)}$$

Where Q_i is the dependent variable indicating whether (or not) the household i purchased seed (1=yes, 0 otherwise), W_i is the intensity of purchase for seed purchasers measured as the quantities (in kgs/ha) in Ethiopia and Malawi and value (in local currency) of purchased seed per hectare in Tanzania. In all our regression models, the intensity of purchase variables are log-transformed to reduce heteroscedasticity and make our data more normal. S_v is the vector of variables capturing long-term climate variables and drought shock in the farmer's village v. W_i is a vector of household wealth variables, while F_i represent a dummy variable for female primary decision maker on input acquisition within the household. D_i is the distance to the nearest main market; vector C_i is the vector of other household control variables we include in our models. i is the household identifier; θ, and β are parameters to be estimated. Lastly ε_i and ϵ_i are normally distributed error terms.

We also perform heterogeneity analysis by assessing whether drought shocks affect (i) relatively resource-poor versus resource-richer households (based on household asset wealth) or (ii) female-led vs. male-led households (in terms of making decisions on input acquisition within the household) households differently with respect to seed purchasing decisions. We do this by estimating the effect of the covariate lagged drought shock variable in (i) relatively rich vs. poor households (elaborated below) and (ii) male vs. female-led households sub-samples. The dummy variable for relatively high asset wealth endowments is defined from the continuous asset wealth index: we start by defining five quintile categories of asset wealth distribution (1=poorest; and 5=richest) from the continuous asset wealth index and then define a dummy variable for high asset endowments equal to 1 for households belonging to the fourth and fifth quintiles (two richest quintiles) and zero otherwise.

3.2 Data

3.2.1 Household data

The study uses pooled data from multiple rounds of rich and representative cross-sectional household survey data sets from available agricultural household surveys from the respective countries Ethiopia (Ethiopia Socioeconomic Survey (ESS)), Malawi (Integrated Household Survey (IHS)), and Tanzania (Tanzania National Panel Survey (TNPS)). For Ethiopia, we use a combined sample of 4 987 rural households from the two latest cross-sectional rounds of the ESS data (2 873 rural households from ESS3 of 2015/16 and 2 114 rural households from ESS4 of 2018/19). In addition, we use pooled data from three rounds of the Malawi IHS (3rd 2010/11, 4th 2016/17, and 5th 2019/2020 rounds), making a total of 26 627 households (9 467, 8 862, and 8 298 from IHS3, 4, and 5 respectively). Additionally, we use three rounds of the TNPS data making a pooled sample of 6 665 households (2 214, 2 709, and 1 742 from TNPS 2, 3, and 4, respectively). The survey data are available through the Living Standards Measurement Study-Integrated surveys on agriculture (LSMS-ISA) program of the World Bank in collaboration with national governments. The LSMS-ISA data collect comprehensive information on agricultural activities and household socioeconomic conditions in respective countries. This study specifically uses data from rural households who engaged in agricultural activity with complete and usable information on seed use and seed purchasing.

3.2.2 Weather data and definitions of climate variables and shocks

In addition to the LSMS-ISA data, we also use historical monthly weather data from WorldClim[13] (Fick and Hijmans 2017; Masarie and Tans 1995) to define historical climate variables (precipitation and maximum temperature) and lagged drought shocks. The LSMS-ISA household data provide the approximate location (longitude and latitude) of clusters (villages) from which interviewed households reside. We used the georeferenced data to extract historical monthly climate data for 38 years (1980-2018) that we combine with household data for analysis. We used WorldClim data at the spatial resolution of 2.5 minutes (approximately ~21 km2).

We start by defining the main crop growing season in the respective countries (May to September for Ethiopia, November to April for Tanzania and Malawi) and then we define our variables of interest for the main crop growing season. We define climate variables for the main growing season to better reflect on conditions during the most important season for food production. We aim to test the effects of a growing season lagged drought shock (1 year lag) and long-term average rainfall, average rainfall variability, and temperature on seed purchasing decisions.

We define long-term climate variables (both rainfall and temperature) as averages and their standard deviations for the period 1980 to 2018. The distribution of average rainfall (1980-2018) for the growing season for the studied countries is shown in Figure 1.

We follow related literature (e.g., Michler et al. (2019), Ward and Shively (2015), and Letta et al. (2018)) and define lagged drought shocks as normalized (negative standard) deviations in a single season's rainfall from the seasonal climate variable over the reference period (1980-2018). Thus, we define rainfall shocks as normalized rainfall deviations (Z-scores): $Rain\_Zscore_{vt} = \left[\frac{rain_{vt} - \overline{rain}_v}{\sigma_{rain_v}}\right]$, where $Rain\_Zscore_{vt}$ is a rainfall shock measure for a village (v) in the year (t), $rain_{vt}$ is the observed amount of rainfall for the defined period (growing season), $\overline{rain}_v$ is the average seasonal rainfall for the village (v) over the reference period (1980-2018), and σ_{rain_v} is the standard deviation of rainfall during the same period.

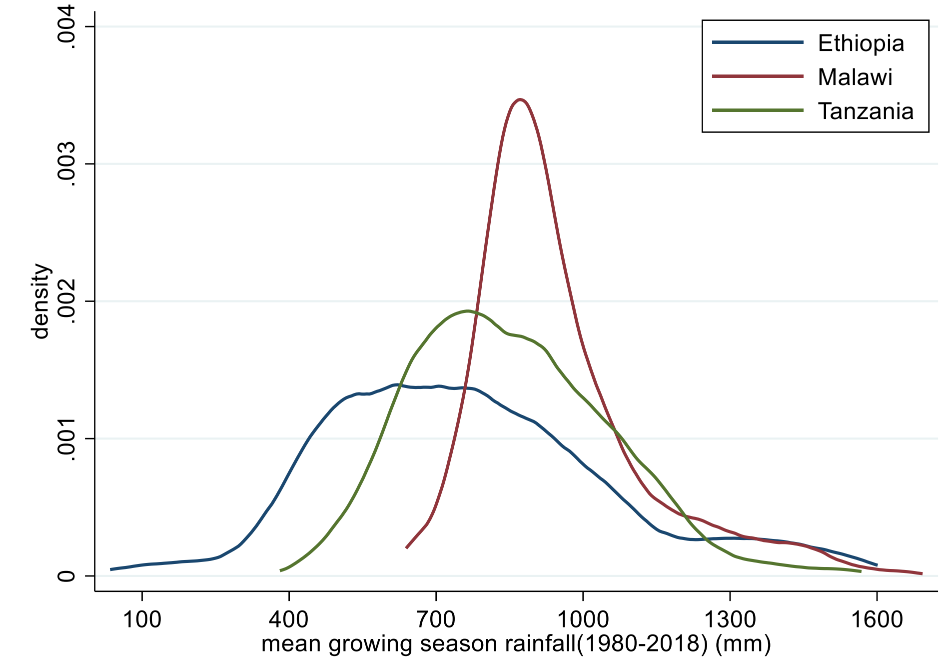

Figure 1: Distribution of historical mean rainfall of the main crop growing season in the reference period (1980-2017) in the studied countries.

We then define a lagged drought shock ($\underline{DS}_{vt}$) (which shows the intensity of the rainfall shortfall from the long-term average) as follows: $\underline{DS}_{vt} = \left\{ \left[\frac{rain_{vt} - \overline{rain}_v}{\sigma_{rain_v}} \right] if \ \ rain_{vt} < \overline{rain}_v, and\ 0\ otherwise. \right.$

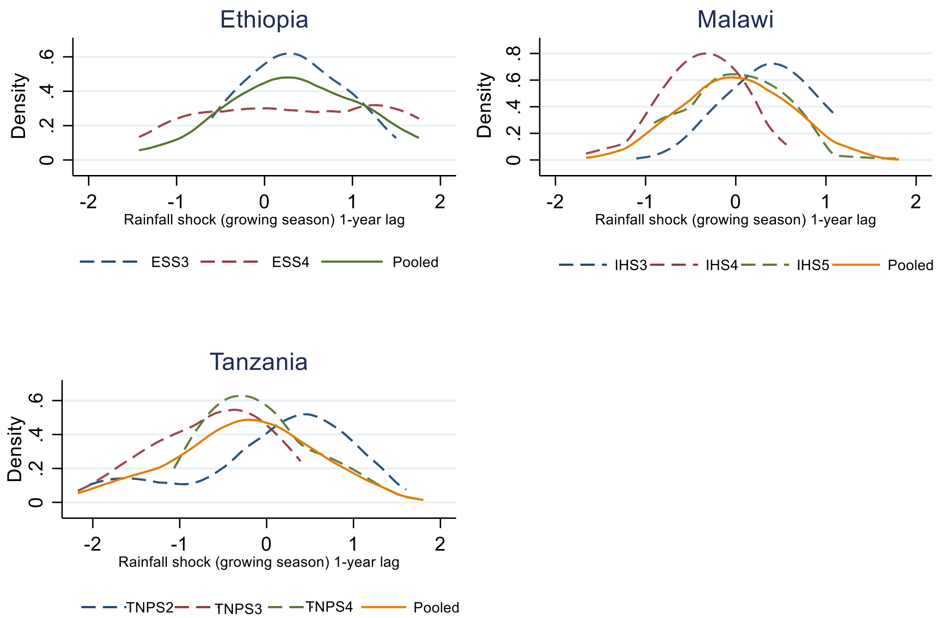

Figure 2: The distribution of the rainfall shock variable (Z-scores) for the growing season (GS) in pooled samples and by survey year for studied countries; ESS=Ethiopia Socioeconomic survey; IHS=Malawi Integrated Household Survey; TNPS=Tanzania National Panel Survey.

3.3 Potential study limitations

As much as our study gives important insights on the crucial relationships between socioeconomic inequality, climate shocks, and seed purchasing, we rely on farmer self-reported data, which could be associated with recall and response biases. It is for example important to note that the classification of varieties as either local or improved is somewhat ambiguous due to the extensive on farm seed saving or 'recycling' of improved varieties. Respondents may after a while refer to such farm-saved improved seeds as local varieties (Westengen et al. 2014). However, this is not a major caveat for the current study as our focus here is on seed purchasing and not on improved variety adoption *per-se*. Also, our data only allow us to understand the impact of past shock exposure and vulnerability on current farmer actions (ex-post), not what they do before exposure (ex-ante risk management). However, by studying the impacts of past exposure on current farmer practices, we can shed light on future responses to shock exposure by farmers in coping with them, which could be helpful for adaptation policy. Despite the noted concerns, the strength of our data is that it is large and representative of smallholder farmers, covering many rural households in the respective countries studied.

4 Results

4.1 Descriptive statistics

We first measure seed purchasing as dummy variables equal to one for farmers who have purchased seed in the reference seasons and zero otherwise. On average across survey years, approximately 48%, 46%, and 39% used purchased seeds in Ethiopia, Malawi, and Tanzania (*Table 2*). Assessing trends over time (from one survey to another), we see that the proportion of farmers using purchased seeds have increased overtime from 47 to 49% in the Ethiopian sample (ESS3 to ESS4), from 40% to 51% in the Malawian sample (from IHS3 to IHS5), and from 33% to 47% in the Tanzanian sample (from TNPS2 to TNPS4) (*Table 2*). In terms of quantity, purchased seeds constitute, on average, a quarter of total seeds used in Ethiopia and nearly a third in Malawi and Tanzania (Figure 3). Purchased seed is the main source of seed used among the seed purchasers, constituting about 53%, 60%, and 62% of this group's seed use in Ethiopia, Malawi, and Tanzania (Figure 3).

Table 2: Descriptive statistics of main variables used in the analysis

	Ethiopia			**Malawi**				**Tanzania**			
Variable	Pooled mean	ESS3 mean	ESS4 mean	Pooled mean	IHS3 mean	IHS4 mean	IHS5 mean	Pooled mean	TNPS2 mean	TNPS3 mean	TNPS4 mean
Seed purchasing variables											
Incidence of seed purchasing†	0.476	0.466	0.490	0.460	0.395	0.481	0.513	0.389	0.332	0.383	0.471
Incidence of seed purchasing (improved)†	0.175	0.173	0.177	0.276	0.246	0.284	0.300	0.258	0.107	0.383	0.256
Incidence of seed purchasing (local)†	0.393	0.383	0.408	0.285	0.227	0.295	0.342	0.284	0.263	.	0.310
Quantity purchased(kg/ha) [**unconditional**]	21.692	20.876	22.758	10.631	6.541	11.225	14.796	7.239	.	3.939	12.371
Quantity purchased(kg/ha) [**conditional**]	44.297	42.843	46.176	22.853	16.536	23.322	27.937	16.63	.	9.01	26.28
Value of seed purchased (local currency/ha) [**unconditional**]	327.281	309.143	350.992	4179.16	735.95	5596.25	6673.59	11076.8	6319.13	10391.1	18604.4
Value of seed purchased (local currency/ha) [**conditional**]	668.342	634.448	712.148	8984.78	1860.04	11626.81	12600.61	27405.32	18542.28	25788.77	37796.62
Share purchased seed (%) [**unconditional**]	25.297	23.379	27.903	27.668	19.776	32.313	31.711	16.499	.	8.280	29.278
Share purchased seed (%) [**conditional**]	53.118	50.163	56.937	60.086	50.044	67.141	61.842	39.524	.	21.61	62.199
Socioeconomic variables											
Female (Female decision maker) †	0.210	0.221	0.195	0.290	0.253	0.303	0.320	0.237	0.225	0.232	0.259
Farm size (ha)	1.126	1.311	0.874	0.656	0.735	0.591	0.637	2.165	2.116	2.277	2.056
Rich (Household is in the top 40% of the sample asset wealth index[14] distribution(1=yes)) †	0.388	0.282	0.533	0.324	0.324	0.312	0.337	0.290	0.353	0.338	0.135
Distance to market(km)	64.1	66.4	61.0	22.015	17.000	24.810	24.751	12.270	13.092	12.799	10.403
Climate variables and shocks											
Historical rainfall growing season (1980-2017)	769.017	763.881	775.996	949.372	957.738	950.474	938.651	852.134	864.499	856.894	828.977
Historical temperature growing season (1980-2017)	25.877	25.810	25.968	28.198	28.173	28.233	28.189	28.461	28.581	28.497	28.253
Rainfall shock growing season (1-year lag)	0.339	0.352	0.321	-0.001	0.382	-0.378	-0.035	-0.257	0.053	-0.633	-0.067
Drought shock growing season (1-year lag)	-0.429	-0.183	-0.705	-0.441	-0.205	-0.447	-0.547	-0.697	-0.884	-0.761	-0.425
Observations	4987	2873	2114	26627	9467	8862	8298	6665	2214	2709	1742

Notes: Climate variables and shocks are shown for the main rainy season of respective countries; Statistics are not weighted; source (own calculation from LSMS-ISA data); †denotes dummy variable; conditional and unconditional refer to stats defined for seed purchasers and full sample respectively; Share purchased seeds is a percentage computed as the proportion of quantities of seed purchased to total seeds used for each crop grown and averaged for all crops grown by the household; ESS=Ethiopia Socioeconomic Survey, TNPS=Tanzania National Panel Survey, IHS=Malawi Integrated Household Survey.

Figure 4 shows the share of purchased seeds for maize only. Improved seeds dominate total maize seed purchases, and purchasing is the main maize seed source among purchasers (Figure 4). For the second crop considered for each country (bean for Tanzania, pigeon pea for Malawi, and sorghum for Ethiopia), seed purchases were mainly local seed varieties.

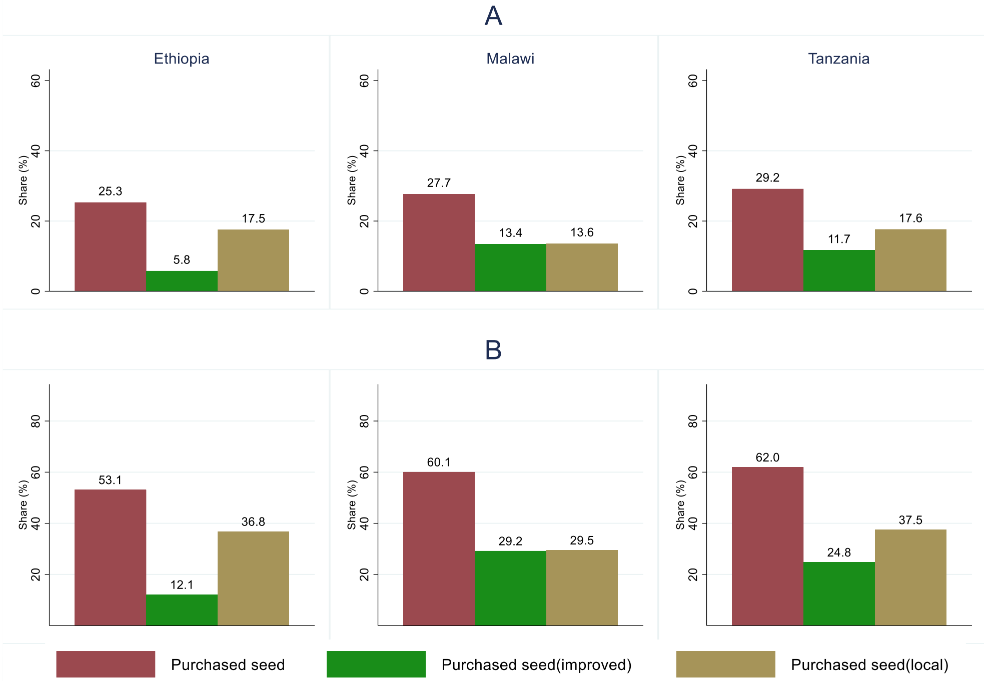

Figure 3: Share of purchased seed as a percentage of the total amount of seed used averaged for all crops grown. Panels A & B shows the unconditional (purchaser+non-purchasers) and conditional statistics (purchasers only) respectively. The figures are made from the pooled data except for Tanzania where figures are based on the latest survey (TNPS4), which permits computing shares of purchased seeds by variety type.

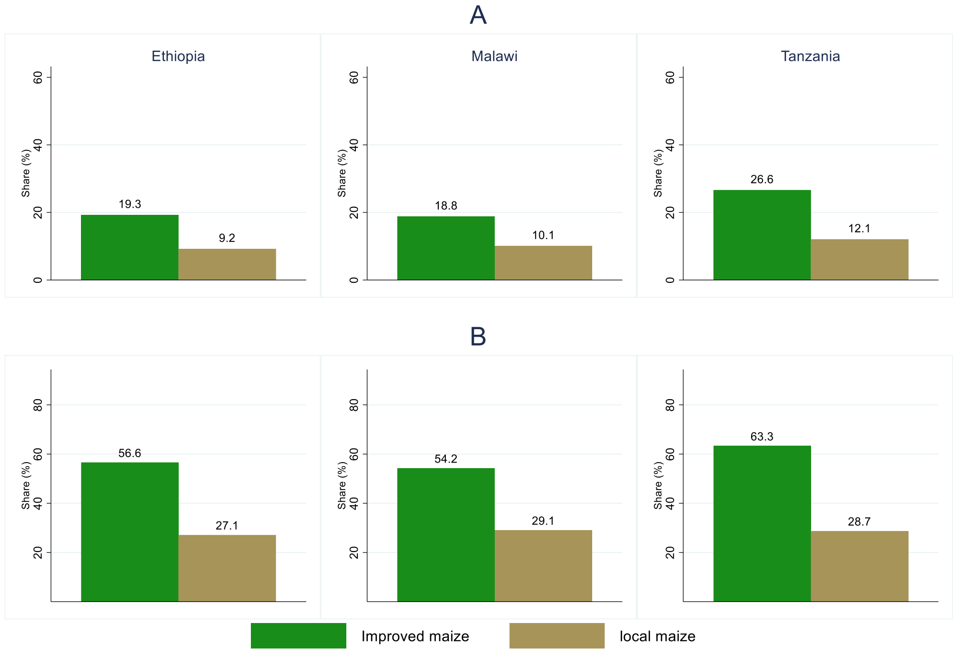

Figure 4:Share of purchased maize seed as a percentage of the total amount of maize seed used; Panels A & B shows the unconditional (purchasers+non-purchasers) and conditional statistics (purchasers only) respectively; The figures are made from the pooled data except for Tanzania where figures are based on the latest survey (TNPS4), which permits computing shares of purchased seeds by variety type.

Women are the primary decision-maker in between 21-29% of the household in the studied countries. Average farm sizes vary greatly among countries, from 0.7 ha in Malawi to 1.1 and 2.2 ha in Ethiopia and Tanzania. We show the distribution of the household wealth index in the respective countries in the attached supplementary material (Figure A). Also, farmers in Ethiopia, on average, travel much longer distances to the nearest main market (64 km) compared to Malawi (22 km) and Tanzania (12 km) (*Table 2*). Descriptive statistics for the other control variables are not shown here for brevity but are available in the supplementary material (Table III).

4.2 Results

This sub-section presents results from the estimated Cragg's double hurdle models for seed purchasing. We present summarized results on the main effects of our key test variables on seed purchasing decisions in Ethiopia, Malawi, and Tanzania in Table 3, Table 4, and Table 5. Full tables showing the full spectrum of explanatory variables used are shown in the supplementary material (Tables IV-V).

4.2.1 Impact of drought shocks

The results show that in Ethiopia (Table 3), a drought shock (1-year lag) positively influences the probability of purchasing seed, local seed, and local maize and sorghum. Also, the intensity of purchasing local maize and sorghum for purchasers increases with previous drought shock exposure in Ethiopia. On the contrary, the probability and intensity of purchasing improved seeds and the likelihood of purchasing improved maize decrease with drought shock exposure (Table 3).

In Malawi (Table 4), a drought shock (1-year lag) increases the probability of purchasing seed (in general) and notably improved seeds. Also, the intensity of local pigeon pea seed purchases increases with previous drought shock exposure in Malawi. In addition, the intensity of purchase for improved seeds, improved maize, and local maize reduces with previous drought exposure.

Results for Tanzania show a 1-year lag drought shock exposure to increase the likelihood of seed purchasing. The likelihood of purchasing seed (in general), improved seeds, and local maize seed increase with previous drought shock exposure (Table 5). Additionally, the intensity of local bean seed purchases increases with drought shock exposure (Table 5).

Overall, the results for the three countries seem to suggest that drought shocks experienced in the recent past encourage participation by farmers in the seed market as buyers (and for different types of seeds) in the following seasons in studied countries. In Ethiopia, drought shock exposure encourages mainly local seed purchases, while in Malawi and Tanzania, drought shock exposure promotes both local and improved seed purchases.

4.2.2 Household asset wealth and seed purchasing decisions

The descriptive statistics in (Figure 5) show that relatively wealthier households have higher average shares for purchased seeds, especially for improved seeds in studied countries. The regression results conform with the descriptive statistics, as asset wealth is a significant factor in explaining seed purchasing decisions in all studied countries (Table 3, Table 4, and Table 5). In Ethiopia, seed purchasing decisions are positively associated with the household asset wealth index (Table 3). In Malawi, the asset wealth index also positively associates with the probability of purchasing seed (in general), improved seeds, improved maize, and the intensity of purchasing local pigeon pea seed. Besides, the asset wealth index negatively associates with the likelihood of purchasing local seed, local maize, and local pigeon pea in Malawi (Table 4). In the Tanzanian sample, the probability and intensity of purchasing seed (in general), improved seed, and improved maize increase with household asset wealth. However, the chances of purchasing local maize reduce, with household wealth in Tanzania.

We also control for farm size as an additional proxy for household wealth and a measure of space available for the farmer to carry out her/his farming activities. We also find chances and intensity of using purchased seed increasing with farm size in studied countries (Table 3, Table 4, and Table 5). For example, in Ethiopia, farm size positively influences both the probability and intensity of seed purchases (Table 3).

Table 3: Double Hurdle model Estimates of seed purchase decisions in Ethiopia: Impact of climate shocks and other socioeconomic factors

	All purchased seed		Improved seed		Local seed		Improved Maize		Local Maize		Local Sorghum	
VARIABLES	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
Drought shock (growing season) 1-year lag	0.148**	-0.142	-0.466***	-0.414**	0.255***	-0.006	-0.589***	-0.084	0.333**	0.502**	0.561***	0.531***
	(0.0723)	(0.1327)	(0.0943)	(0.1694)	(0.0723)	(0.1666)	(0.1380)	(0.1251)	(0.1302)	(0.2095)	(0.1558)	(0.1852)
Rainfall (GS) mm (1980-2018) (log)	0.334***	-0.158	0.725***	-0.340*	0.207***	-0.133	1.180***	-0.147	-0.280***	-0.655***	0.016	-0.461**
	(0.0633)	(0.1120)	(0.1107)	(0.1796)	(0.0616)	(0.1324)	(0.1845)	(0.1160)	(0.0957)	(0.1670)	(0.1407)	(0.2056)
Temperature (GS) °C (1980-2018)	-0.022***	-0.032***	0.024***	-0.034***	-0.031***	-0.032***	0.005	0.035***	-0.043***	0.019	-0.011	0.073***
	(0.0037)	(0.0067)	(0.0045)	(0.0120)	(0.0037)	(0.0077)	(0.0068)	(0.0077)	(0.0070)	(0.0130)	(0.0086)	(0.0144)
Female decision maker(1=yes)	-0.037	-0.134	-0.061	-0.230	-0.053	-0.031	-0.068	-0.079	0.104	0.134	0.277*	-0.053
	(0.0668)	(0.1160)	(0.0829)	(0.1686)	(0.0676)	(0.1357)	(0.1131)	(0.1112)	(0.1174)	(0.1867)	(0.1450)	(0.1639)
Household asset wealth index	0.038***	0.010	0.032*	0.053*	0.031**	-0.004	0.039*	0.046**	-0.017	0.040	-0.002	0.088*
	(0.0134)	(0.0262)	(0.0181)	(0.0296)	(0.0137)	(0.0335)	(0.0234)	(0.0188)	(0.0208)	(0.0382)	(0.0256)	(0.0508)
Log of Farm size(ha)	0.176***	0.973***	0.168***	0.837***	0.136***	0.982***	0.041	0.327***	0.035	0.958***	-0.127	0.766***
	(0.0518)	(0.0938)	(0.0581)	(0.1261)	(0.0513)	(0.1143)	(0.0767)	(0.0830)	(0.0868)	(0.1799)	(0.1049)	(0.1830)
Distance to market (km, log)	-0.055*	0.070	-0.122***	0.006	-0.012	0.175***	-0.212***	0.044	-0.024	0.016	-0.095	-0.088
	(0.0286)	(0.0506)	(0.0380)	(0.0684)	(0.0285)	(0.0633)	(0.0506)	(0.0432)	(0.0494)	(0.0807)	(0.0623)	(0.0709)
Sigma constant		1.322***		1.031***		1.420***		0.656***		0.936***		0.746***
		(0.0253)		(0.0389)		(0.0314)		(0.0358)		(0.0493)		(0.0501)
Other household controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,987	2,375	4,987	873	4,987	1,962	2,923	683	2,923	350	1,978	254

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability purchasing seed(1=yes,0Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha)., All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for maize and sorghum; Full tables showing the full spectrum of household control variables is given in supplementary material (Table III), GS=growing season (Meher season).

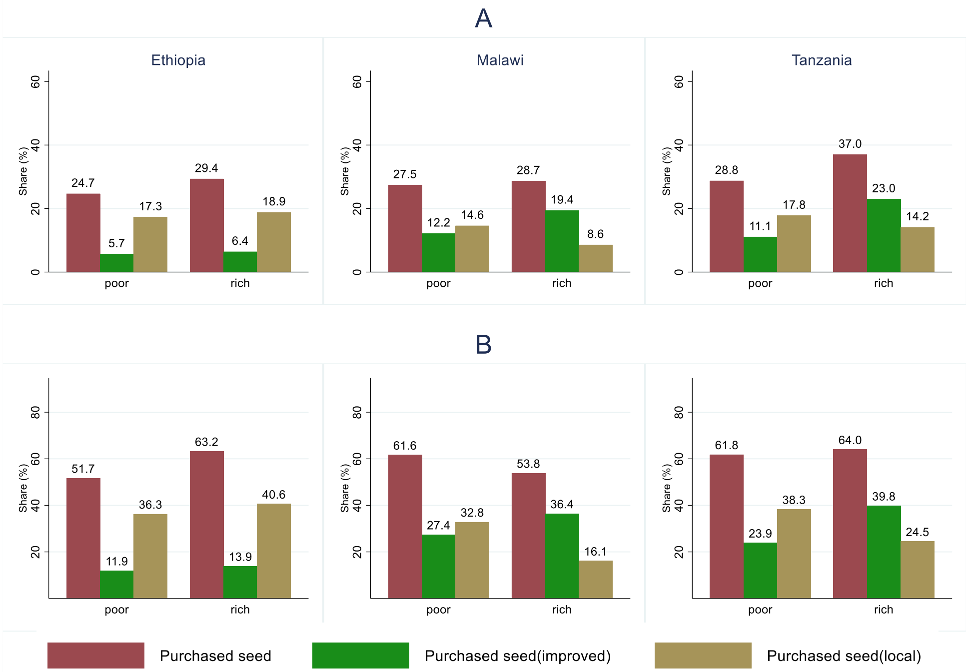

Figure 5:Average share of purchased seed by asset wealth status in studied countries. Panels A & B shows the unconditional and conditional statistics, respectively. The figures' stats for Tanzania are based on the latest survey (TNPS4), where we can compute shares of purchased seeds by variety type; Rich and poor are as described in Table 2.

In Malawi, farm size also positively associates with the probability and intensity of purchasing seed (in general), improved seed, local seed, and improved maize. Additionally, farm size reduces the likelihood of purchasing local maize and local pigeon pea but positively associates with purchase intensity. Similarly, in Tanzania, farm size is positively associated with the probability and intensity of purchasing seed (in general, improved and local seed) but negatively associated with the likelihood of purchasing local maize and beans (Table 5). The implication is that farmers with relatively larger farm sizes are more likely to participate in the seed market as buyers in studied countries.

Table 4: Double Hurdle model Estimates of seed purchase decisions in Malawi: Impact of climate shocks and other socioeconomic factors

	All purchased seed		Improved seed (all crops)		Local seed (all crops)		Improved Maize		Local Maize		Local pigeon pea	
	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
Drought shock (GS) 1-year lag	0.106***	-0.058	0.083**	-0.108**	0.002	0.034	0.058	-0.056*	-0.035	-0.168***	-0.008	0.241*
	(0.0340)	(0.0387)	(0.0357)	(0.0428)	(0.0393)	(0.0631)	(0.0385)	(0.0298)	(0.0544)	(0.0626)	(0.1588)	(0.1413)
Rainfall (GS) mm (1980-2018) (log)	-0.330***	0.055	-0.236***	0.418***	-0.560***	-0.254**	0.048	-0.133**	-0.347***	-0.110	-0.321**	0.412***
	(0.0632)	(0.0678)	(0.0681)	(0.0774)	(0.0701)	(0.1091)	(0.0740)	(0.0594)	(0.0912)	(0.0944)	(0.1492)	(0.1217)
Temperature (GS) °C (1980-2018)	0.015***	0.017***	0.003	-0.009***	0.003	0.023***	0.012***	-0.009***	0.036***	-0.002	0.058***	0.010
	(0.0028)	(0.0031)	(0.0030)	(0.0036)	(0.0031)	(0.0050)	(0.0033)	(0.0029)	(0.0042)	(0.0047)	(0.0086)	(0.0073)
Female decision maker(1=yes)	-0.121***	-0.068***	-0.208***	-0.014	-0.019	-0.060**	-0.223***	-0.025	-0.005	0.015	-0.118***	-0.022
	(0.0188)	(0.0202)	(0.0207)	(0.0251)	(0.0202)	(0.0280)	(0.0222)	(0.0204)	(0.0256)	(0.0259)	(0.0349)	(0.0275)
Household asset wealth index	0.047***	0.035***	0.078***	0.057***	-0.031***	-0.011	0.104***	0.036***	-0.120***	0.025	-0.053**	0.026**
	(0.0080)	(0.0071)	(0.0084)	(0.0086)	(0.0085)	(0.0130)	(0.0089)	(0.0066)	(0.0217)	(0.0172)	(0.0207)	(0.0131)
Log of farmsize (ha)	0.218***	0.711***	0.289***	0.461***	0.126***	0.957***	0.191***	0.616***	-0.163***	0.920***	-0.317***	0.811***
	(0.0322)	(0.0396)	(0.0338)	(0.0469)	(0.0334)	(0.0584)	(0.0350)	(0.0378)	(0.0465)	(0.0584)	(0.0735)	(0.0708)
Distance to nearest ADMARC (km)	-0.002*	-0.006***	0.000	-0.001	-0.005***	-0.005**	-0.002	0.000	-0.003	0.000	-0.002	-0.001
	(0.0013)	(0.0015)	(0.0014)	(0.0017)	(0.0015)	(0.0023)	(0.0015)	(0.0013)	(0.0020)	(0.0020)	(0.0031)	(0.0025)
Sigma constant		0.886***		0.826***		0.955***		0.575***		0.592***		0.561***
		(0.0081)		(0.0104)		(0.0107)		(0.0077)		(0.0100)		(0.0136)
Other household controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,627	11,261	26,627	7,337	26,627	7,597	24,945	5,894	24,945	2,943	7,468	2,453

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** $p<0.01$, ** $p<0.05$, * $p<0.1$; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha). All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for maize and local pigeon pea; Full tables showing the full spectrum of control variables is given in supplementary material (Table IV).

4.2.3 *Gender of decision-maker on input acquisition and seed purchasing decisions*

We also assessed the influence of gender of the main decision-maker on input acquisition within the household on seed purchasing decisions in the three countries. In Ethiopia, having a female prime decision-maker on input acquisition within the household seems not to significantly influence seed purchase decisions except for enhancing chances of purchasing local sorghum seed (Table 3). However, in Malawi, results show that having a female primary decision-maker on input acquisition within the household reduces the chances of purchasing seed (in general), improved seeds, improved maize, and local pigeon pea (Table 4). In Tanzania, having a female primary decision-maker significantly reduces the intensity of seed purchases, particularly for local seed and local maize (Table 5). We plot average shares of purchased seeds by gender of primary decision-maker on input acquisition in Figure 6.

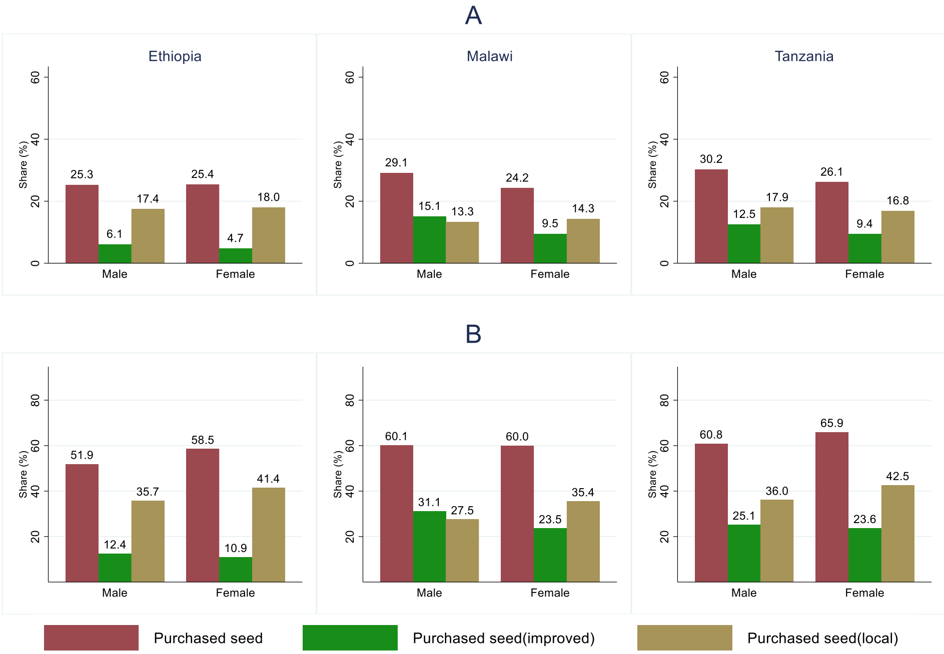

Figure 6: Average shares of purchased seed by male vs. female decision-makers in studied countries. Panels A & B shows the unconditional and conditional statistics, respectively. Stats for Tanzania in the figure are based on the latest survey (TNPS4), where we can compute shares of purchased seeds by variety type, Male and female decision-makers are as defined in Table 2.

Table 5: Double Hurdle model Estimates of seed purchase decisions in Tanzania: Impact of climate shocks and other socioeconomic factors

	All purchased seed		Improved seed(all crops)		Local seed(all crops)		Improved Maize		Local Maize		Local bean	
	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
Drought shock (GS) 1-year lag	0.103***	0.001	0.117***	-0.072	0.049	0.114	0.076	-0.101	0.336***	0.106	-0.126	0.562***
	(0.0306)	(0.0445)	(0.0341)	(0.0684)	(0.0449)	(0.0715)	(0.0467)	(0.0798)	(0.0697)	(0.0969)	(0.1312)	(0.1594)
Rainfall (GS) mm (1980-2018)(log)	-0.743***	-0.810***	-0.815***	-0.863***	-0.511***	-0.642***	-0.751***	-0.688***	-0.337**	-1.054***	-0.649***	-0.443
	(0.0768)	(0.1083)	(0.0853)	(0.1601)	(0.1040)	(0.1577)	(0.1021)	(0.1741)	(0.1426)	(0.2308)	(0.2047)	(0.2824)
Temperature (GS) °C (1980-2018)	-0.009	-0.088***	-0.010	-0.097***	0.001	-0.097***	-0.026**	-0.086***	0.071***	-0.058**	0.021	-0.076**
	(0.0074)	(0.0108)	(0.0082)	(0.0161)	(0.0103)	(0.0154)	(0.0106)	(0.0175)	(0.0173)	(0.0238)	(0.0255)	(0.0357)
Female decision maker(1=yes)	-0.050	-0.116**	-0.037	-0.049	0.011	-0.238***	-0.081	-0.077	-0.145*	-0.374***	-0.021	-0.167
	(0.0401)	(0.0549)	(0.0445)	(0.0777)	(0.0537)	(0.0792)	(0.0560)	(0.0808)	(0.0786)	(0.1100)	(0.1116)	(0.1432)
Household asset wealth index	0.026***	0.100***	0.062***	0.094***	-0.019	0.057***	0.114***	0.079***	-0.108***	0.011	0.001	0.052
	(0.0082)	(0.0111)	(0.0089)	(0.0140)	(0.0119)	(0.0187)	(0.0116)	(0.0141)	(0.0213)	(0.0356)	(0.0236)	(0.0337)
Log farm size(ha)	0.079***	0.319***	0.081***	0.339***	0.057*	0.245***	-0.144***	0.371***	-0.204***	0.313***	-0.141*	0.240**
	(0.0252)	(0.0372)	(0.0274)	(0.0518)	(0.0348)	(0.0529)	(0.0352)	(0.0539)	(0.0554)	(0.0854)	(0.0775)	(0.0976)
Distance to the nearest market(km)	0.001	0.003***	-0.001	0.004*	0.002*	0.001	-0.001	0.006***	-0.003	0.000	-0.003	0.003
	(0.0010)	(0.0012)	(0.0012)	(0.0022)	(0.0012)	(0.0018)	(0.0016)	(0.0023)	(0.0022)	(0.0040)	(0.0032)	(0.0062)
Sigma constant		1.110***		1.025***		1.070***		0.944***		0.910***		1.013***
		(0.0190)		(0.0233)		(0.0233)		(0.0238)		(0.0347)		(0.0355)
Other household controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,665	2,593	6,665	1,722	3,956	1,122	4,306	1,112	2,561	383	908	336

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (value of purchased seed/ha)). All crops model is estimated by defining seed purchasing variables averaged for all crops grown by the household. Then variables are defined for maize and bean crops, Equations for local varieties are estimated for two surveys (TNPS2 and 4) where seed purchasing variables were defined by seed variety type; Full tables showing the full spectrum of control variables is given in supplementary material (Table V).

4.2.4 Proximity to markets

We also test the influence of proximity to main input and output markets on seed purchasing decisions in studied countries. Results show distance to the nearest main market to negatively associate with seed purchasing decisions (in general) and for improved seed and maize in Ethiopia (Table 3). Farmers further away from main agricultural markets are less likely to participate in the market for seed as buyers in Ethiopia. Distance to the nearest main market is negatively associated with seed purchasing decisions in general for Malawi and does not significantly deter seed purchasing decisions in Tanzania (Table 4 and Table 5).

4.2.5 Heterogeneity effects: Drought shocks and socioeconomic inequality

In addition to the main effects of drought shock exposure, we also attempted to explore the heterogeneity of drought shock effects on seed purchasing in different household socioeconomics groups (relatively rich vs. relatively poor households based on asset wealth index, and female vs. male-led households in decision making on inputs acquisition within the household). Our primary interest is to test whether the impact of a drought shock is the same for farmers in different strata of socioeconomic status regarding seed purchasing.

From the heterogeneity analysis, we learn that our measure for drought shock enhances seed purchasing in general and for local seeds in the group of relatively richer households in Ethiopia while significantly reducing improved seeds purchase and enhancing local maize seed purchasing in poorer households (Table 6). In Malawi and Tanzania, we also learned that drought shock exposure significantly enhances seed purchasing in general and particularly for improved seeds in the group of relatively wealthier households (Table 7 and Table 8). On the contrary, drought shock exposure in the relatively less affluent group of smallholder farmers reduces the intensity of improved seed purchases and enhances local maize seed purchases. Therefore, wealthier households are more likely to source their seed through purchase following drought shock exposure than their poorer counterparts in studied countries. In Figure 7 we plot local polynomial regressions that summarize the raw relationships between seed purchasing intensity and rainfall shocks for relatively richer vs. poorer households. The plots support the gaps in the influence of drought shocks on seed purchasing between poor and richer households in studied countries.

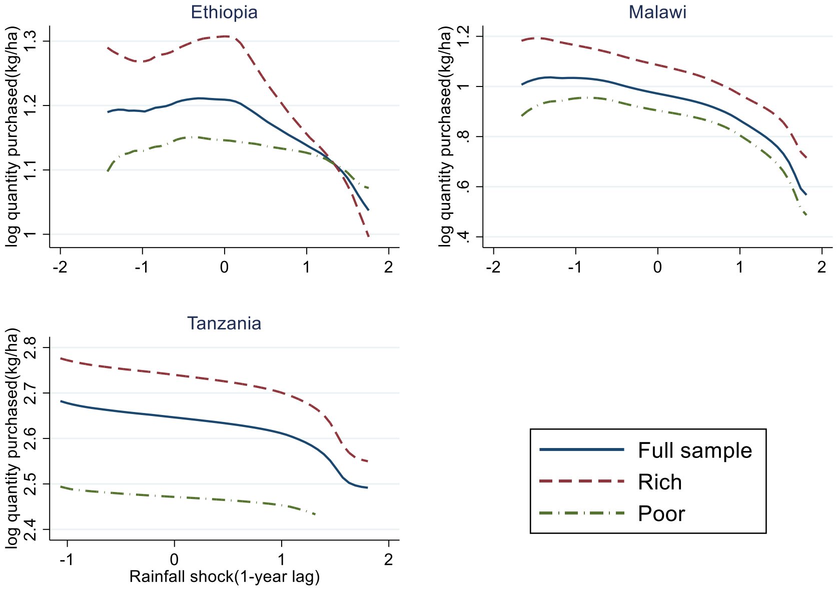

Figure 7: The distribution of seed purchasing intensity by rainfall shock exposure by relatively more affluent vs. relatively poorer households and in the full sample. The figure plots local polynomial regressions of the intensity of seed purchasing on the one-year lag of rainfall shock. Thus, the graph shows the raw relationship between the intensity of seed purchasing and a 1-year lag of rainfall shock for poor versus rich households in studied countries.

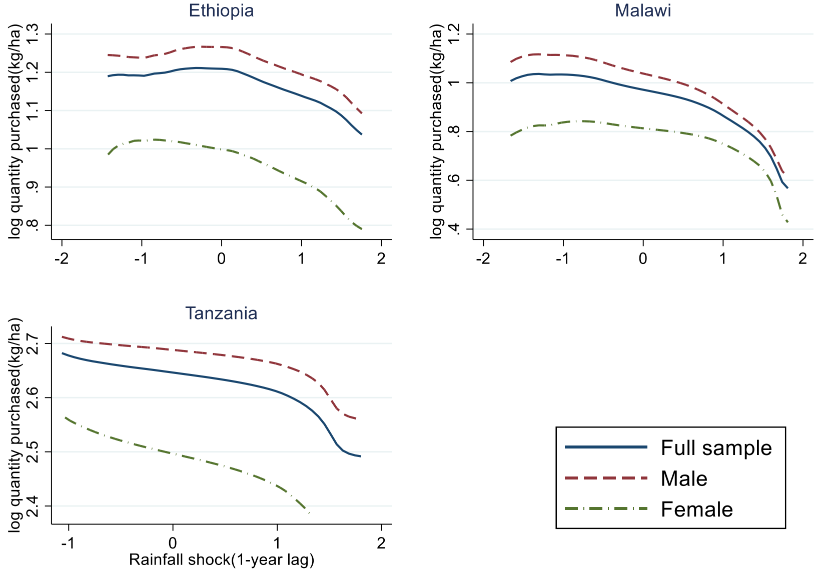

Figure 8:The distribution of seed purchasing intensity by rainfall shock exposure by male-led vs. female-led households (wrt making decisions on input sourcing) and in the full sample. The figure plots local polynomial regressions of the intensity of seed purchasing on the one-year lag of rainfall shock. Thus, the graph shows the raw relationship between the intensity of seed purchasing and a 1-year lag of rainfall shock for male-led and female-led households on input sourcing decisions in studied countries.

Table 6: Double Hurdle model Estimates of seed purchase decisions in Ethiopia: Heterogeneity effects

	All purchased seed		Improved seed		Local seed		Improved Maize		Local Maize		Local Sorghum	
	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
A-resource-poor versus resource-richer households (based on the asset wealth index)												
Rich sub-sample												
Drought shock (GS) 1-year lag	0.329***	-0.271	-0.169	-0.216	0.386***	-0.163	-0.361**	-0.090	0.016	0.403	0.714***	-0.201
	(0.1047)	(0.2033)	(0.1197)	(0.2270)	(0.1054)	(0.2765)	(0.1726)	(0.1748)	(0.1858)	(0.3767)	(0.2622)	(0.4605)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2090	920	2090	393	2090	690	1197	299	1197	135	823	85
Poor sub-sample												
Drought shock (GS) 1-year lag	-0.018	0.063	-0.877***	-0.276	0.140	0.186	-0.943***	0.119	0.524***	0.631***	0.427*	0.481**
	(0.1079)	(0.1748)	(0.1584)	(0.2911)	(0.1072)	(0.2040)	(0.2265)	(0.2097)	(0.1877)	(0.2272)	(0.2203)	(0.2019)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2897	1222	2897	427	2897	990	1726	359	1726	183	1155	156
B-Male versus female-led households on input acquisition (decision making)												
Male sub-sample												
Drought shock (GS) 1-year lag	0.190**	-0.173	-0.453***	-0.371**	0.302***	-0.041	-0.606***	-0.050	0.509***	0.562**	0.581***	0.600***
	(0.0821)	(0.1450)	(0.1038)	(0.1779)	(0.0816)	(0.1858)	(0.1517)	(0.1397)	(0.1486)	(0.2273)	(0.1704)	(0.2004)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3940	1749	3940	695	3940	1371	2345	558	2345	258	1652	184
Female sub-sample												
Drought shock (GS) 1-year lag	-0.022	0.037	-0.528**	-0.267	0.022	0.172	-0.478	-0.148	-0.322	0.139	0.684*	0.429
	(0.1577)	(0.2838)	(0.2321)	(0.3905)	(0.1598)	(0.3261)	(0.3412)	(0.2332)	(0.2984)	(0.4582)	(0.3989)	(0.4697)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1047	393	1047	125	1047	309	488	100	578	60	326	57

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** $p<0.01$, ** $p<0.05$, * $p<0.1$; Hurdle 1 is the probability purchasing seed(1=yes,0Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha)., All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for maize and sorghum.

Table 7: Double Hurdle model Estimates of seed purchase decisions in Malawi: Heterogeneity effects

	All purchased seed		Improved seed		Local seed		Improved Maize		Local Maize		Local pigeon pea	
	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
A-Resource-poor versus resource-richer households (based on the asset wealth index)												
Rich sub-sample												
Drought shock (GS) 1-year lag	0.178***	-0.095*	0.163***	-0.096*	-0.046	0.015	0.102*	0.005	0.028	-0.198	-0.219	0.411
	(0.0504)	(0.0563)	(0.0515)	(0.0579)	(0.0598)	(0.1074)	(0.0546)	(0.0380)	(0.0926)	(0.1332)	(0.2639)	(0.2577)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9867	4580	9867	3259	9867	2368	9432	2733	9432	607	2520	685
Poor sub-sample												
Drought shock (GS) 1-year lag	0.043	-0.025	0.002	-0.118*	0.033	0.057	0.005	-0.139***	-0.071	-0.160**	0.127	0.199
	(0.0466)	(0.0528)	(0.0499)	(0.0622)	(0.0520)	(0.0767)	(0.0549)	(0.0478)	(0.0666)	(0.0727)	(0.2017)	(0.1609)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16413	6689	16413	3435	16413	4678	15439	2661	15439	2107	4905	1613
B-Male versus female-led households on input acquisition (decision making)												
Male sub-sample												
Drought shock (GS) 1-year lag	0.141***	-0.069	0.119***	-0.141***	0.020	0.043	0.071	-0.061*	-0.015	-0.189**	-0.023	0.335*
	(0.0392)	(0.0444)	(0.0404)	(0.0482)	(0.0455)	(0.0750)	(0.0436)	(0.0326)	(0.0654)	(0.0785)	(0.1853)	(0.1776)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18629	8357	18629	5306	18629	4941	17573	4321	17573	1795	4793	1556
Female sub-sample												
Drought shock (GS) 1-year lag	0.003	-0.037	-0.047	0.052	-0.044	-0.023	-0.008	-0.023	-0.075	-0.118	0.055	-0.010
	(0.0698)	(0.0758)	(0.0767)	(0.0885)	(0.0794)	(0.1142)	(0.0827)	(0.0738)	(0.1001)	(0.1043)	(0.3130)	(0.2222)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7651	2912	7651	1388	7651	2105	7298	1073	7298	919	2635	742

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha). All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for maize and local pigeon pea.

Table 8: Double Hurdle model Estimates of seed purchase decisions in Tanzania: Heterogeneity effects

	All purchased seed		Improved seed		Local seed		Improved Maize		Local Maize		Local bean	
	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
A-resource-poor versus resource-richer households (based on the asset wealth index)												
Rich sub-sample												
Drought shock (GS) 1-year lag	0.138***	0.069	0.172***	-0.019	0.033	0.240**	0.114*	0.040	0.248**	0.235*	-0.049	0.574***
	(0.0427)	(0.0587)	(0.0462)	(0.0860)	(0.0647)	(0.0969)	(0.0603)	(0.0999)	(0.0995)	(0.1345)	(0.1555)	(0.1839)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3859	1686	3859	796	2303	722	2564	614	1540	220	560	233
Poor sub-sample												
Drought shock (GS) 1-year lag	0.082*	-0.110	0.055	-0.162	0.095	-0.017	0.022	-0.355***	0.443***	-0.028	-0.229	0.377
	(0.0448)	(0.0694)	(0.0515)	(0.1130)	(0.0652)	(0.1123)	(0.0741)	(0.1294)	(0.1051)	(0.1340)	(0.2459)	(0.3019)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2763	884	2763	332	1626	385	1740	233	1020	154	348	102
B-Male versus female-led households on input acquisition (decision making)												
Male sub-sample												
Drought shock (GS) 1-year lag	0.098***	0.014	0.114***	-0.031	0.035	0.106	0.045	-0.070	0.320***	0.132	-0.025	0.554***
	(0.0342)	(0.0503)	(0.0378)	(0.0768)	(0.0504)	(0.0830)	(0.0520)	(0.0934)	(0.0797)	(0.1132)	(0.1416)	(0.1863)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5052	2037	5052	921	2984	860	3300	691	1951	290	677	252
Female sub-sample												
Drought shock (GS) 1-year lag	0.124*	-0.048	0.129	-0.182	0.101	0.302**	0.204	-0.212	0.399***	0.109	-0.735*	0.693**
	(0.0692)	(0.0985)	(0.0797)	(0.1387)	(0.0997)	(0.1417)	(0.1086)	(0.1445)	(0.1457)	(0.1724)	(0.3835)	(0.2863)
All baseline controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1570	533	1553	207	945	247	993	156	609	84	231	83

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** $p<0.01$, ** $p<0.05$, * $p<0.1$; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (value of purchased seed/ha)). All crops model is estimated by defining seed purchasing variables averaged for all crops grown by the household. Then variables are defined for maize and bean crops, Equations for local varieties are estimated for two surveys (TNPS2 and 4) where seed purchasing variables were defined by seed variety type

In the bottom panels of Table 6, Table 7, and Table 8, we show results from heterogeneity analysis by gender of the main decision-maker on input sourcing within the household for Ethiopia, Malawi, and Tanzania, respectively. We seek to determine whether previous drought shock exposure has heterogeneous effects for male-led and female-led households regarding their seed purchasing decisions. The results establish that drought shock exposure significantly enhances seed purchasing decisions in the male sub-sample compared to the female sub-sample in all the studied countries. In Ethiopia, previous exposure to a drought shock enhance seed purchasing in general and for local seeds in the male sub-sample, while the same drought shock reduces improved seed purchase and enhances local sorghum purchase in the female sub-sample (Table 6). In Malawi, previous exposure to a drought shock enhance seed purchasing in general and for improved seeds in the male sub-sample and does not explain seed purchasing in the female sub-sample (Table 7). Additionally, in Tanzania, we learn that previous drought shock exposure enhances seed purchasing for both local and improved seeds in the male sub-sample and mainly enhances local seed purchases in the female sub-sample in Tanzania (Table 8). In Figure 8, we show plots from local polynomial regressions that summarize the raw relationships between seed purchasing intensity and rainfall shocks for male-led and female-led households (concerning making decisions on input sourcing). The plots (Figure 8) support the gaps in the influence of drought shocks on seed purchasing between female-led and male-led households in studied countries.

5 Discussion

Our results show that seed purchasing is a common practice in studied countries and that for seed purchasers, more than half of the total seeds used on the farm are purchased. Seed purchasing can allow the farmer to overcome challenges associated with complete reliance on farm-saved seed and to take advantage of new varieties (Almekinders et al. 1994; Almekinders et al. 2007; Nordhagen and Pascual 2013). As such, seed purchasing means enhanced access (both physical and economic access) to a larger diversity of seed, one of the most important factors of seed security (FAO 1998; FAO and ECHA 2015; Sperling 2020). Results further reveal interesting associations between drought shocks and socioeconomic inequality on seed access through purchasing, which we discuss more in-depth in the following sub-sections.

5.1 Impact of drought shocks

Exposure to drought shocks in the previous season, on average, encourages seed purchasing decisions in studied countries. In Ethiopia, drought shock exposure encourages mainly local seed purchases whilst in Malawi and Tanzania, drought shock exposure promotes both local and improved seed purchases. The results support our hypotheses. More purchased seed use after shock exposure is in accordance with the state-contingent theory of technology adoption that reflects that farming households learn from shocks and may become willing to adopt technologies that help them deal with future shocks (Holden and Quiggin 2017). Also, past exposure to drought shocks reduces the possibility of farmers saving enough seed for the following seasons from their own harvest; hence, with depleted on-farm seed stocks, farmers may be forced to purchase off-farm seed (Howden et al. 2007; McGuire 2008; Nordhagen and Pascual 2013).

At the same time, it is also possible that drought shocks, through their effects on household economies, lead to less resource allocation to purchasing of seed. This is evident in some of our findings. For instance, in Ethiopia, previous drought shocks discourage seed purchasing for improved seeds, and for purchasers in Malawi, drought shock exposure discourages the intensity of purchase for improved seeds. This finding also follows literature that reveals that shocks might significantly disrupt agricultural-based livelihoods, which may increase hunger and poverty, promoting the use of inferior technologies that may render them inefficient and more vulnerable (Dercon 2005; Kubik and Maurel 2016; Yesuf and Bluffstone 2009).

The contrasting effects of shocks on the type of seeds purchased in studied countries could also reflect important differences in seed systems (Table 1) and other factors that might have implications for seed availability, preferences, and accessibility to farmers for purchasing in good and bad seasons. For instance, in Malawi and Tanzania, the use of formal seed systems in accessing seed (improved seed) by farmers has accelerated more rapidly in the recent past than in other countries such as Ethiopia (Crawford et al. 2003; Jayne and Rashid 2013; Sheahan and Barrett 2017). This phenomenon is partly attributed to successful revitalized government support programs such as FISP in Malawi and the National Agriculture Input Voucher (NAIVS) program in Tanzania. Such programs have increased awareness, availability, and access to improved seeds and enhanced input market development in Sub-Sahara Africa, including in countries such as Malawi and Tanzania (Jayne and Rashid 2013). For instance, in Malawi, Katengeza et al. (2019) found that exposure to drought shocks combined with the provision of subsidized seeds after shock exposure leads to higher uptake of improved (drought-tolerant) varieties. Hence, access to seed from off-farm sources can generate experiences on the performance of different crop varieties under shocks, which may increase their propensity to purchase seed.

However, the enhanced(reduced) likelihood of purchasing local(improved) seeds post-drought shock exposure in Ethiopia could also reflect farmer perceptions of local and improved varieties and the availability of local and improved seeds for purchasing in bad and good seasons. Farmer perceptions are important in influencing crop variety choice (Tripp 1996; Tripp 1997), and if farmers view local varieties as better adapted to erratic rainfall and drought shocks, they are more likely to purchase local varieties post-drought shock exposure, which could explain the findings in Ethiopia. Also, given the dominance of the informal seed system and local variety use in Ethiopia (Atilaw and Korbu 2011; Wale 2012), local crop varieties could be more readily accessible for purchase post-shock exposure which could explain why chances of purchasing local varieties are high post-exposure in Ethiopia. The pluralistic seed system strategy implemented in Ethiopia in recent years is designed to enhance local availability of farmer preferred varieties, be they of formal or informal origin (Mulesa et al. 2021).

5.2 Association of gender, asset wealth, and other socioeconomic variables with seed purchasing

In addition to drought shock exposure, our results have shown the importance of other socioeconomic factors in explaining seed purchasing decisions in studied countries. For instance, the female decision-maker dummy is associated with lower chances of purchasing seed particularly improved seeds in Malawi, while in Tanzania, the female dummy is negatively associated with the intensity of seed purchasing for purchasers. On the one hand, the results possibly suggest a gender gap favoring male decision-makers purchasing seeds (particularly improved seeds). Women farmers, because of their underlying challenges, including low resource endowments among other inequalities in accessing agricultural training and markets (UN-Women 2015; World Bank and Campaign 2014), often have more inadequate access to new technologies compared to men. On the other hand, the result could partly reflect on unique preferences between men and women decision-makers on seed sourcing and seed type to use. Women as custodians of household food security tend to prefer local varieties for food crops because of their better culinary traits (taste, ease of processing, ease of storage, etc.) (Lunduka et al. 2012). The overall implication is that the gender coefficient we find could be due to both endowment and structural factors. Analyzing gender-disaggregated data on seed purchasing for specific crops could allow for further exploration of the gender gap in seed purchasing decisions and the relative contributions of differences in endowment factors and structural factors on the gap. We leave such an investigation for further studies.

As expected, wealthier households are more likely to purchase improved seeds and less likely to purchase local seeds. Improved seeds usually fetch a higher price on the market when compared to local seeds. Wealthier farmers are, hence, more likely to afford improved seeds compared to poorer farmers. This view is plausible because seed cost is a significant barrier to purchasing modern seeds through formal channels (Gemeda et al. 2001; Louwaars 2005). Also, farmers with larger farm sizes are more likely to use purchased seeds in studied countries. Land is an important resource for farmers, which directly influences the space available for the farmer to carry out her/his farming activities and access to complimentary farming inputs (e.g., access to credit). Farmers with larger land sizes are therefore more likely to diversify their seed sources by using purchased seed.

Overall, results conform to the literature showing that access to productive assets is central to inspiring market participation by smallholder farmers and subsequent escape from poverty traps (Barrett 2008). However, the social safety-net function of the input subsidy programs seems not to counter-balance the inequality in such access.

5.3 Heterogeneity in the impact of drought shocks by socioeconomic inequality

The effects of drought exposure are heterogeneous across households of different socioeconomic statuses. Drought shock exposure significantly enhances seed purchasing decisions in the male sub-sample compared to the female sub-sample in all the studied countries. This result is particularly true for local seed purchasing in Ethiopia and improved seed purchases in Malawi and Tanzania. Women farmers or decision-makers are particularly at risk of increased marginalization when there is climate change-induced competition for resources (Eastin 2018). This view could explain why female decision-making is associated with lower chances and or intensity of seed purchasing post-shock exposure when compared to their male counterparts in studied countries.

Also, the more affluent households (in terms of household asset wealth) in all the studied countries are more likely to purchase seeds particularly improved seeds post-drought shock exposure, compared to their poorer counterparts. A possible explanation for the findings is that the poorer farmers are more vulnerable, and drought shocks reduce their propensity to purchase seeds, especially more expensive off-farm seeds, such as improved seeds. Assets are essential as informal insurance for rural households, and hence a strong asset base is an important characteristic of drought-resilient households (Dercon 2005; Gerber and Mirzabaev 2017). Farmers better endowed with household assets are hence more resilient, and they can still get some means of purchasing seed which is unlikely amongst the relatively poorer farmers. Nonetheless, the reliability of assets as informal insurance reduces as the scale and frequency of climate-induced risk increase (Dercon 2005; Fafchamps et al.

1998). Overall, results show that disadvantaged farmers, including women farmers and asset-poor households, are likely to be seed insecure post-drought exposure compared to their opposite counterparts. These findings is a basis for “realism about which farming households can be served by current approaches to seed system development”(Almekinders et al. 2021), confirming the view that a number of models for variety development and dissemination is needed to cater for the heterogenous farming communities in the study countries (Mausch et al. 2021).

6 Conclusions and implications

Drought shocks experienced in the recent past encourage seed purchasing in the following season in studied countries. On average, drought shock exposure increases seed purchasing for both improved and local seeds in Malawi and Tanzania while encouraging(discouraging) local(improved) seed purchases in Ethiopia. Also, drought shock exposure reduces the intensity of purchase for seed purchasers (as shown in some results for Malawi and Ethiopia). The implication is that drought shocks have contrasting effects on seed purchasing decisions in smallholder farming: it can motivate more seed purchasing (for different types of seed) to increase climate resilience, and the drought itself may have reduced the supply of farm-produced seeds. But, the income loss induced by the drought may also impose liquidity constraints and limit the opportunities for buying seeds, an effect that may dominate in Ethiopia (for improved seeds purchases) and Malawi (for the intensity of improved seed purchases).

Following this line of reasoning, we find – as expected – that better-endowed farmers increase the purchase of seeds and particularly of improved seeds after drought exposure. Also, women's decision-making on input acquisition is associated with lower (higher) chances of purchasing improved (local) seeds following drought shock exposure. Hence, we conclude that drought shock exposure and socioeconomic disadvantages, amongst other structural factors, increase seed insecurity in smallholder farming.

The findings support the call for the promotion of integrated approaches to seed system development. Besides using farmer saved seeds, farmers also use seeds purchased through both formal and informal channels. Access to off-farm seed through purchasing potentially helps improve the resilience of farmer seed systems to climate risk. As formal channels represent the most likely primary source for new varieties, policy should address the inequality in access to such channels through supply-side measures such as increasing production of affordable quality-controlled seed and demand-side measures such as social protection programs. At the same time, informal seed systems continue to be the backbone of the seed systems farmers use, and seed policies and regulations should enable the co-existence of formal and informal systems. An integrated seed systems approach supported with policies that will reduce inequities in accessing seed from commercial sources will improve the accessibility of improved and local seed varieties and serve the poor and vulnerable groups.

Acknowledgments

This research work is an output of a research project, “*Access to seeds: from Emergencies to Seed System development (ACCESS)”* which is fully supported by the Research Council of Norway, project number (RCN 288493). The authors are also grateful to the World Bank for making the Livelihood Standard Measurement survey data for Ethiopia, Malawi and Tanzania available for this research.

References

Acevedo M et al. (2020) A scoping review of adoption of climate-resilient crops by small-scale producers in low-and middle-income countries Nature plants 6:1231-1241

Aguilar A, Carranza E, Goldstein M, Kilic T, Oseni G (2015) Decomposition of gender differentials in agricultural productivity in Ethiopia Agricultural Economics 3:311-334

Akpo E, Muricho G, Lukurugu GA, Opie H, Ojiewo CO, Varshney R (2020) Legume seed production for sustainable seed supply and crop productivity: case of groundnut in Tanzania and Uganda Journal of Crop Improvement 34:518-539

Almekinders CJ, Louwaars NP, De Bruijn G (1994) Local seed systems and their importance for an improved seed supply in developing countries Euphytica 78:207-216

Almekinders CJ, Mausch K, Donovan J (2021) Editorial introduction: Design issues and practical questions for demand-oriented seed systems vol 50. SAGE Publications Sage UK: London, England,

Almekinders CJ, Thiele G, Danial DL (2007) Can cultivars from participatory plant breeding improve seed provision to small-scale farmers? Euphytica 153:363-372

Atilaw A, Korbu L (2011) Recent development in seed systems of Ethiopia Improving farmers' access to seed:13-30

Barrett CB (2008) Smallholder market participation: Concepts and evidence from eastern and southern Africa Food policy 33:299-317

Bellon MR, Adato M, Becerril J, Mindek D (2006) Poor farmers' perceived benefits from different types of maize germplasm: the case of creolization in lowland tropical Mexico World development 34:113-129

Bellon MR, Hodson D, Hellin J (2011) Assessing the vulnerability of traditional maize seed systems in Mexico to climate change Proceedings of the National Academy of Sciences 108:13432-13437

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge university press,

Challinor AJ, Koehler A-K, Ramirez-Villegas J, Whitfield S, Das B (2016) Current warming will reduce yields unless maize breeding and seed systems adapt immediately Nature Climate Change 6:954-958

Chambers RG, Quiggin J (2000) Uncertainty, production, choice, and agency: the state-contingent approach. Cambridge University Press,

Coomes OT et al. (2015) Farmer seed networks make a limited contribution to agriculture? Four common misconceptions Food Policy 56:41-50

Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods Econometrica: Journal of the Econometric Society:829-844

Crawford E, Kelly V, Jayne TS, Howard J (2003) Input use and market development in Sub-Saharan Africa: an overview Food Policy 28:277-292

Croppenstedt A, Demeke M, Meschi MM (2003) Technology adoption in the presence of constraints: the case of fertilizer demand in Ethiopia Review of Development Economics 7:58-70

Dercon S (2005) Risk, poverty and vulnerability in Africa Journal of African Economies 14:483-488

Dercon S, Christiaensen L (2011) Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia Journal of development economics 96:159-173

Eastin J (2018) Climate change and gender equality in developing states World Development 107:289-305

Enfors EI, Gordon LJ (2008) Dealing with drought: The challenge of using water system technologies to break dryland poverty traps Global Environmental Change 18:607-616 doi:https://doi.org/10.1016/j.gloenvcha.2008.07.006
Erenstein O, Kassie GT (2018) Seeding eastern Africa's maize revolution in the post-structural adjustment era: a review and comparative analysis of the formal maize seed sector International Food and Agribusiness Management Review 21:39-52
Fafchamps M, Udry C, Czukas K (1998) Drought and saving in West Africa: are livestock a buffer stock? Journal of Development economics 55:273-305
FAO (1998) Developing seed security strategies and programmes for food security in developing countries. FAO, Rome, Italy
FAO (2015) Voluntary Guide for National Seed Policy Formulation.
FAO, ECHA (2015) Building Capacity for Seed Security Assessments: Household Seed Security Concepts and Indicators, Discussion Paper. FAO, and European Commission Humanitarian Aid. .
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas International journal of climatology 37:4302-4315
Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of india* Demography 38:115-132
Gebregziabher G, Holden ST (2011) Distress rentals and the land rental market as a safety net: contract choice evidence from Tigray, Ethiopia Agricultural Economics 42:45-60
Gemeda A, Aboma G, Verkuijl H, Mwangi W (2001) Farmers' maize seed systems in Western Oromia, Ethiopia. . International Maize and Wheat Improvement Center (CIMMYT) and Ethiopian Agricultural Research Organization (EARO), Ethiopia
Gerber N, Mirzabaev A (2017) Benefits of action and costs of inaction: Drought mitigation and preparedness–a literature review Drought and Water Crises: Integrating Science, Management, and Policy:95
Heckman J (1979) Sample selection bias as a specification error Econometrica 47:153–161
Holden S, Shiferaw B (2004) Land degradation, drought and food security in a less-favoured area in the Ethiopian highlands: a bio-economic model with market imperfections Agricultural Economics 30:31-49
Holden ST, Quiggin J (2017) Climate risk and state-contingent technology adoption: shocks, drought tolerance and preferences European Review of Agricultural Economics 44:285-308
Howden SM, Soussana J-F, Tubiello FN, Chhetri N, Dunlop M, Meinke H (2007) Adapting agriculture to climate change Proceedings of the national academy of sciences 104:19691-19696
Humphreys BR (2013) Dealing with zeros in economic data University of Alberta, Department of Economics:1-27
Jayne TS, Rashid S (2013) Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence Agricultural economics 44:547-562
Katengeza SP, Holden ST, Lunduka RW (2019) Adoption of drought tolerant maize varieties under rainfall stress in Malawi Journal of Agricultural Economics 70:198-214 doi: https://doi.org/10.1111/1477-9552.12283
Kubik Z, Maurel M (2016) Weather shocks, agricultural production and migration: Evidence from Tanzania The Journal of Development Studies 52:665-680
Langyintuo AS, Mwangi W, Diallo AO, MacRobert J, Dixon J, Banziger M (2010) Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private-public intervention to promote growth Food Policy 35:323-331 doi:10.1016/j.foodpol.2010.01.005
Letta M, Montalbano P, Tol RS (2018) Temperature shocks, short-term growth and poverty thresholds: evidence from rural Tanzania World Development 112:13-32
Louwaars N (2005) Seed laws: biases and bottlenecks Seedling:4-9

Louwaars NP, de Boef WS (2012) Integrated Seed Sector Development in Africa: A Conceptual Framework for Creating Coherence Between Practices, Programs, and Policies Journal of Crop Improvement 26:39-59 doi:10.1080/15427528.2011.611277
Lunduka R, Fisher M, Snapp S (2012) Could farmer interest in a diversity of seed attributes explain adoption plateaus for modern maize varieties in Malawi? Food Policy 37:504-510 doi:https://doi.org/10.1016/j.foodpol.2012.05.001
Masarie KA, Tans PP (1995) Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record Journal of Geophysical Research: Atmospheres 100:11593-11610
Mausch K, Almekinders CJ, Hambloch C, McEwan MA (2021) Putting diverse farming households' preferences and needs at the centre of seed system development Outlook on Agriculture:00307270211054111
McGuire S, Sperling L (2011) The links between food security and seed security: facts and fiction that guide response Development in Practice 21:493-508
McGuire S, Sperling L (2013) Making seed systems more resilient to stress Global Environmental Change 23:644-653
McGuire S, Sperling L (2016) Seed systems smallholder farmers use Food Security 8:179-195
McGuire SJ (2008) Securing Access to Seed: Social Relations and Sorghum Seed Exchange in Eastern Ethiopia Human Ecology 36:217-229 doi:10.1007/s10745-007-9143-4
McKenzie DJ (2005) Measuring inequality with asset indicators Journal of population economics 18:229-260
Michler JD, Baylis K, Arends-Kuenning M, Mazvimavi K (2019) Conservation agriculture and climate resilience Journal of environmental economics and management 93:148-169
MoA (2014) Productive Safety Net Programme Phase IV Programme Implementation Manual. Ministry of Agriculture(MOA), Government of Ethiopia, Addis Ababa, Ethiopia. doi:https://www.usaid.gov/sites/default/files/documents/1866/psnp_iv_programme_implementation_manual_14_dec_14.pdf
Mortimore MJ, Adams WM (2001) Farmer adaptation, change and 'crisis' in the Sahel Global Environmental Change 11:49-57 doi:https://doi.org/10.1016/S0959-3780(00)00044-3
Mulesa TH, Dalle SP, Makate C, Haug R, Westengen OT (2021) Pluralistic Seed System Development: A Path to Seed Security? Agronomy 11:372
Munthali W, Okori P (2018) Community seed banks in Malawi: an informal approach for seed delivery. International Crops Research Institute for the Semi-Arid Tropics(ICRISAT). doi:https://cgspace.cgiar.org/bitstream/handle/10568/96202/malawi_model.pdf?sequence=1&isAllowed=y
Nordhagen S, Pascual U (2013) The impact of climate shocks on seed purchase decisions in Malawi: implications for climate change adaptation World Development 43:238-251
Peterman A, Behrman JA, Quisumbing AR (2014) A review of empirical evidence on gender differences in nonland agricultural inputs, technology, and services in developing countries. In: Gender in agriculture. Springer, pp 145-186
Sheahan M, Barrett CB (2017) Ten striking facts about agricultural input use in Sub-Saharan Africa Food Policy 67:12-25 doi:https://doi.org/10.1016/j.foodpol.2016.09.010
Sisay DT, Verhees FJHM, van Trijp HCM (2017) Seed producer cooperatives in the Ethiopian seed sector and their role in seed supply improvement: A review Journal of Crop Improvement 31:323-355 doi:10.1080/15427528.2017.1303800
Slavchevska V (2015) Gender differences in agricultural productivity: the case of Tanzania Agricultural Economics 46:335-355
Speranza CI, Wiesmann U, Rist S (2014) An indicator framework for assessing livelihood resilience in the context of social–ecological dynamics Global Environmental Change 28:109-119

Sperling L (2020) Seed security response during COVID-19: building on evidence and orienting to the future Food Security:1-5
Sperling L, Cooper HD, Remington T (2008) Moving towards more effective seed aid The Journal of Development Studies 44:586-612
Sperling L, Gallagher P, McGuire S, March J, Templer N (2020) Potential Seed Traders: The Backbone of Seed Business and African Smallholder Seed Supply Sustainability 12:7074
Tripp R (1996) Biodiversity and modern crop varieties: sharpening the debate Agriculture and human values 13:48-63
Tripp R (1997) New seed and old laws: regulatory reform and the diversification of national seed systems. Intermediate Technology Publications Ltd (ITP),
Wale E (2012) Farmers' perceptions on replacement and loss of traditional crop varieties: examples from Ethiopia and implications. In: The Economics of Managing Crop Diversity On-farm. Routledge, pp 79-104
Ward PS, Shively GE (2015) Migration and Land Rental as Responses to Income Shocks in Rural C hina Pacific Economic Review 20:511-543
Westengen OT, Haug R, Guthiga P, Macharia EJFiSFS (2019) Governing seeds in East Africa in the face of climate change: Assessing political and social outcomes 3:53
Westengen OT, Ring KH, Berg PR, Brysting AK (2014) Modern maize varieties going local in the semi-arid zone in Tanzania BMC Evolutionary Biology 14:1-12
Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press,
Yesuf M, Bluffstone RA (2009) Poverty, Risk Aversion, and Path Dependence in Low-Income Countries: Experimental Evidence from Ethiopia American Journal of Agricultural Economics 91:1022-1037 doi:10.1111/j.1467-8276.2009.01307.x %J American Journal of Agricultural Economics

Notes

---

[1] The LSMA-ISA data are available publicly on the LSMS website (www.worldbank.org/lsms)
[2] FAOSTAT 2019
[3] Access to Seed Index (ASI) https://www.accesstoseeds.org/
4 The CGIAR's Diffusion and Impact of Improved Varieties in Africa (DIIVA) project https://asti.cgiar.org/diiva#0
[5] Sisay et al. (2017)
[6] Akpo et al. (2020)
[7] MoA (2014)
[8] Munthali and Okori (2018)
[9] The support for farmer-led seed systems in African seed laws (2017) https://www.bioversityinternational.org/fileadmin/user_upload/The_support_Herpers_2017.pdf
[10] The African Seed ACCESS index https://tasai.org/
[11] https://www.asa.go.tz/about-us/
[12] Such preferences may also have been affected by past shocks and experiences.
[13] https://www.worldclim.org/data/monthlywth.html
[14] Five quintiles are derived from the household asset wealth index (derived from principal components analysis) and a dummy variable (Rich) with 1=yes is derived to indicate proportion of households in the top two richest quintiles (quintiles 4 and 5) (top 40% of the asset wealth distribution).

Appendix: ***Rainfall shocks and inequality have heterogeneous effects on farmers' seed purchase decisions in East Africa***

Clifton Makate[*] , Arild Angelsen, Stein Terje Holden, & Ola Tveitereid Westengen

A. Descriptive statistics

- ***The full table of descriptive statistics of variables***

Table A: Descriptive statistics of variables used in the analysis

	Ethiopia			**Malawi**				**Tanzania**			
Variable	Pooled mean	ESS3 mean	ESS4 mean	Pooled mean	IHS3 mean	IHS4 mean	IHS5 mean	Pooled mean	TNPS2 mean	TNPS3 mean	TNPS4 mean
Seed purchasing variables											
Incidence of seed purchasing†	0.476	0.466	0.490	0.460	0.395	0.481	0.513	0.389	0.332	0.383	0.471
Incidence of seed purchasing (improved)†	0.175	0.173	0.177	0.276	0.246	0.284	0.300	0.258	0.107	0.383	0.256
Incidence of seed purchasing (local)†	0.393	0.383	0.408	0.285	0.227	0.295	0.342	0.284	0.263	.	0.310
Quantity purchased(kg/ha) [**unconditional**]	21.692	20.876	22.758	10.631	6.541	11.225	14.796	7.239	.	3.939	12.371
Quantity purchased(kg/ha) [**conditional**]	44.297	42.843	46.176	22.853	16.536	23.322	27.937	16.63	.	9.01	26.28
Value of seed purchased (local currency/ha) [**unconditional**]	327.281	309.143	350.992	4179.16	735.95	5596.25	6673.59	11076.8	6319.13	10391.1	18604.4
Value of seed purchased (local currency/ha) [**conditional**]	668.342	634.448	712.148	8984.78	1860.04	11626.81	12600.61	27405.32	18542.28	25788.77	37796.62
Share purchased seed (%) [**unconditional**]	25.297	23.379	27.903	27.668	19.776	32.313	31.711	16.499	.	8.280	29.278
Share purchased seed (%) [**conditional**]	53.118	50.163	56.937	60.086	50.044	67.141	61.842	39.524	.	21.61	62.199
Socioeconomic variables											
Female (Female decision maker) †	0.210	0.221	0.195	0.290	0.253	0.303	0.320	0.237	0.225	0.232	0.259
Farm size (ha)	1.126	1.311	0.874	0.656	0.735	0.591	0.637	2.165	2.116	2.277	2.056
Rich (Household is in the top 40% of the sample asset wealth index distribution(1=yes) †	0.388	0.282	0.533	0.324	0.324	0.312	0.337	0.290	0.353	0.338	0.135
Household asset wealth index	-0.392	-0.439	-0.329	-0.254	-0.218	-0.292	-0.255	-0.651	-0.331	-0.408	-1.437
Received assistance with seed from government or NGOs (e.g. relief or coupon seed) (1=yes)	0.182	0.216	0.135	0.144	0.229	0.119	0.071	0.202	0.392	0.125	0.080
Distance to main market(km)	64.133	66.435	61.004	22.015	17.000	24.810	24.751	12.270	13.092	12.799	10.403
Distance to paved road(km)	16.124	15.302	17.240	10.420	10.060	10.852	10.371	3.347	3.233	3.557	3.164
Tropical Livestock Units(TLU)	3.404	3.868	2.775	-	-	-	-	2.366	2.123	2.498	2.469
Age of household head(years)	46.999	47.967	45.684	44.152	43.169	44.567	44.835	48.111	48.764	48.489	46.692
Education level attained lower than 12th grade(Ethiopia) †, at least JCE(Malawi) and higher than D7(Tanzania) †	0.294	0.296	0.292	0.331	0.365	0.317	0.307	0.395	0.408	0.391	0.385
Household size	5.896	6.423	5.179	4.573	4.706	4.435	4.569	5.670	5.786	5.717	5.449
Access to market information(extension) †	0.405	0.377	0.444	0.277	0.258	0.324	0.249	0.119	0.141	0.096	0.129
Labor(number of hired men labor(Ethiopia), number of weeks spent in an agricultural season(Malawi), and hours spend in agriculture in a week(Tanzania)	3.783	1.738	6.563	21.579	28.413	15.067	20.737	59.796	56.659	62.285	59.913
Climate variables and shocks											
Historical rainfall growing season (1980-2018)	769.017	763.881	775.996	949.372	957.738	950.474	938.651	852.134	864.499	856.894	828.977
Historical temperature growing season (1980-2018)	25.877	25.810	25.968	28.198	28.173	28.233	28.189	28.461	28.581	28.497	28.253
Rainfall shock growing season (1-year lag)	0.339	0.352	0.321	-0.001	0.382	-0.378	-0.035	-0.257	0.053	-0.633	-0.067
Observations	4987	2873	2114	26627	9467	8862	8298	6665	2214	2709	1742

Notes: Climate variables and shocks are shown for the main rainy season of respective countries; Statistics are not weighted; source (own calculation from LSMS-ISA and WordClim data); †denotes dummy variable; conditional and unconditional refer to stats defined for seed purchasers and full sample respectively; Share purchased seeds is a percentage computed as the proportion of quantities of seed purchased to total seeds used for each crop grown and averaged for all crops grown by the household; ESS=Ethiopia Socioeconomic Survey, TNPS=Tanzania National Panel Survey, IHS=Malawi Integrated Household Survey.

- ***Descriptive characteristics of key variables by gender of prime decision-maker on input acquisition within the household***

Table B: Descriptive statistics by gender of the main decision-maker on inputs acquisition within the household

	Ethiopia			***Malawi***			***Tanzania***		
	Female	**Male**	**p-value**	**Female**	**Male**	**p-value**	**Female**	**Male**	**p-value**
Incidence of seed purchasing†	0.434	0.488	0.002	0.403	0.484	0.000	0.339	0.405	0.000
Incidence of seed purchasing (improved)†	0.130	0.187	0.000	0.193	0.309	0.000	0.217	0.271	0.000
Incidence of seed purchasing (local)†	0.365	0.401	0.033	0.290	0.283	0.262	0.263	0.290	0.104
Quantity purchased(kg/ha) [unconditional]	20.872	21.903	0.643	9.715	11.006	0.000	.	.	.
Value of seed purchased (local currency/ha) [unconditional]	335.749	325.103	0.710	3311.666	4534.084	0.000	10863.906	11143.198	0.771
Share purchased seed (%) [unconditional]	25.355	25.282	0.956	24.184	29.093	0.000	.	.	.
Distance to main market(km)	63.454	64.314	0.605	21.629	22.173	0.001	11.169	12.612	0.002
Distance to paved road(km)	15.684	16.241	0.354	10.517	10.381	0.360	3.191	3.395	0.254
Received assistance with seed from government or NGOs (e.g. relief or coupon seed) (1=yes)	0.227	0.170	0.000	0.138	0.146	0.097	0.208	0.200	0.468
Household asset wealth index	-0.345	-0.405	0.230	-0.757	-0.048	0.000	-1.125	-0.504	0.000
Rich (Household is in the top 40% of the sample asset wealth index distribution(1=yes) †	0.391	0.387	0.844	0.183	0.382	0.000	0.197	0.319	0.000
Tropical Livestock Units(TLU)	2.182	3.729	0.000	.	.	.	1.534	2.624	0.000
Farm size (ha)	0.744	1.227	0.000	0.532	0.707	0.000	1.556	2.354	0.000
Age of household head(years)	50.284	46.126	0.000	48.548	42.346	0.000	53.544	46.424	0.000
Education level attained lower than 12th grade(Ethiopia) †, at least JCE(Malawi) and higher than D7(Tanzania) †	0.100	0.346	0.000	0.407	0.300	0.000	0.542	0.350	0.000
Household size	4.581	6.245	0.000	3.869	4.861	0.000	4.659	5.984	0.000
Access to market information(extension) †	0.334	0.424	0.000	0.262	0.284	0.000	0.079	0.132	0.000
Labor(number of hired men labor(Ethiopia), number of weeks spent in an agricultural season(Malawi), and hours spend in agriculture in a week(Tanzania)	1.963	4.267	0.000	15.983	23.868	0.000	46.514	63.920	0.000
Historical rainfall growing season (1980-2018)	766.617	769.654	0.764	947.429	950.167	0.244	858.782	850.067	0.126
Historical temperature growing season (1980-2018)	25.836	25.888	0.658	28.320	28.148	0.000	28.494	28.451	0.508
Rainfall shock growing season (1-year lag)	0.292	0.351	0.023	0.008	-0.004	0.085	-0.242	-0.262	0.357
Observations	1,047	3,940		7,730	18,897		1,579	5,083	

Notes: Notes: p-value is from a t-test mean comparison test; Statistics are not weighted

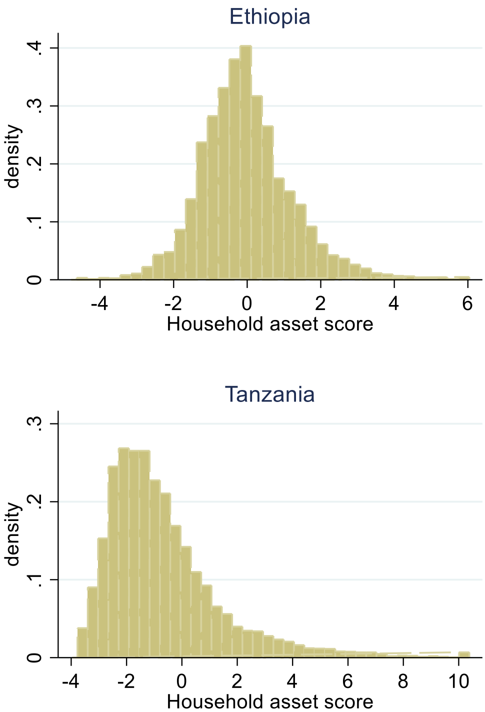

Figure A:Distribution of the Household asset wealth index in respective countries. The household asset index is defined as a weighted sum of durable household assets and other home-dwelling features common in each country. The weights are derived from principal components analysis (Filmer and Pritchett 2001).

B. Full Tables of Main Results shown in Manuscript: Double Hurdle (Cragg) model

Table C: Double Hurdle model Estimates: Impact of climate shocks and other socioeconomic characteristics on household seed purchasing decisions in rural Ethiopia

	All purchased seed		Improved seed		Local seed		Improved Maize		Local Maize		Local Sorghum	
VARIABLES	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
drought shock(growing season) 1-year lag	0.148**	-0.142	-0.466***	-0.414**	0.255***	-0.006	-0.589***	-0.084	0.333**	0.502**	0.561***	0.531***
	(0.0723)	(0.1327)	(0.0943)	(0.1694)	(0.0723)	(0.1666)	(0.1380)	(0.1251)	(0.1302)	(0.2095)	(0.1558)	(0.1852)
Rainfall (GS) mm (1980-2018)(log)	0.334***	-0.158	0.725***	-0.340*	0.207***	-0.133	1.180***	-0.147	-0.280***	-0.655***	0.016	-0.461**
	(0.0633)	(0.1120)	(0.1107)	(0.1796)	(0.0616)	(0.1324)	(0.1845)	(0.1160)	(0.0957)	(0.1670)	(0.1407)	(0.2056)
Temperature (GS) °C (1980-2018)	-0.022***	-0.032***	0.024***	-0.034***	-0.031***	-0.032***	0.005	0.035***	-0.043***	0.019	-0.011	0.073***
	(0.0037)	(0.0067)	(0.0045)	(0.0120)	(0.0037)	(0.0077)	(0.0068)	(0.0077)	(0.0070)	(0.0130)	(0.0086)	(0.0144)
Female decision maker(1=yes)	-0.037	-0.134	-0.061	-0.230	-0.053	-0.031	-0.068	-0.079	0.104	0.134	0.277*	-0.053
	(0.0668)	(0.1160)	(0.0829)	(0.1686)	(0.0676)	(0.1357)	(0.1131)	(0.1112)	(0.1174)	(0.1867)	(0.1450)	(0.1639)
Household asset wealth index	0.038***	0.010	0.032*	0.053*	0.031**	-0.004	0.039*	0.046**	-0.017	0.040	-0.002	0.088*
	(0.0134)	(0.0262)	(0.0181)	(0.0296)	(0.0137)	(0.0335)	(0.0234)	(0.0188)	(0.0208)	(0.0382)	(0.0256)	(0.0508)
Log Farm size(ha)	0.176***	0.973***	0.168***	0.837***	0.136***	0.982***	0.041	0.327***	0.035	0.958***	-0.127	0.766***
	(0.0518)	(0.0938)	(0.0581)	(0.1261)	(0.0513)	(0.1143)	(0.0767)	(0.0830)	(0.0868)	(0.1799)	(0.1049)	(0.1830)
Relief inputs(1=yes)	0.007	0.297***	-0.020	0.525***	-0.005	0.210*	0.018	0.095	0.074	0.132	-0.027	-0.010
	(0.0535)	(0.0928)	(0.0791)	(0.1591)	(0.0532)	(0.1119)	(0.1234)	(0.1109)	(0.0883)	(0.1539)	(0.1001)	(0.1337)
Distance to market(km, log)	-0.055*	0.070	-0.122***	0.006	-0.012	0.175***	-0.212***	0.044	-0.024	0.016	-0.095	-0.088
	(0.0286)	(0.0506)	(0.0380)	(0.0684)	(0.0285)	(0.0633)	(0.0506)	(0.0432)	(0.0494)	(0.0807)	(0.0623)	(0.0709)
Distance to paved road(km)	-0.000	-0.003	-0.007***	-0.005	0.002*	-0.003	-0.006**	-0.002	0.000	0.002	0.009***	0.001
	(0.0013)	(0.0022)	(0.0018)	(0.0031)	(0.0012)	(0.0026)	(0.0025)	(0.0022)	(0.0022)	(0.0041)	(0.0025)	(0.0030)
Access to market information (extension)	0.390***	-0.055	0.982***	-0.156	0.066	-0.119	0.928***	-0.040	-0.302***	-0.105	-0.225***	-0.422***
	(0.0399)	(0.0653)	(0.0508)	(0.0965)	(0.0401)	(0.0819)	(0.0656)	(0.0634)	(0.0664)	(0.1286)	(0.0851)	(0.1317)
Tropical Livestock Units(TLU)	-0.030***	0.012	-0.002	0.002	-0.034***	0.021	-0.002	0.026***	-0.038***	0.036**	-0.021	0.004
	(0.0065)	(0.0126)	(0.0016)	(0.0158)	(0.0067)	(0.0144)	(0.0015)	(0.0097)	(0.0120)	(0.0161)	(0.0129)	(0.0218)
Age of household head(years)	-0.004***	-0.008***	-0.005***	-0.006*	-0.002	-0.009***	-0.005**	0.001	-0.001	-0.001	-0.000	0.000
	(0.0013)	(0.0024)	(0.0018)	(0.0029)	(0.0013)	(0.0029)	(0.0023)	(0.0020)	(0.0023)	(0.0043)	(0.0026)	(0.0043)
Education level attained lower than 12th	0.022	-0.002	0.038	0.060	-0.011	-0.063	-0.002	-0.003	0.084	0.067	0.005	-0.128
grade(1=yes)	(0.0446)	(0.0744)	(0.0561)	(0.0949)	(0.0444)	(0.0934)	(0.0738)	(0.0677)	(0.0723)	(0.1329)	(0.0927)	(0.1454)
Household size	0.037***	0.037**	0.025**	0.042**	0.035***	0.028	0.026*	0.048***	0.015	-0.030	-0.004	-0.021
	(0.0092)	(0.0163)	(0.0116)	(0.0198)	(0.0091)	(0.0203)	(0.0147)	(0.0136)	(0.0153)	(0.0262)	(0.0174)	(0.0272)
Hired labor (days,log)	0.062***	0.116***	0.100***	0.077**	0.051***	0.105***	0.145***	0.107***	0.023	0.019	0.119***	0.176***
	(0.0188)	(0.0286)	(0.0216)	(0.0337)	(0.0184)	(0.0362)	(0.0280)	(0.0238)	(0.0278)	(0.0665)	(0.0313)	(0.0540)
Family labor in a week (hours, log)	0.023*	-0.011	0.055***	0.013	0.012	-0.027	0.080***	0.036*	-0.029	-0.051	-0.001	-0.025
	(0.0126)	(0.0226)	(0.0173)	(0.0295)	(0.0126)	(0.0273)	(0.0228)	(0.0216)	(0.0199)	(0.0386)	(0.0249)	(0.0416)
Household head is single(1=yes)	0.092	0.047	0.044	0.150	0.061	0.028	0.057	0.100	-0.116	-0.312	0.048	-0.209
	(0.0672)	(0.1188)	(0.0817)	(0.1650)	(0.0680)	(0.1407)	(0.1104)	(0.1134)	(0.1207)	(0.2042)	(0.1418)	(0.1460)
Sigma constant		1.322***		1.031***		1.420***		0.656***		0.936***		0.746***
		(0.0253)		(0.0389)		(0.0314)		(0.0358)		(0.0493)		(0.0501)
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,987	2,375	4,987	873	4,987	1,962	2,923	683	2,923	350	1,978	254

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability purchasing seed(1=yes,0Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha)., All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for Maize and sorghum.

Table D: Double Hurdle model estimates: Impact of climate shocks and other socioeconomic characteristics on household seed purchasing decisions in rural Malawi

	All purchased seed		Improved seed(all crops)		Local seed(all crops)		Improved Maize		Local Maize		Local pigeon pea	
VARIABLES	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
drought shock (GS) 1-year lag	0.106***	-0.058	0.083**	-0.108**	0.002	0.034	0.058	-0.056*	-0.035	-0.168***	-0.008	0.241*
	(0.0340)	(0.0387)	(0.0357)	(0.0428)	(0.0393)	(0.0631)	(0.0385)	(0.0298)	(0.0544)	(0.0626)	(0.1588)	(0.1413)
Rainfall (GS) mm (1980-2018)(log)	-0.330***	0.055	-0.236***	0.418***	-0.560***	-0.254**	0.048	-0.133**	-0.347***	-0.110	-0.321**	0.412***
	(0.0632)	(0.0678)	(0.0681)	(0.0774)	(0.0701)	(0.1091)	(0.0740)	(0.0594)	(0.0912)	(0.0944)	(0.1492)	(0.1217)
Temperature (GS) °C (1980-2018)	0.015***	0.017***	0.003	-0.009***	0.003	0.023***	0.012***	-0.009***	0.036***	-0.002	0.058***	0.010
	(0.0028)	(0.0031)	(0.0030)	(0.0036)	(0.0031)	(0.0050)	(0.0033)	(0.0029)	(0.0042)	(0.0047)	(0.0086)	(0.0073)
Female decision maker(1=yes)	-0.121***	-0.068***	-0.208***	-0.014	-0.019	-0.060**	-0.223***	-0.025	-0.005	0.015	-0.118***	-0.022
	(0.0188)	(0.0202)	(0.0207)	(0.0251)	(0.0202)	(0.0280)	(0.0222)	(0.0204)	(0.0256)	(0.0259)	(0.0349)	(0.0275)
Household asset wealth index	0.047***	0.035***	0.078***	0.057***	-0.031***	-0.011	0.104***	0.036***	-0.120***	0.025	-0.053**	0.026**
	(0.0080)	(0.0071)	(0.0084)	(0.0086)	(0.0085)	(0.0130)	(0.0089)	(0.0066)	(0.0217)	(0.0172)	(0.0207)	(0.0131)
Agricultural implement access index	0.042***	0.006	0.069***	0.010	-0.005	-0.009	0.061***	0.019***	-0.069***	0.019	-0.031**	0.016
	(0.0062)	(0.0066)	(0.0065)	(0.0074)	(0.0067)	(0.0102)	(0.0067)	(0.0053)	(0.0107)	(0.0125)	(0.0148)	(0.0137)
Log of farmsize(ha)	0.218***	0.711***	0.289***	0.461***	0.126***	0.957***	0.191***	0.616***	-0.163***	0.920***	-0.317***	0.811***
	(0.0322)	(0.0396)	(0.0338)	(0.0469)	(0.0334)	(0.0584)	(0.0350)	(0.0378)	(0.0465)	(0.0584)	(0.0735)	(0.0708)
Coupon seed(1=yes)	-0.221***	-0.225***	-0.249***	-0.154***	-0.121***	-0.231***	-0.336***	-0.099***	-0.299***	-0.099***	-0.033	-0.013
	(0.0191)	(0.0223)	(0.0209)	(0.0276)	(0.0208)	(0.0313)	(0.0221)	(0.0211)	(0.0274)	(0.0338)	(0.0365)	(0.0298)
Distance to nearest ADMARC (km)	-0.002*	-0.006***	0.000	-0.001	-0.005***	-0.005**	-0.002	0.000	-0.003	0.000	-0.002	-0.001
	(0.0013)	(0.0015)	(0.0014)	(0.0017)	(0.0015)	(0.0023)	(0.0015)	(0.0013)	(0.0020)	(0.0020)	(0.0031)	(0.0025)
Log Distance to paved road(km)	-0.031***	0.034***	-0.028***	0.041***	-0.035***	0.040***	-0.040***	0.030***	0.005	0.025**	-0.005	0.027*
	(0.0082)	(0.0089)	(0.0087)	(0.0109)	(0.0088)	(0.0127)	(0.0093)	(0.0085)	(0.0110)	(0.0122)	(0.0182)	(0.0158)
Access to market information (extension)	0.036**	-0.013	0.099***	0.001	-0.054***	-0.016	0.110***	-0.001	-0.146***	0.027	-0.053	-0.010
	(0.0180)	(0.0193)	(0.0190)	(0.0227)	(0.0194)	(0.0282)	(0.0201)	(0.0171)	(0.0261)	(0.0270)	(0.0354)	(0.0283)
Age of household head (years)	-0.009***	-0.001**	-0.008***	-0.001	-0.006***	0.001	-0.009***	0.001	-0.001**	0.002**	-0.006***	0.000
	(0.0005)	(0.0006)	(0.0006)	(0.0008)	(0.0006)	(0.0008)	(0.0006)	(0.0006)	(0.0007)	(0.0008)	(0.0010)	(0.0009)
Education level attained, at least JCE)	0.006	0.040**	0.019	0.076***	-0.032*	0.025	0.048**	0.020	-0.022	-0.034	-0.068**	-0.013
	(0.0174)	(0.0188)	(0.0185)	(0.0225)	(0.0187)	(0.0273)	(0.0197)	(0.0175)	(0.0239)	(0.0260)	(0.0340)	(0.0285)
Household size	0.026***	0.025***	0.022***	0.016***	0.023***	0.018***	0.027***	0.017***	0.026***	0.022***	0.043***	0.006
	(0.0042)	(0.0046)	(0.0044)	(0.0058)	(0.0045)	(0.0069)	(0.0047)	(0.0042)	(0.0059)	(0.0065)	(0.0084)	(0.0070)
Family labor(weeks)	0.001**	0.000	0.001**	0.000	0.001*	0.000	0.001**	-0.000	0.000	-0.001	0.001*	0.001
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0005)	(0.0003)	(0.0003)	(0.0004)	(0.0005)	(0.0006)	(0.0006)
Sigma constant		0.886***		0.826***		0.955***		0.575***		0.592***		0.561***
		(0.0081)		(0.0104)		(0.0107)		(0.0077)		(0.0100)		(0.0136)
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,627	11,261	26,627	7,337	26,627	7,597	24,945	5,894	24,945	2,943	7,468	2,453

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (Quantity of seeds purchased kg/ha). All crops model is estimated by considering seed purchasing variables averaged for all crops grown; Then variables are defined for Maize and local pigeon pea

Table E: Double Hurdle model Estimates: Impact of climate shocks and other socioeconomic characteristics on household seed purchasing decisions in rural Tanzania

	All purchased seed		Improved seed(all crops)		Local seed(all crops)		Improved Maize		Local Maize		Local bean	
VARIABLES	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdle2	Hurdle1	Hurdel2	Hurdle1	Hurdel2	Hurdle1	Hurdle2
Drought shock (GS) 1-year lag	0.103***	0.001	0.117***	-0.072	0.049	0.114	0.076	-0.101	0.336***	0.106	-0.126	0.562***
	(0.0306)	(0.0445)	(0.0341)	(0.0684)	(0.0449)	(0.0715)	(0.0467)	(0.0798)	(0.0697)	(0.0969)	(0.1312)	(0.1594)
Rainfall (GS) mm (1980-2018)(log)	-0.743***	-0.810***	-0.815***	-0.863***	-0.511***	-0.642***	-0.751***	-0.688***	-0.337**	-1.054***	-0.649***	-0.443
	(0.0768)	(0.1083)	(0.0853)	(0.1601)	(0.1040)	(0.1577)	(0.1021)	(0.1741)	(0.1426)	(0.2308)	(0.2047)	(0.2824)
Temperature (GS) °C (1980-2018)	-0.009	-0.088***	-0.010	-0.097***	0.001	-0.097***	-0.026**	-0.086***	0.071***	-0.058**	0.021	-0.076**
	(0.0074)	(0.0108)	(0.0082)	(0.0161)	(0.0103)	(0.0154)	(0.0106)	(0.0175)	(0.0173)	(0.0238)	(0.0255)	(0.0357)
Female decision maker(1=yes)	-0.050	-0.116**	-0.037	-0.049	0.011	-0.238***	-0.081	-0.077	-0.145*	-0.374***	-0.021	-0.167
	(0.0401)	(0.0549)	(0.0445)	(0.0777)	(0.0537)	(0.0792)	(0.0560)	(0.0808)	(0.0786)	(0.1100)	(0.1116)	(0.1432)
Household asset wealth index	0.026***	0.100***	0.062***	0.094***	-0.019	0.057***	0.114***	0.079***	-0.108***	0.011	0.001	0.052
	(0.0082)	(0.0111)	(0.0089)	(0.0140)	(0.0119)	(0.0187)	(0.0116)	(0.0141)	(0.0213)	(0.0356)	(0.0236)	(0.0337)
Log farm size(ha)	0.079***	0.319***	0.081***	0.339***	0.057*	0.245***	-0.144***	0.371***	-0.204***	0.313***	-0.141*	0.240**
	(0.0252)	(0.0372)	(0.0274)	(0.0518)	(0.0348)	(0.0529)	(0.0352)	(0.0539)	(0.0554)	(0.0854)	(0.0775)	(0.0976)
Value of relief received(log)	-0.018	-0.044	-0.054**	-0.023	0.009	-0.015	-0.048*	-0.011	0.037	0.033	0.042	0.136***
	(0.0175)	(0.0303)	(0.0226)	(0.0348)	(0.0214)	(0.0407)	(0.0281)	(0.0409)	(0.0264)	(0.0422)	(0.0602)	(0.0236)
Distance to the nearest market(km)	0.001	0.003***	-0.001	0.004*	0.002*	0.001	-0.001	0.006***	-0.003	0.000	-0.003	0.003
	(0.0010)	(0.0012)	(0.0012)	(0.0022)	(0.0012)	(0.0018)	(0.0016)	(0.0023)	(0.0022)	(0.0040)	(0.0032)	(0.0062)
Log distance to paved road	0.007	0.017	-0.012	-0.038	0.010	0.054	-0.103***	-0.044	0.004	0.036	-0.035	0.138**
	(0.0195)	(0.0268)	(0.0218)	(0.0397)	(0.0260)	(0.0394)	(0.0269)	(0.0423)	(0.0384)	(0.0534)	(0.0534)	(0.0649)
Family labor (hours spent in agric in a week, log)	0.052***	0.063***	0.050***	0.073***	0.043***	0.048**	0.026*	0.047**	0.028	-0.048	-0.048	0.057
	(0.0106)	(0.0147)	(0.0119)	(0.0215)	(0.0145)	(0.0217)	(0.0146)	(0.0231)	(0.0210)	(0.0315)	(0.0310)	(0.0417)
Access to information (extension)	0.248***	0.297***	0.396***	0.145*	0.018	0.171*	0.290***	0.207**	-0.190*	0.460***	-0.015	0.061
	(0.0496)	(0.0629)	(0.0527)	(0.0829)	(0.0644)	(0.0967)	(0.0631)	(0.0848)	(0.1041)	(0.1750)	(0.1175)	(0.1412)
Tropical Livestock Units(TLU)	0.001	0.003	0.004	0.009**	-0.010***	0.001	0.002	0.007*	-0.027***	0.028**	-0.041***	0.028*
	(0.0024)	(0.0039)	(0.0026)	(0.0038)	(0.0039)	(0.0066)	(0.0030)	(0.0042)	(0.0090)	(0.0115)	(0.0121)	(0.0166)
Age of household head(years)	-0.008***	-0.007***	-0.007***	-0.006***	-0.009***	-0.005**	-0.007***	-0.003	-0.004**	-0.005*	-0.013***	-0.008*
	(0.0011)	(0.0016)	(0.0012)	(0.0023)	(0.0015)	(0.0023)	(0.0015)	(0.0024)	(0.0021)	(0.0032)	(0.0031)	(0.0042)
Number of years in formal education	-0.007	-0.008	0.007	-0.027*	-0.015	-0.008	-0.004	-0.019	-0.052***	-0.017	-0.021	0.012
	(0.0078)	(0.0109)	(0.0085)	(0.0136)	(0.0105)	(0.0157)	(0.0111)	(0.0142)	(0.0177)	(0.0235)	(0.0235)	(0.0315)
Household size	0.012**	0.018**	0.010	0.015	0.022***	0.018	0.021**	0.018	0.023*	0.027	0.036**	-0.030
	(0.0058)	(0.0083)	(0.0064)	(0.0111)	(0.0081)	(0.0111)	(0.0084)	(0.0130)	(0.0128)	(0.0218)	(0.0171)	(0.0218)
Sigma constant		1.110***		1.024***		1.070***		0.944***		0.910***		1.012***
		(0.0190)		(0.0233)		(0.0233)		(0.0238)		(0.0347)		(0.0355)
Region Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,665	2,593	6,665	1,722	3,956	1,122	4,306	1,112	2,561	383	908	336

Notes: In parenthesis are cluster robust standard errors at the primary sampling unit(village); *** p<0.01, ** p<0.05, * p<0.1; Hurdle 1 is the probability of purchasing seed (1=yes,0 Otherwise), and Hurdle2 is the intensity of purchase (log (value of purchased seed/ha)). All crops model is estimated by defining seed purchasing variables averaged for all crops grown by the household. Then variables are defined for maize and bean crops; equations for local varieties are estimated for two surveys (TNPS2 and 4), where seed purchasing variables were defined by seed variety type.

Paper III

Evolution of farm-level crop diversification and response to rainfall shocks in smallholder farming: Evidence from Malawi and Tanzania

Clifton Makate[1*] , Arild Angelsen[1], Stein Terje Holden[1], & Ola Tveitereid Westengen[2]

[1*]School of Economics and Business, Norwegian University of Life Sciences, P.O.Box 5003, 1432 Ås, Norway; clifton.makate@nmbu.no (C.M); arild.angelsen@nmbu.no (A.A); stein.holden@nmbu.no (S.T.H)
[2]Department of International Environment and Development Studies, Noragric, Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway; ola.westengen@nmbu.no (O.T.W.)
* Corresponding author: clifton.makate@nmbu.no (C.M)

Abstract
Crop diversification is a key on-farm strategy for smallholders facing climate and market risks. This study investigates how recent exposure to recent rainfall shocks, long-term rainfall variability, and knowledge and experience from past crop diversification decisions influence farm-level crop diversification in Malawi and Tanzania. We use balanced household panel data combined with corresponding historical monthly weather data from the two countries to achieve our aim. Crop diversification at the household level is studied as a state-contingent risky investment decision that households make before the state of nature is revealed. Crop diversification decisions are modeled using correlated and dynamic random effects panel Poisson and Tobit models that control for unobserved heterogeneity in household crop diversification decisions plus initial conditions that may influence crop diversification across space and time. We establish that smallholder farmers in Malawi and Tanzania respond to rainfall shocks by diversifying their crop portfolio and that crop diversification decisions are state-dependent. Thus, farmers with knowledge and experience gained from past crop diversification have elevated chances to intensify subsequent crop diversification and adapt to recurrent rainfall shocks, unlike those without prior experience. Heterogeneity analysis reveals that it is the relatively better-off farmers with sufficient land and non-land household assets who are more likely to achieve successful crop diversification post-drought shock exposure as an adaptation mechanism. Our results support pro-poor policies that address both supply and demand-side constraints faced by smallholder farmers in accessing a diversity of seeds and planting material that meets their needs and preferences, important to support crop diversification as a strategy to improve resilience to recurrent rainfall shocks.

Keywords: crop diversification; rainfall shocks; state dependency; state-contingent framework; dynamic & correlated random effect models; heterogeneity analysis

1. Introduction

Climate risk is a central part of livelihoods in Sub-Saharan Africa. Recurrent rainfall and, or temperature variations associated with climate change expose farming households to production shocks that shape their farming practices and strategies (Kubik and Maurel 2016; Katengeza et al. 2019a; IPCC 2022). Agriculture is a key pillar in rural livelihoods, and exposure to production shocks affects access to agricultural inputs, production, and livelihoods. Characteristics of the rural settings such as high reliance on agriculture, lack of functional insurance markets, and the dire consequences of a bad season (Rose 2001; Dercon 2005) complicate both *ex-post* (risk coping mechanisms) and *ex-ante* (adaptation) response to shocks (Angelsen and Dokken 2018). However, literature shows that farming households, when exposed to climate shocks such as droughts or floods, adopt diverse strategies to cope or adapt (Dercon 2002; Dercon 2005; Alobo Loison 2015). Given that the vulnerability of households is closely associated with their resource endowment poverty (Dercon 2005), the choice of coping and adaptation strategies is usually a function of their land, labor, and asset endowments and the institutions that govern access to such resources (Dercon 2005; Winters et al. 2009).

Smallholder farmers may respond to shocks by diversifying their livelihood strategies both on and off-farm, and crop diversification is one important on-farm diversification strategy used by farmers (Ellis 2000; Morton 2007; Alobo Loison 2015). Crop diversification through the cultivation of a diverse collection of crop species and or varieties well-suited to local conditions and that meet farmers' preferences make up important strategies used by farmers in Sub-Sahara Africa (SSA) to cope and adapt to socio-economic and environmental risks (Lin 2011; Alobo Loison 2015; Labeyrie et al. 2021). Crop diversity is important in improving the resilience of cropping systems to climate risk. For instance, a diversified cropping portfolio can suppress pest and disease outbreaks that may worsen with increased climate variability (Chakraborty and Newton 2011; Lin 2011). When farmers grow multiple crop species on the farm, it makes the productive exploitation of synergies among crops and niche differentiation possible (Di Falco and Chavas 2009; Di Falco et al. 2010). Crop diversification also influences food production outcomes by reducing the risk of total crop failure and increasing production and production stability with increased climate risk(Di Falco et al. 2010; Lin 2011; Makate et al. 2016; Renard and Tilman 2019; Bellon et al. 2020). With a diversified cropping portfolio, farmers increase their chances of dealing with the uncertainty created by climate variability and change, reducing the risk of total crop failure and providing themselves with alternative means of generating income.

The importance of crop diversification in managing risks over time also comes through its contribution to the conservation of crop diversity *in situ* (Bellon 1996; Love and Spaner 2007; Bezabih 2008). Resource-poor farmers in marginal environments often manage agrobiodiversity to meet a multitude of agroecological and use objectives(Bellon 1996; Wood 1997; Jarvis et al. 2008). Over time, crop diversification thus increases the pool of options and lower transaction costs in accessing crop diversity locally. This notion is possible given that crop genetic diversity is an impure public good with both private and public economic attributes (Smale et al. 2003; Di Falco et al. 2010).

Crop-based adaptation, such as crop diversification, switching towards varieties resistant to climate stress and adopting new crop cultivars, require seed systems that are fit for purpose (McGuire and Sperling 2013; Nordhagen and Pascual 2013). Such seed systems must support farmers by ensuring continuous access to sufficient quantities of good quality seed and planting materials of preferred crop varieties (FAO and ECHA 2015). When farmers have access to quality seeds of well-adapted varieties that meet their needs and preferences, over time, they can be said to be seed secure (FAO 2018). In addition, maintaining a highly diversified cropping portfolio over time also requires that farmers face lower transaction costs in accessing seeds through available channels, including seed markets. In contexts where seed markets are thin, missing, or incomplete, as with factor markets in SSA (Barrett 2008; Markelova et al. 2009; Ricker-Gilbert and Chamberlin 2018), farmers face high transaction costs that constrain access to seed. Thus, crop-based adaptation to climate change relies on well-adapted crops and varieties (technology) and well-functioning seed systems that enable access to a wide variety of preferred seeds at lower transaction costs. The seed systems farmers use often mix crop types and crop varieties to serve different needs and risk-reducing strategies in the households (Bellon and Hellin 2011; Westengen et al. 2014). Thus, understanding the evolution of crop diversification and crop diversification responses to rainfall shocks is important for policies aimed at building smallholder farming systems that are resilient to shocks.

This paper focuses on (a) smallholder farm-level crop diversification responses to rainfall shocks (b) and the influence of knowledge and experience gained from past crop diversification on subsequent crop diversification. Despite crop diversification being recognized as an effective crop-based adaptation method to climate risk, little attention in terms of empirical investigation has been given to the inter-temporal effects of rainfall shock exposure on household crop diversification decisions. There is also limited knowledge on the impacts of past crop diversification decisions on later crop diversification decisions (dynamics of crop diversification) under recurrent shock exposure. Accumulation of knowledge and experience from previous crop diversification decisions over time may reduce the transaction costs required to implement a diversified cropping portfolio in subsequent years and help the farmer achieve successful diversification to buffer future risk. Hence, it is interesting to find out to what extent previous crop diversification decisions (knowledge and experience) enhance subsequent crop diversification decisions using panel data. By so doing, we can investigate indirectly to what extent transaction costs in implementing a diversified cropping portfolio and related factors constrain or enhance later crop diversification decisions under increased climate risk. This study is an important contribution to the growing literature on household crop diversification decisions, including its key drivers and constraints (see Alobo Loison (2015), and Tacconi et al. (2022) for recent reviews) and livelihood benefits (see Feliciano (2019), Waha et al. (2022), for recent reviews), and investigate the evolution of crop diversification over time and the effects of both short- and long-term exposure to rainfall shocks on current crop diversification decisions. By using national-level panel datasets, we are able to address the role of temporal drivers of crop diversification at a large spatial scale. Furthermore, we do a heterogeneity analysis to explore how differential resource endowments (land and household assets) shape farmers' crop diversification responses to rainfall shocks and lagged crop diversification decisions (detailed in section 3). We believe empirical findings from such an evolution and heterogeneity analysis study will inform policies that will help smallholder farmers to enhance or maintain highly diversified cropping portfolios over time to help them adapt to climate risk.

To achieve our objectives, we focus on smallholder farmers in Malawi and Tanzania. We use balanced household panel survey data for rural farmers in Malawi and Tanzania from the Living Standards Measurement Surveys -Integrated Surveys on Agriculture (LSMS-ISA) combined with historical weather data from WorldClim (Fick and Hijmans 2017). For Malawi, we constructed a four-round balanced panel household survey data from three survey rounds 2010, 2013, 2016, and 2019. Similarly, we construct a three-round balanced household panel data set from the Tanzania LSMS-ISA data collected in 2009, 2011, and 2013. The household panel survey data is suitable for this study as it collects comprehensive information on crop production activities and can be combined with weather data for analysis. The crop count and Simpson index are used to measure crop diversity. We define short-term measures of rainfall shocks (one-year lags of drought and flood shocks) and long-term rainfall variability (standard deviation average for 38 years) for the main crop growing season and use them as measures of rainfall shocks. In addition, we define lagged crop diversification indices and use them as proxies for experience and knowledge in implementing a diversified cropping portfolio (gained from past experiences) that marginally reduces transaction costs over time. We analyze balanced household panel data using household correlated random effects and dynamic random effects estimation methods that control for unobserved heterogeneity in household crop diversification decisions plus unobservable initial conditions that may influence crop diversification across space and time.

The rest of the article is organized as follows: the next section (2) outlines the study's theoretical framework. Section 3 outlines the empirical approach, while section 4 presents the results. Section 5 discusses the results, while section 6 concludes the article and proffers policy implications.

2. Theoretical framework

Household farming decisions such as crop diversification responses under rainfall shocks can be studied within the state-contingent production framework proposed by Chambers and Quiggin (2000) and Quiggin and Chambers (2006). This is evidenced in previous literature that has successfully studied the adoption of agricultural technologies under climate risk, including improved seeds (Holden and Quiggin 2017; Katengeza et al. 2019a; Gebru et al. 2021), cash crops (Gebru et al. 2021), integrated soil fertility management technologies (Katengeza et al. 2019b), and land rental market participation (Gebru et al. 2019; Tione and Holden 2021b), to mention a few examples. Farming households endowed with assets, labor, and land who aim to maximize crop production utility based on beliefs about the likelihood and production outcomes under different states of nature make state-contingent decisions (Quiggin and Chambers 2006; Holden and Quiggin 2017). According to this theory, farmers make input decisions before weather conditions for that season are revealed, based on their beliefs, preferences, and expectations (Quiggin and Chambers 2006; Holden and Quiggin 2017). Over time, exposure to different states of nature helps the farmer build more realistic expectations about the performance of alternative farming technologies that may influence adoption and adaptation processes. In other words, households gain experience over time that helps them shape their subjective production risk assessment, farming input choices, and consumption decisions, ex-ante and ex-post the production period (Quiggin and Chambers 2006; Dercon and Christiaensen 2011).

As in the state-contingent theory, we consider a farming household making crop production decisions under uncertainty. Rainfall variations present different states of nature (s), which are not known to the farmer when making input decisions. Based on the probability (π_s) of state (s) occurring, and on input choices (v), farming households can realize state-contingent output (q_s) in a set of positive real numbers ($\Re_+^s$). Following Holden and Quiggin (2017), the household production function (K), where different input allocations can lead to different production costs and output amounts, can be expressed as follows:

$$K = [(v, q): v \text{ can produce q}] \qquad \text{(Eq1)}$$

Under different states of nature (s), the farming household aims to minimize production costs subject to the state-contingent output (q_s). Assuming the price of inputs is denoted by p, we can present the household cost function as in equation 2 (Eq2):

$$C(p, q) = \min\,[p \times q: (v, q)] \qquad \text{(Eq2)}$$

Using first-order necessary conditions (Kuhn-Tucker conditions) and minimizing the cost function, the input demand function can be expressed as follows (Eq3):

$$v(p, q) = argmin[p \times q: (v, q) \text{ is in K}] \qquad \text{(Eq3)}$$

The input demand function in (Eq3) provides a complete characterization of the farm household production function(Holden and Quiggin 2017). If we assume a simple case of only two states of nature, where a household faces an unfavorable state in the form of flood or drought shocks (q_1), compared to a normal rainfall season (q_2), the cost minimization function for the less favorable expected output function is specified as in equation 4 (Eq4), where $q_1 < q_2$.

$$q^* = min\left[C(p, q): \pi_1 \times q_1 + \pi_2 \times q_2 = \underline{q}\right] \qquad \text{(Eq4)}$$

According to Holden and Quiggin (2017), if the vectors of state-contingent output have the same mean, the order may depend on associated risk in input choices. In such a setting, it is possible to distinguish between input choices that are risk-substituting and those that are risk-complementary. The implication is that farm households allocate input resources to manage risk, given their preferences and price expectations (Holden and Quiggin 2017). An exogenous increase in the probability of having a less favorable state of nature (drought or flood) will likely increase the share of risk-substituting inputs in the input mix for a given expected output. In the context of this paper, an increase in the probability of a less favorable state of nature, such as drought or flood, may increase crop diversification. Crop diversification is a state-contingent risky farming decision, given that households make their costly investment decision before the state of nature is revealed. Following the state-contingent theory discussed above, we aim to test the hypothesis that:

H1: Recent (past) exposure to drought or flood shocks and long-term rainfall variability increase crop diversification.

However, the decision by the farmer to diversify crop production will not be influenced by production factors alone but also by consumption factors. This notion is also in line with other theoretical frameworks often used to study farming decisions and outcomes in contexts where markets are missing or imperfect, including the Sustainable Livelihoods Framework(SLF) (Carney 1998; Scoones 1998; Ellis 2000) and Agricultural Household Models(AHM) (Singh et al. 1986; De Janvry et al. 1991; Behrman 2000). In brief, the SLF comprises interactions of various components, including household assets, livelihood strategies/activities, contextual factors(e.g., institutions and social relations), vulnerability context (i.e., exposure to external shocks and internal household changes), and livelihood outcomes (e.g., food security) (Carney 1998; Scoones 1998; Ellis 2000). Households use their assets (physical, natural, human, social, and financial) in both on-farm and off-farm activities to make a living while contextual factors (e.g., social relations, institutions) and exposure to covariate (e.g., climate shocks), and idiosyncratic shocks (e.g., changes in asset base) directly influence how these assets influence livelihood strategies and outcomes. In the balance, the combinations and recombinations of household resources (assets), institutions that govern access to those assets, and shocks (covariate and idiosyncratic shocks) determine production relations in smallholder farming (Binswanger and Rosenzweig 1986). Likewise, Agricultural household models (Singh et al. 1986; De Janvry et al. 1991; Behrman 2000) stress that semi-commercial farms that produce multiple crops and livestock often combine two fundamental units of economic analysis (the household and the firm). These agricultural household models can be independent (recursive) or dependent (simultaneous) (Singh et al. 1986). Recursive agriculture household models treat production and consumption decisions as independent, implying that farm households can be modeled as pure profit maximizers. However, household production and consumption decisions are inseparable indicating that a production or profit maximization model would not adequately describe the decision-making process (Singh et al. 1986; De Janvry et al. 1991; Caviglia-Harris 2004). Non-separability of production and consumption decisions, in other words, implies that asset distribution and consumption needs may significantly impact production decisions and the management of land and labor (Caviglia-Harris 2004). Hence, production and consumption-related factors are essential when analyzing smallholder farming decisions such as crop diversification.

The smallholder farmer's decision to diversify crop production could be driven by the need to respond to market imperfections common in SSA (Ellis 2000; Dercon and Christiaensen 2011; Alobo Loison 2015). For instance, crop diversification decisions may respond to high transaction costs that characterize factor markets in SSA. Due to market imperfections in SSA, market access is not uniform because households may face different transaction costs (Renkow et al. 2004; Barrett 2008). Non-linear transactions costs in SSA are high, and they emanate from policies, institutions, and social factors that influence the degree of information asymmetry and access to productive resources (Fafchamps 2004; Holden et al. 2010; Ricker-Gilbert and Chamberlin 2018; Gebru et al. 2019; Tione and Holden 2021a). By definition, transaction costs are the costs incurred in making a market transaction, excluding the actual price paid for the commodity(Coase 1960; North 1987). More precisely, the costs might include three cost components including (i) searching and attracting potential trading partners, including pre-sale inspection, (ii) negotiation, contracting, and fulfillment costs, and (iii) monitoring and implementation costs(Coase 1937; North 1987). In implementing a diversified cropping portfolio, such transaction costs may include all costs incurred in acquiring crop seeds and complementary inputs required to implement a diversified cropping portfolio. For instance, in acquiring crop seed through formal or informal markets, such transaction costs may include the costs of searching and obtaining information on production and consumption traits of the seed of different crops, costs of searching and locating them, and the costs of negotiating for the seeds. Given this background, we aim to test our second hypothesis that:

H2: Crop diversification decisions are state-dependent. This implies that lagged crop diversification decisions strongly and positively explain later crop diversification decisions.

Lagged crop diversification decisions may proxy experience and knowledge gained from implementing crop diversification in the past, which could prove to be important in marginally reducing transaction costs in enhancing crop diversification in the future. Also, the contribution of on-farm agrobiodiversity to future access to seed may contribute to state dependency in crop diversification decisions. Farmers who have maintained a diversified cropping portfolio in the past may have better access to seeds to support later crop diversification, unlike their opposite counterparts. We, hence, expect to find state-dependent crop diversification decisions when analyzing farm household panel data capturing crop diversification over time.

The impact of recent rainfall shocks, long-term rainfall variability, and lagged crop diversification decisions are likely to be heterogeneous for farmers in different socioeconomic strata. This is a plausible view for a few reasons. First, the vulnerability of smallholder farming households to shocks is closely associated with their resource poverty, such that resource-poor households without social safety nets are considered vulnerable (Dercon 2004; Dercon 2005). Second, evidence from the literature suggests that it's the relatively better-off smallholder farmers with sufficient assets who achieve successful livelihood diversification(Ellis 2000; Alobo Loison 2015). Hence, the behavioral impact of rainfall shocks and past diversification decisions may significantly differ among the rich and the poor (as defined by their relative land and non-land asset endowments). Following the state-contingent theory and the brief discussions above, we aim to test the sub-hypotheses that:

H3: Households better endowed with assets (land and household assets) are more likely to (a) intensify crop diversification to help them deal with rainfall shocks (i.e., drought shocks) and (b) capitalize on the experience gained from past diversification decisions and use it to diversify production overtime to achieve multiple objectives including dealing with climate risk, relative to their poorer counterparts.

3. Methods

3.1. Data sources

In this study, we use a combination of household survey data and historical weather data to study household crop diversification responses to rainfall shocks in Tanzania and Malawi, which we describe below.

3.1.1. Household Survey Data

We use the household survey data from the Living Standards Measurement Study- Integrated Surveys on Agriculture (LSMS-ISA) available through the World Bank. The household LSMS-ISA surveys adopt a multi-sector approach which allows understanding the links between agricultural production activities, socio-economic status, and off-farm activities. The LSMS-ISA data collect comprehensive information on agricultural activities and household socio-economic conditions in respective countries. In this study, we use household survey data from Malawi (Malawi (Integrated Household Survey (IHS)), and Tanzania (Tanzania National Panel Survey (TNPS)). We specifically use four rounds of the Malawi LSMS-ISA collected in 2009/10, 2012/13, 2015/16, and 2018/19. The panel surveys for Malawi started with 1 619 households, with about 71% (1 144) of them being rural households traced in four successive rounds. We focus on rural households, and from the initial 1 144 rural households, we construct a balanced panel data of 971 households, with consistent household information and usable information on cropping activities in all four rounds. For Tanzania, we use three rounds of the household survey data conducted in 2008/9, 2010/11, and 2012/13. The initial round of Tanzania LSMS-ISA data surveyed about 3 265 households, of which about 63% (2 063) were rural households. We trace rural households with consistent household information and usable data on cropping activities to construct a balanced panel of 1 675 rural households, which we use for analysis. Overall, we have 3 884 and 5 025 total observations for Malawi and Tanzania, respectively. Loss of respondents from baseline to subsequent rounds, if systematic, could lead to bias in estimation. We hence test and control for possible attrition bias in our estimation. In later sections (estimation strategy), we give more details on steps taken to test and control potential attrition bias in our analysis. We use the LSM-ISA household data to define the socio-economic characteristics of studied households and, most importantly, to define crop diversification indices.

We use the different crops grown by the household and the approximate area allocated to each crop to compute our indices for crop diversification. The LSMS-ISA for Malawi and Tanzania collects information on the different crops grown by farmers following a consistent crop classification across survey rounds. The information on crop production is collected for the latest completed primary crop growing season. We compute two indices of crop diversification: the crop count index (number of crops grown) and the Simpson index of crop diversification. We use both indices to define and describe crop diversification in studied countries. The Simpson index (SI) of crop diversification (Simpson 1949) is a composite and commonly used measure of diversification, and it measures both crop species richness (count) and evenness (relative abundance of crop species). We run our regression models using both the Simpson and count index as our main outcome variables. We estimate the Simpson index (SI) of crop diversification as follows: $\mathrm{SI_j}=1-\sum_{c=1}^{n}\left[\frac{\mathrm{fs_c}}{\mathrm{fs}}\right]^2$, Where; SI_j is the Simpson index for farmer j, fs_c is the approximate area devoted to crop c, and fs is the total farm size. The Simpson index ranges from 0 to 1, and larger(lower) indices indicate high(low) levels of crop diversification. The index equals one under complete diversification and zero under complete specialization. We provide summary statistics of crop diversification indices in *Table* 1.

3.1.2. Weather data

We use georeferenced data for primary sampling units(clusters) available with LSMS-ISA household survey data to extract historical monthly time series weather data from WorldClim[1] (Masarie and Tans 1995; Fick and Hijmans 2017), which we used to define rainfall variability and lagged rainfall shocks. The LSMS-ISA household data provide the approximate location (longitude and latitude) of clusters (villages) from which interviewed households live. We take advantage of such georeferenced data to extract and process rainfall data,

[1] https://www.worldclim.org/data/monthlywth.html

which we combine with household data for analysis. The distribution of clusters from which households included in our balanced panels were sampled is shown in *Figure 1*.

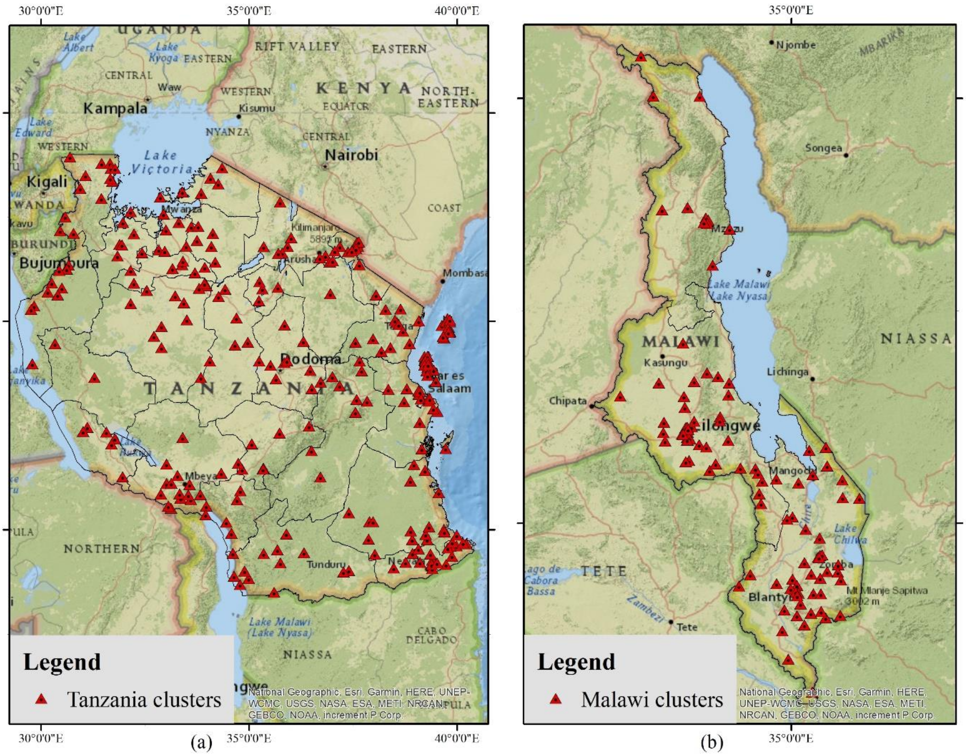

Figure 1:Geographical distribution of rural clusters included in the Tanzanian(a), and Malawian(b) balanced panel samples analyzed in this study.

The WorldClim data are available at high resolutions[2] and are bias-corrected by the Climatic Research Unit, University of East Anglia (Harris et al. 2014; Fick and Hijmans 2017). We focus on lagged rainfall shocks and rainfall variability which we define using the steps described below. We start by defining the main crop growing season, which runs from November to April in both countries, and then we define our variables of interest for the main crop growing season. To better reflect on conditions during the most important season for food production we define rainfall shock variables for the main growing season. We use one-year lags of flood and drought shocks, and long-term rainfall variability as our main variables measuring rainfall shocks. We define long-term rainfall variability as the average rainfall variability (standard deviation) for rainfall for the reference period (1980 to 2018). The distribution of average rainfall and its standard deviation for the growing season for the studied countries is shown in the top panel of *Figure 2*. From *Figure 2*, we can see more variation in average rainfall and its variability in Tanzania compared to Malawi. Descriptive statistics of rainfall shock variables are presented with descriptive statistics of other variables in Table 1. We define rainfall shocks for the main growing season as normalized rainfall deviations last season (1 -year lag) from seasonal rainfall variables over a reference period following previous studies(Michler et al. 2019; Bora 2022). We define lagged drought(flood) shocks as normalized negative(positive) standard deviations in a single season's rainfall from the seasonal climate variable over the reference period (1980-2018). We hence first define a rainfall shock variable that shows both positive and negative normalized rainfall standard deviations (Z-Scores) as follows: $Rain\_shock_{vt} = \left[\frac{Rain_{vt}-\overline{Rain_v}}{\sigma_{Rain_v}}\right]$, where $Rain\_shock_{vt}$ is a rainfall shock measure for a village (v) in the year (t), $Rain_{vt}$ is the observed amount of rainfall for the defined period (season), $\overline{Rain_v}$ is the average seasonal rainfall for the village (v) over the reference period (1980-2018), and σ_{Rain_v} is the standard deviation of rainfall during

[2] The WorldClim data we use is at a spatial resolution of 2.5 minutes (approximately ~21 km2)

the same period. Given our interest in testing the influence of specific rainfall deviations (negative and positive deviations) we split the rainfall shock variable into negative and positive rainfall deviations, which we term drought and flood shocks, respectively. We do the splitting as follows: (a) Drought shock: $\underline{DS}_{vt} = \left\{\left[\frac{Rain_{vt}-\overline{Rain}_v}{\sigma_{Rain_v}}\right] if\ Rain_{vt} < \overline{Rain}_v, and\ 0\ otherwise\right.$, and, (b) Flood shock: $\overline{FS}_{vt} = \left\{\left[\frac{Rain_{vt}-\overline{Rain}_v}{\sigma_{Rain_v}}\right] if\ Rain_{vt} > \overline{Rain}_v, and\ 0\ otherwise\right.$, where $\underline{DS}_{vt}$ ($\overline{FS}_{vt}$) is a measure for a drought (flood) shock in a village (v) at a time (t).

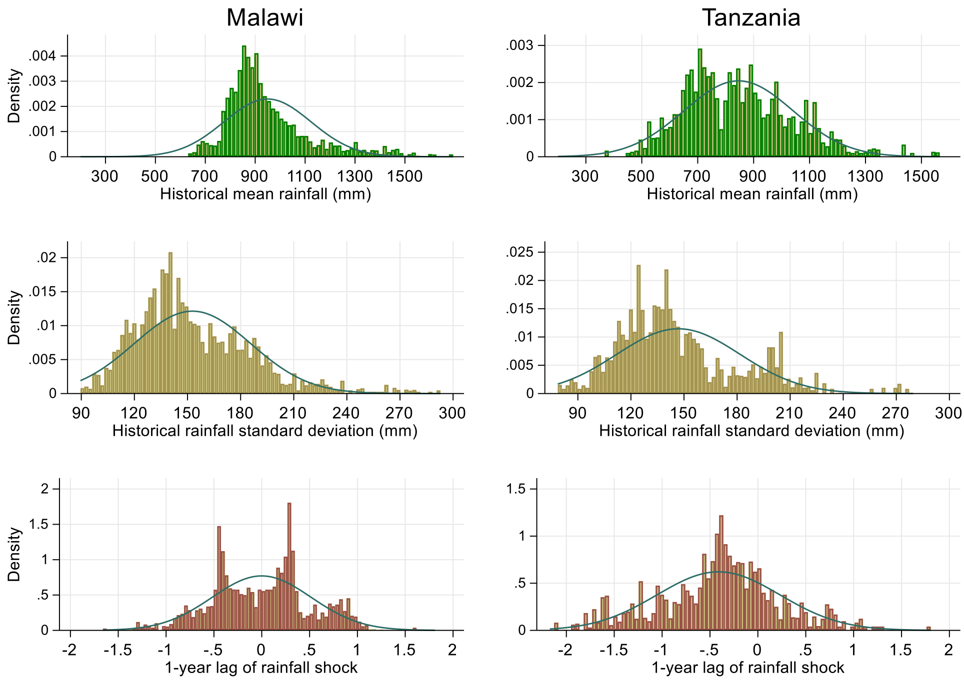

Figure 2: Distribution of historical (1980-2018) mean rainfall and its standard deviation and 1-year lag of rainfall shock (Z-score) for the main crop Growing Season (GS) (November to April) in the studied countries based on WorldClim Data. Plotted are histograms(with hundred bins) with a normal curve overlay.

In the bottom panel of Figure 2 we show the distribution of the 1 -year lag of normalized rainfall Z-score (rainfall shock) in the respective country pooled samples. We also see a larger variation in rainfall shocks in Tanzania than in Malawi, as with average rainfall and its standard deviation. Additional summary statistics for climate variables and shocks are given with descriptive statistics in Table 1.

3.2. Empirical Estimation Strategies

We model crop diversification decisions (D_{it}) using both Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) models (Wooldridge 2005; Wooldridge 2010). DRE and CRE models apply to limited dependent variables, such as the count index and Simpson index of crop diversification. For the count index, we specify DRE and CRE Poisson models, while for the Simpson index, we specify DRE and CRE Tobit models. The CRE and DRE models incorporate our key variables of interest, from which we test our hypothesis.

3.2.1. The Correlated Random Effects (CRE) approach

We specify crop diversification (D_{it}) in the CRE framework as shown in equations 5:

$$D_{it} = \alpha + \beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta \bar{X}_i + \gamma \check{X}_i + a + \mu_i + \epsilon_{it} \qquad \text{[Eq5]}$$

Where D_{it} is the dependent variable measuring crop diversification index (Simpson index or count index) of household i at time t. DS_{t-1} , FS_{t-1}, and $V_i$are measures of one-year lag drought shock (DS), 1-year lag flood shock (FS), and long-term rainfall standard deviation(variability) respectively. $\bar{X}_i, \check{X}_i$, are respectively the means and deviations of observed household and farm characteristics (X_{it})(specified below). μ_i and ϵ_{it} are respectively time constant unobserved heterogeneity at the household level, and idiosyncratic error that is independent and identically distributed. In line with specified random effect models, we assume the errors to be additive(Wooldridge 2010). The vector a is for regional and time (panel year) dummies. The CRE framework in Equation 5(Eq5), was first suggested by Mundlak (1978) and Chamberlain (1982), and it is equivalent to using household fixed effects with continuous dependent variables. The CRE framework applies to nonlinear models and our study applies CRE Poisson for the count index, and CRE Tobit for the Simpson index(Wooldridge 2010). The CRE approach has the advantage that it helps avoid the incidental variables problem that fixed effects introduce in non-linear models. In line with the CRE approach, we assume that the unobserved heterogeneity can be replaced with its linear projection onto the time averages of all household level regressors(X_{it}) (Mundlak 1978; Chamberlain 1982). In implementing our CRE Poisson and Tobit regression, we, hence, control for means ($\bar{X}_i$), and deviations from means ($\check{X}_i$) of farm and household characteristics in Eq5. In CRE specifications (Eq 5) our parameters of interest are β, θ and δ for 1-year lag drought shock (DS), 1-year lag flood shock (FS), and growing season long-term rainfall variability (V) respectively.

3.2.2. The Dynamic Random Effects (DRE) approach

We also model crop diversification using the DRE approach. The DRE model specification is important in assessing the effects of covariate rainfall shocks, as it also controls for initial unobserved crop diversification conditions that may influence diversification across time and space. Following Wooldridge (2005), we specify DRE Poisson and Tobit models, as shown in equations 6 and 7.

DRE Poisson:

$$E\left(D_{it} | D_{i,t-n}, D_{i0}, DS_{t-1}, FS_{t-1}, V_i, X_{it}, a, \mu_i\right) = u_i exp\left[\beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta X_{it} + g\left(D_{i,t-n}\right)\rho\right] \quad \text{[Eq6]}$$

DRE Tobit:

$$D_{it} = max\left[0, \beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta X_{it} + \rho D_{i,t-n} + a + \mu_i + \epsilon_{it}\right] \qquad \text{[Eq7]}$$

Where D_{it} , DS_{t-1} , FS_{t-1}, V_i, and X_{it} are as described prior. $D_{i,t-n}$ represents initial (D_0) and previous survey round crop (D_{t-1}) diversification indices. The initial crop diversification index (D_0) remains the same for subsequent panel rounds and previous panel round crop diversification index (D_{t-1}) is the households' crop diversification index for the previous survey round. For Malawi, we use the survey years 2016, 2013, and 2010 as previous panel rounds for 2019, 2016, and 2013 panel rounds respectively. Similarly, for Tanzania, panel years 2011 and 2008 are used as previous panel rounds for 2013, and 2010 panel rounds respectively. In the DRE models, the statistical significance of ρ assess whether there is state dependency in crop diversification decisions. The initial hypothesis is that there is no state dependency (i.e., $\rho = 0$). In the DRE models, with limited dependent variables, the household time-invariant unobserved heterogeneity (μ_i) is also additive and can be expressed as follows:

$$\mu_i = \tau_0 + \tau_1 D_{i0} + \tau_2 X_i + \epsilon_i. \qquad \text{[Eqn8]}$$

Where $\epsilon_i \sim Normal(0, \sigma_\epsilon^2)$ and is independent of $(\tau_1 D_{i0} + \tau_2 X_i)$. τ_0 is a constant. The DRE approach, as mentioned earlier allows us to control for unobserved effect (μ_i) and initial household conditions that are likely to influence crop diversification in subsequent years, including transactions costs. Hence in addition to the parameters of interest β, θ , and δ we have an additional parameter of interest ρ in the DRE specifications which allows us to test for state dependence in crop diversification decisions.

Unlike with the CRE specifications where we use data for all panel rounds (four in Malawi, and three in Tanzania), in the DRE approach, we lose the baseline survey round per country because we do not have lagged crop diversification indices for them. Hence total observations in the DRE models in Malawi and Tanzania are respectively 2 913, and 3 350, compared to 3884 and 5025 in the CRE approach.

The household-level characteristics we include in vector (X_{it}) in our CRE and DRE models include household wealth endowments (farm size(ha), household labor units (elaborated below), and asset wealth index (elaborated below)), characteristics of the household head (age(years), education (at least secondary education(1=yes)), sex (1=female; 0 otherwise), marital status (1= single; 0 otherwise), the household age dependency ratio(%), and distance to agricultural markets(km). For labor units, we define male adult-equivalent labor units where we assign 1, 0.8, and 0.5 to an adult male, adult female, and children between 5 and 15 years of age, respectively. We consider household members available within the household for at least a month within a calendar year. For household asset wealth, we combine information on household ownership of durable non-land assets (e.g., agricultural equipment and machinery) and household dwelling characteristics common in each country to create the household asset wealth index, using Principal Components Analysis (PCA). The first principal component of a set of variables in PCA is the linear index of all the variables that captures the largest amount of information common to all the variables (asset indicators) and is then kept and used as a proxy for household asset wealth. The resulting asset wealth index (or score) can take both negative and positive values, and the increasing(decreasing) value of the index shows higher(lower) relative households' asset wealth endowments. However, for ease of comparison and interpretation of the index, we normalize the household wealth index from PCA using the unity-based normalization method as follows: $Vnorm = \frac{V - \text{minimum } (V)}{\text{maximum}(V) - \text{minimum } (V)}$, where V is the original wealth index from PCA(which includes both negative and positive values), and $Vnorm$ is the normalized version of the index which is bounded between 0 (min) and 1(max). For more details on the technical explanation of such asset-based household wealth indicators, readers can refer to Filmer and Pritchett (2001b) and Filmer and Pritchett (2001c); Filmer and Pritchett (2001b); Filmer and Pritchett (2001a); McKenzie (2005).

3.2.3. Robustness checks: Testing and controlling for attrition bias in estimation

Attrition, if systematic, may bias our results. As a robustness check to our main results, we test and control for potential attrition bias because of dropping out of households in subsequent rounds. To handle possible attribution bias effect, we follow the following steps: First; we estimate probit attrition models for respective countries with dummy variable (1=yes) for households not observed in the follow-up surveys (2013 for Malawi, and 2011 for Tanzania), and zero otherwise, using household characteristics at baseline as explanatory variables. We present results from the attrition probit models as part of the Appendix (Table F). From the attrition probit results, we see that some household characteristics were significant in explaining the probability of dropping out, showing that attrition was not purely random, which could lead to bias. Second, we construct an Inverse mills ratio (IMR) from the attrition probit models. The IMR we construct becomes a time-invariant variable in our balanced panel data set, as households keep the same value of IMR across panel rounds. Third, we use the constructed IMR to test and control for the potential attrition bias effect by including it as an additional explanatory variable in our dynamic random effect probit and Tobit models. The IMR was not significant in any of the models in Malawi, which suggests that attrition bias was not an issue to worry about. However, in Tanzania, the IMR was significant, showing that attrition bias was significant, and we hence report estimates adjusted for attrition bias in the manuscript for Tanzania. We present full tables of results from this exercise (where we test and control for potential attrition bias) in Appendix (Tables G to H). In both Malawi and Tanzania, comparing results with and without attrition bias correction leads to the same conclusions, showing that our findings are all robust to attrition.

3.2.4. Heterogeneity analysis

Besides the main effects of rainfall shocks and lagged crop diversification decisions on subsequent diversification, we explore heterogeneity effects. In heterogeneity analysis, we seek to understand whether households' responses to rainfall shocks and past diversification decisions vary with the households' wealth as defined by their land and non-land asset endowments. We believe, going beyond the "average" farmer and understanding how heterogeneity in constraints and opportunities faced by well-endowed and poorly endowed farmers shape their behavioral response to risk over time is crucial for the design and implementation of better-targeted interventions aimed at improving resilience. We, hence, estimate our CRE and DRE models specified earlier in equations 5, 6, and 7 in split samples of household's resource endowment quintiles (q). For each of the variables that we use to signify resource endowments (total farm size (ha), and asset wealth index (PCA)). We define three quintile categories (q) (1=low, 2=medium, 3=high) asset wealth endowments which we then use to split our samples into subgroups. In other words, we specify our CRE and DRE specifications by resource endowment quintile category as follows:

CRE Models by resource endowment quintile (q):

$$D_{it}^{q} = \alpha + \beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta \bar{X}_i + \gamma \check{X}_i + a + \mu_i + \varepsilon_{it} \qquad \text{[Eq9]}$$

DRE Poisson by resource endowment quintile (q):

$$E\left(D_{it}^{q} | D_{i,t-n}^{q}, D_{io}^{q}, DS_{t-1}, FS_{t-1}, V_i, X_{it}, a, \mu_i\right) = u_i exp\left[\beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta X_{it} + g\left(D_{i,t-n}^{q}\right)\rho\right] \quad \text{[Eq10]}$$

DRE Tobit by resource endowment quintile (q):

$$D_{it}^{q} = max\left[0, \beta DS_{t-1} + \theta FS_{t-1} + \delta V_i + \vartheta X_{it} + \rho D_{i,t-n}^{q} + a + \mu_i + \varepsilon_{it}\right] \qquad \text{[Eq11]}$$

Where the superscript (q) in Equations 9, 10, and 11 is 1, 2, or 3 for low, medium, and high resource endowments (based on total farm size, or household asset wealth score) in all our CRE and DRE specifications. Variables and key parameters of interest are as described earlier. We follow the same procedure for farm size and household asset wealth quintiles.

4. Results

4.1. Descriptive stats

In Table 1, we present descriptive statistics of our key variables in this study. We present average crop diversification indices per survey round. Based on the number of crops grown and the Simpson index, crop diversification has slightly increased over time in Malawi. In Malawi, the number of crops grown increased by about 2 crops from the first panel round (2009/10) to the fourth panel round (2018/19). Based on the trends shown in Figure 3, we can see also that the share of farmers growing over four crops has increased from less than 20% in the 2009/10 panel round to about 40% in the 2018/19 round. The share of households showing to have grown a single crop fluctuates over time but is lowest in the latest round (2018/19). Also, the Simpson index of crop diversification, as given in Table 1, rose from 0.47 to 0.62 within the same period in Malawi, which shows that crop diversification in the analyzed sample has increased. The most common crops in the Malawi balanced panel analyzed include: maize, groundnut, pigeon pea, pumpkin leaves (nhakwani), sorghum, tobacco, common bean, soybean, sweet potato, rice, finger millet, pearl millet, cowpeas, cotton, and sunflower (Table B in Appendix)).

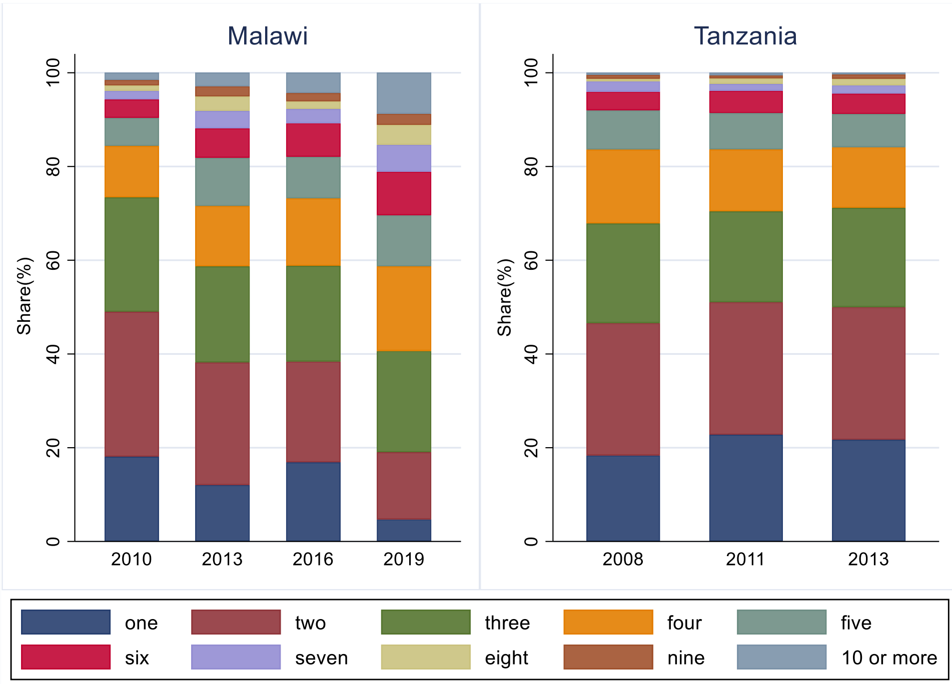

Figure 3: The number of crops grown by survey year in studied countries. Source: Authors' elaboration based on LSMS-ISA data. The figure shows the share of farmers growing a given number of crops from one crop to ten or more crops and changes over time.

In Tanzania, we observe quite uniform trends in the shares of farmers growing a specific number of crops. Over 80% of farmers in the Tanzania sample grew between one and four crops, and the average number of crops grown per household remained stable at about 3 crops across panel rounds (Table 1). The Simpson index slightly fell from 0.44 in the 2008/9 round to about 0.4 in the 2012/13 round. The most common crops grown by households in the Tanzania panel are maize, common bean, rice, groundnut, sorghum, sweet potato, sunflower, pigeon pea, cowpeas, cassava, sesame, pearl millet, finger millet, cotton bambara nuts, and tobacco (Table C in Appendix).

Table 1: Descriptive statistics of key variables used in the analysis

	Malawi				Tanzania		
	2010	2013	2016	2019	2009	2011	2013
VARIABLES	mean	mean	mean	mean	mean	mean	mean
Crop diversification indices (D_{it})							
Count Index (Number of crops grown)	3.09	3.71	3.72	4.88	2.99	2.90	2.91
Simpson index	0.47	0.55	0.46	0.62	0.44	0.39	0.40
Lagged Crop diversification indices ($D_{i,t-n}$)							
Count index at baseline (D_0)	3.09	3.09	3.09	3.09	2.99	2.99	2.99
Count index in previous panel round (D_{t-1})	.	3.09	3.71	3.72	.	2.99	2.90
Simpson index at baseline (D_0)	0.47	0.47	0.47	0.47	0.44	0.44	0.44
Simpson index in previous panel round (D_{t-1})	.	0.47	0.55	0.46	.	0.44	0.39
Rainfall shocks							
Rainfall shock growing season (1-year lag)(z-score)	0.75	-0.33	-0.32	-0.03	-0.04	0.45	-0.52
Longterm growing season(Nov-April) rainfall variability (1980-2018)(mm)	143.60	143.60	143.60	143.60	146.65	146.65	146.65
Long-term early season(Nov-Jan) rainfall variability (1980-2018)(mm)	87.01	87.01	87.01	87.01	98.99	98.99	98.99
Household resource endowments							
Farm size (ha)	0.73	0.74	0.78	0.77	2.05	2.11	1.89
Household labor units	2.88	3.24	3.18	3.30	3.28	3.51	3.34
Household asset wealth index(normalized)	0.13	0.13	0.12	0.14	0.30	0.29	0.28
Observations	971	971	971	971	1675	1675	1675

Statistics are not weighted; source (own calculation from LSMS-ISA data for Malawi and Tanzania): The rest of the control variables used are not shown here for brevity but are available with the supplementary material (Table A).

Descriptive statistics for the rest of the key explanatory variables used are shown in the bottom panel of Table 1. On average, farming households work on smaller farm sizes of about 0.76 ha in Malawi compared to about 2.0 ha in Tanzania. In addition, farmers in the Tanzanian sample have slightly more labor units (between 3.3 and 3.5) compared to Malawi (between 2.9 and 3.3). In terms of household asset wealth, farmers in Tanzania, on average, have more household assets, as shown by a normalized asset score of 0.287 compared to 0.132 in Malawi (Table 1). Descriptive statistics for the rest of the control variables considered in the analysis are shown in Appendix (Table A).

In Table 2, we show descriptive statistics of capital endowments (farm size, and household asset wealth) and crop diversification indices (count and Simpson index) by quintile of asset endowments. From the descriptive statistics on crop diversification indices in Table 2, and Figure 4, we see that average diversification indices increase with quintiles of asset wealth (Q1 to Q3) suggesting that better endowed (Q3) households in both Malawi and Tanzania sample are highly diversified compared to their poorer counterparts (Q1).

Table 2:Descriptive characteristics of resource endowments, and crop diversification indices in different quintile categories

	Malawi				Tanzania			
	Q1	Q2	Q3	Pooled	Q1	Q2	Q3	Pooled
	mean	mean	mean	mean	mean	mean	mean	mean
Farm size quintiles (Q1=low; Q2=medium; Q3=High)								
Farm size (ha)	0.26	0.63	1.38	0.75	0.26	1.16	4.73	2.02
Count Index (Number of crops grown)	2.85	3.79	4.91	3.85	2.30	2.94	3.52	2.93
Simpson index	0.43	0.54	0.60	0.52	0.32	0.42	0.49	0.41
Observations	*1295*	*1297*	*1292*	*3884*	*1746*	*1634*	*1645*	*5025*
Household wealth quintiles (Q1=low; Q2=medium; Q3=High)								
Household asset wealth index (score)	-1.35	-1.04	0.69	-0.57	-2.17	-0.66	2.59	-0.08
Household asset wealth index (normalized)	0.01	0.06	0.32	0.13	0.08	0.23	0.55	0.29
Count Index (Number of crops grown)	3.43	3.98	4.15	3.85	2.59	3.03	3.18	2.93
Simpson index	0.49	0.54	0.54	0.52	0.39	0.42	0.42	0.41
Observations	*1313*	*1277*	*1294*	*3884*	*1675*	*1675*	*1675*	*5025*

Notes: source (own calculation from LSMS-ISA data for Malawi and Tanzania):

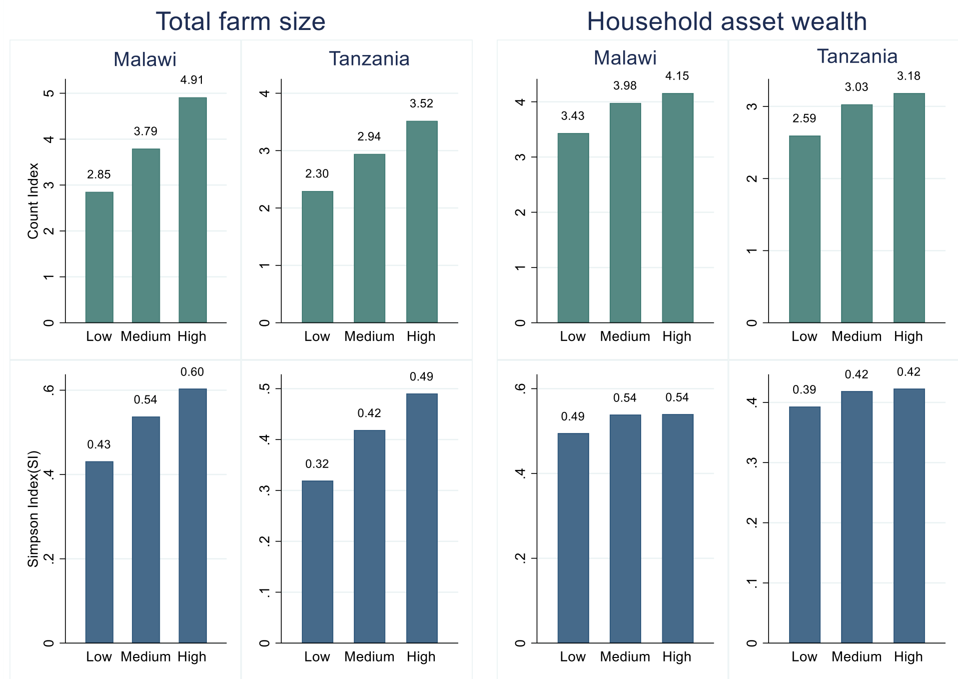

Figure 4: Mean crop diversification indices by quintiles of farm size and household asset wealth endowments. The left panel compares average crop diversification indices (count and Simpson Index) by farm size quintiles, while the right panel compares the two indices of diversification by household asset wealth quintiles. Low, medium, and high denote the first, second, and third quintiles of increasing endowments as described earlier in the manuscript.

4.2. Main results

4.2.1. The impact of lagged rainfall shocks and variability on crop diversification

We present the main results from our Dynamic Random Effects (DRE), and Correlated Random Effects (CRE) Poisson and Tobit Models of crop diversification for Malawi and Tanzania, in Table 3, and Table 4, respectively. We show average partial effects (APE) here and present more detailed results showing coefficients results in the attached appendix.

Results show that a 1-year lag of drought shock enhances crop diversification in both Malawi and Tanzania. On average, a unit increase in the 1-year lag drought shock increases the number of crops grown by 0.03, and the Simpson index by 0.02 (Table 3). In Tanzania, on average, we find a unit increase in the 1-year lag drought shock to enhance the number of crops grown by about 0.04 and 0.06 and the Simpson index 0.01 and 0.02 in the CRE and DRE models respectively (Table 4).

Table 3: Crop diversification decisions in Malawi: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models reporting Average partial effects (APE)

	Poisson models		Tobit Models	
Variables	**DRE**	**CRE**	**DRE**	**CRE**
Rainfall shocks and variability				
Flood shock growing season 1-year lag (z-score)	0.049	0.003	0.029	0.014
	(0.0478)	(0.0462)	(0.0207)	(0.0198)
Drought shock growing season 1-year lag (z-score)	0.004	0.025^{*}	0.003	0.016^{***}
	(0.0176)	(0.0146)	(0.0077)	(0.0060)
Long-term season (Nov-April) rainfall variability (1980-2018) (mm)	0.008^{***}	0.007^{***}	0.002^{***}	0.002^{***}
	(0.0009)	(0.0005)	(0.0003)	(0.0002)
Lagged Crop diversification indices ($D_{i,t-n}$)				
Count (Simpson) index at baseline (D_0)	0.038^{***}		0.092^{***}	
	(0.0061)		(0.0220)	
Count (Simpson) index in previous panel round (D_{t-1})	0.023^{***}		0.068^{**}	
	(0.0050)		(0.0314)	
Household & farm characteristics				

Observed household characteristics	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes
Year dummies & regional dummies				
2019 (1=yes)	0.243***	0.378***	0.067***	0.110***
	(0.0236)	(0.0371)	(0.0102)	(0.0157)
2016 (1=yes)	-0.029	0.106***	-0.094***	-0.042***
	(0.0264)	(0.0377)	(0.0109)	(0.0151)
2013(1=yes)		0.136***		0.043***
		(0.0351)		(0.0147)
Northern region (1=yes)	0.037	0.051	-0.016	0.003
	(0.0624)	(0.0495)	(0.0228)	(0.0175)
Southern region(1=yes)	-0.002	0.117***	-0.012	0.001
	(0.0406)	(0.0318)	(0.0150)	(0.0115)
constant	0.288***	-0.221**	0.194***	0.013
	(0.1117)	(0.0961)	(0.0489)	(0.0427)
lnalpha/ sigma_u	-3.047***	-2.677***	0.050**	0.098***
	(0.1381)	(0.0848)	(0.0215)	(0.0066)
sigma_e			0.250***	0.254***
			(0.0055)	(0.0038)
Panel households	971	971	971	971
Observations	2913	3884	2913	3884

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively the number of crops grown and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parentheses are robust standard errors; Household and farm characteristics included as controls include farm size (ha), household labor units, asset wealth index (PCA), the household age dependency ratio (%), distance to agricultural markets (km), age (years) of household head, dummies for female and single-headed households, and a dummy for household head attaining at least Junior Certificate Examination (JCE). The reference for regions is the central region. The year 2010 and 2013 are reference dummies in CRE and DRE models, respectively. The full table of coefficients showing all control variables included in the DRE and CRE Poisson and Tobit models is shown in Supplementary material (Table D).*

Furthermore, we did not find the 1-year lag flood shock to significantly explain crop diversification decisions in Malawi, but it enhance crop diversification in Tanzania (Table 4). We also learn that, on average, long-term rainfall variability enhances crop diversification decisions in both Malawi and Tanzania. More precisely, we find that, on average, a unit (1 mm) increase in rainfall variability enhances the number of crops grown, and the Simpson index by about 0.01 and 0.002 respectively in Malawi (Table 3). In Tanzania, we also learn that a marginal increase in long-term rainfall variability enhances the number of crops grown by between 0.002 and 0.003 and the Simpson index of diversification by about 0.001 (Table 4).

Overall, results show that rainfall shocks, particularly drought shocks and long-term rainfall uncertainty (variability) enhance crop diversification in both Malawi and Tanzania. Based on these results, we could not reject our hypothesis that recent (past) exposure to rainfall shocks and long-term rainfall variability increase crop diversification. We expand on this important result in our discussions.

Table 4: Crop diversification decisions in Tanzania: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models reporting Average partial effects (APE)

	Poisson models		**Tobit Models**	
Variables	**DRE**	**CRE**	**DRE**	**CRE**
Rainfall shocks and variability				
Flood shock growing season 1-year lag (z-score)	0.074	0.072**	0.015	0.018*
	(0.0452)	(0.0359)	(0.0141)	(0.0111)
Drought shock growing season 1-year lag (z-score)	0.061***	0.041***	0.019***	0.014***
	(0.0185)	(0.0137)	(0.0057)	(0.0042)
Long-term season (Nov-April) rainfall variability (1980-2018) (mm)	0.002***	0.003***	0.001***	0.001***
	(0.0008)	(0.0006)	(0.0003)	(0.0002)
Lagged Crop diversification indices ($D_{i,t-n}$)				
Count (Simpson) index at baseline (D_0)	0.118***		0.190***	
	(0.0110)		(0.0267)	
Count (Simpson) index in the previous panel round (D_{t-1})	0.021**		0.046	
	(0.0101)		(0.0328)	
Household & farm characteristics (X)				
Observed household characteristics	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes
Year dummies & regional dummies (a)				
2013 (1=yes)	-0.059**	-0.083***	-0.011	-0.041***
	(0.0269)	(0.0235)	(0.0084)	(0.0073)
2011 (1=yes)		-0.051**		-0.038***
		(0.0227)		(0.0071)
Northern region(1=yes)	0.012	-0.001	0.000	0.011
	(0.0776)	(0.0750)	(0.0248)	(0.0235)

Coast region (1=yes)	-0.007	-0.031	-0.018	-0.008
	(0.0772)	(0.0746)	(0.0247)	(0.0234)
Central region (1=yes)	0.119	0.032	0.049*	0.032
	(0.0797)	(0.0773)	(0.0256)	(0.0241)
Lake region (1=yes)	0.008	0.015	0.005	0.027
	(0.0811)	(0.0785)	(0.0259)	(0.0245)
Southern Highlands (1=yes)	0.164**	0.162**	0.056**	0.069***
	(0.0817)	(0.0787)	(0.0263)	(0.0248)
West region (1=yes)	0.179**	0.068	0.053*	0.027
	(0.0868)	(0.0843)	(0.0278)	(0.0263)
Inverse mills Ratio (IMR) from the attrition probit model	0.134***	0.415***	0.015*	0.056***
	(0.0276)	(0.0288)	(0.0087)	(0.0092)
Constant	-0.082	(0.1283)	-0.181	(0.1272)
lnalpha/ sigma_u	-2.314***	(0.1396)	0.225***	(0.0238)
sigma_e			0.364***	(0.0124)
Panel households	1675	1675	1675	1675
Observations	3350	5025	3350	5025

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively the number of crops grown and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parentheses are robust standard errors; Household and farm characteristics included as controls include farm size (ha), household labor units, asset wealth index (PCA), the household age dependency ratio (%), distance to agricultural markets (km), age (years) of household head, dummies for female and single-headed households, and a dummy for household head attaining at least standard 7 (D7). The reference for regions is the Zanzibar zone. The year 2008 and 2011 are reference dummies in CRE and DRE models, respectively. The full table of coefficients showing all control variables included in the DRE and CRE Poisson and Tobit models is shown in Supplementary material (Table E).*

4.2.2. Impact of lagged crop diversification on later diversification

Besides the influence of rainfall variability and shocks, with the DRE specifications, we can also learn the influence of lagged crop diversification decisions on subsequent diversification. In both Malawi and Tanzania, we learn that lagged crop diversification decisions significantly enhance crop diversification. Precisely, in Malawi, a marginal increase in the number of crops grown at baseline(D_0) and in the previous survey round (D_{t-1}) enhances the number of crops grown in subsequent years by 0.04 and 0.02 respectively (Table 3). Similarly, we find a marginal increase in the Simpson index at baseline (D_0) and in the previous survey round (D_{t-1}) to enhance the Simpson index by 0.09 and 0.07 units in Malawi respectively (Table 3). In Tanzania, crop diversification indices at baseline (D_0) also enhance current crop diversification indices (Table 4). For instance, the count index at baseline and in the previous round marginally enhance the number of crops grown by respectively 0.12 and 0.02 respectively (Table 4). Similarly, a marginal increase in the Simpson index at baseline (D_0) enhances the Simpson index by 0.19 in Tanzania (Table 4). Comparing the influence of crop diversification indices at baseline(D_0) and in previous survey rounds (D_{t-1}) we establish that baseline indices enhance crop diversification to a greater extent compared to the previous survey round indices in both Malawi and Tanzania.

Overall, we establish that lagged crop diversification decisions enhance later crop diversification in studied countries. Based on these findings, we could not reject our hypothesis that crop diversification decisions are state-dependent. We expand on this important finding in the next sections (discussions).

4.3. Heterogeneity effects

In addition to the main effects of covariate rainfall shocks, and lagged crop diversification decisions on subsequent crop diversification we also explore heterogeneity in the effect of covariate rainfall shocks, and lagged diversification in different household socioeconomic strata as defined by their relative land and non-land asset endowments. As described earlier (methods section) we consider three quintiles of farm and non-land household asset endowments (low, Medium, and high) and compare the effects of our prime variables of interest for households in different socioeconomic groups. We present results on household asset-wealth and farm size heterogeneities for the studied countries in Table 5 and Table 6, respectively.

4.3.1. Effects of covariate rainfall shocks on diversification in households of different socioeconomic strata

From the results, we learn that the 1-year lag of drought shocks significantly enhances crop diversification for households in high quintiles of household asset-wealth (i.e., quintile 3) compared to relatively less endowed households (i.e., quintiles 1 and 2) in both Malawi and Tanzania (Table 5). We also learn that long-term rainfall variability unanimously enhances crop diversification in all households' socioeconomic groups (low, medium, and high asset wealth endowments) (Table 5). When we consider heterogeneities by total farm size holdings (Table 6) we find similar results: (i) households better endowed with land (highly endowed Quintile 3), respond to the 1-year lag of drought shock by significantly enhancing crop diversification compared to relatively less endowed counterparts (those in medium to low farm size endowment quintiles). (ii) also, long-term rainfall variability significantly enhances crop diversification decisions in all household socioeconomic groups (low, medium, and highly endowed households) (Table 6).

We summarize the relationships between the Simpson index of crop diversification and covariate rainfall shocks (1-year lag of rainfall shock, and long-term rainfall variability) by quintiles of farm and non-farm endowment quintiles in Figure 5. From the plots, we see similar relationships to those we find in parametric regressions on the relationships between crop diversification and rainfall shocks in different household socioeconomic groups.

Table 5: Crop diversification decisions in Malawi and Tanzania: Asset wealth endowment heterogeneities (Low-Medium, High)

	Malawi				**Tanzania**			
	Poisson models		**Tobit Models**		**Poisson models**		**Tobit Models**	
Variables	**DRE**	**CRE**	**DRE**	**CRE**	**DRE**	**CRE**	**DRE**	**CRE**
Low asset wealth endowments (Quintile=1)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.062	-0.014	0.030	0.025	0.032	-0.018	-0.009	-0.002
(z-score)	(0.1015)	(0.0992)	(0.0431)	(0.0415)	(0.0841)	(0.0718)	(0.0260)	(0.0215)
Drought shock growing season 1-year lag	-0.043	-0.024	-0.014	0.007	0.014	0.020	0.012	0.009
(z-score)	(0.0313)	(0.0280)	(0.0137)	(0.0115)	(0.0327)	(0.0263)	(0.0100)	(0.0079)
Long-term season (Nov-April) rainfall	0.010***	0.008***	0.003***	0.002***	0.003**	0.003***	0.001***	0.001***
variability (1980-2018) (mm)	(0.0014)	(0.0007)	(0.0005)	(0.0003)	(0.0012)	(0.0010)	(0.0004)	(0.0003)
Lagged Crop diversification indices ($D_{i,t-n}$)								
Count (Simpson) index at baseline (D_0)	0.021**		0.047		0.065***		0.128***	
	(0.0098)		(0.0333)		(0.0181)		(0.0375)	
Count (Simpson) index in the previous	0.035***		0.108***		0.080***		0.079*	
panel round (D_{t-1})	(0.0090)		(0.0416)		(0.0173)		(0.0431)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1055	1313	1055	1313	1128	1675	1128	1675
Medium asset wealth endowments (Quintile=2)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.046	0.015	0.067*	0.044	0.073	0.123*	0.017	0.031
(z-score)	(0.0898)	(0.0886)	(0.0368)	(0.0367)	(0.0812)	(0.0642)	(0.0271)	(0.0205)
Drought shock growing season 1-year lag	-0.018	0.023	0.001	0.007	0.050	0.033	0.012	0.008
(z-score)	(0.0343)	(0.0273)	(0.0143)	(0.0108)	(0.0311)	(0.0227)	(0.0101)	(0.0072)
Long-term season (Nov-April) rainfall	0.004***	0.006***	0.002***	0.002***	0.003**	0.004***	0.001**	0.001***
variability (1980-2018) (mm)	(0.0015)	(0.0007)	(0.0006)	(0.0003)	(0.0012)	(0.0009)	(0.0004)	(0.0003)
Lagged Crop diversification indices ($D_{i,t-n}$)								
Count (Simpson) index at baseline (D_0)	0.027***		0.113***		0.058***		0.130***	
	(0.0097)		(0.0347)		(0.0172)		(0.0360)	
Count (Simpson) index in previous panel	0.044***		0.055		0.087***		0.173***	
round (D_{t-1})	(0.0089)		(0.0394)		(0.0159)		(0.0414)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	884	1277	884	1277	1135	1675	1135	1675
High asset wealth endowments (Quintile=3)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.032	-0.002	0.013	0.008	0.036	0.051	0.021	0.019
(z-score)	(0.0739)	(0.0715)	(0.0305)	(0.0295)	(0.0739)	(0.0588)	(0.0237)	(0.0184)
Drought shock growing season 1-year lag	0.070**	0.071***	0.015	0.027***	0.127***	0.049*	0.037***	0.020**
(z-score)	(0.0306)	(0.0249)	(0.0129)	(0.0099)	(0.0346)	(0.0250)	(0.0114)	(0.0078)
Long-term season (Nov-April) rainfall	0.007***	0.007***	0.001	0.001***	0.002	0.003***	0.001	0.001***
variability (1980-2018) (mm)	(0.0016)	(0.0009)	(0.0006)	(0.0003)	(0.0016)	(0.0012)	(0.0005)	(0.0003)
Lagged Crop diversification indices ($D_{i,t-n}$)								

Count (Simpson) index at baseline (D_0)	0.036***		0.093***		0.148***		0.138***	
	(0.0093)		(0.0336)		(0.0172)		(0.0458)	
Count (Simpson) index in previous panel round (D_{t-1})	0.029***		0.039		0.003		0.164***	
	(0.0073)		(0.0392)		(0.0151)		(0.0533)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	974	1294	974	1294	1087	1675	1087	1675

*Notes: We run Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models and report Average partial effects (APE), The dependent variables for DRE and CRE Poisson and Tobit models are respectively the number of crops grown and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parentheses are robust standard errors.*

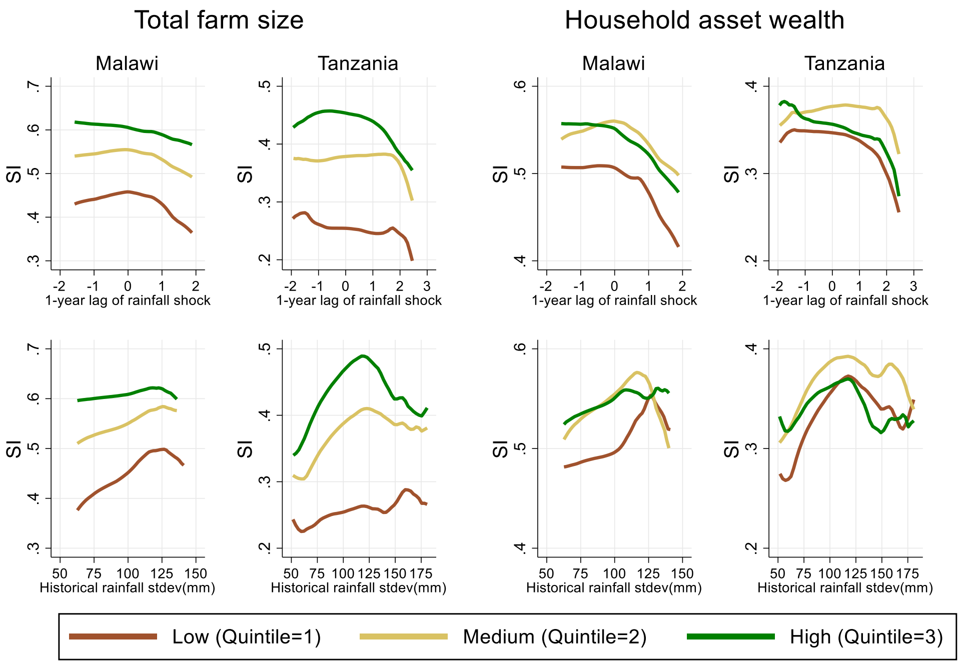

Figure 5: A summary of the relationships between covariate shock exposure (1-year lag rainfall shocks, and long-term rainfall variability) and the Simpson index (SI) of crop diversification by quintiles of farm size and household asset wealth endowments. The figure plots non-parametric local polynomial regressions using the Simpson index of crop diversification as the dependent variable. The top panel of the graph shows the relationships between the SI and the 1-year lag of rainfall shocks, while the bottom panel shows the relationship between the SI and long-term rainfall variability. The left and right panels in the graph show heterogeneities in the relationships by farm size and household asset wealth endowment quintiles, respectively.

Overall results partly support our hypothesis that better resource endowed smallholder farmers achieve successful diversification and are more likely to intensify crop diversification post-drought shock exposure to buffer future risk, unlike their poorer counterparts.

4.3.2. Effects of lagged diversification on subsequent diversification in households of different socioeconomic strata

We also learn that state-dependency in crop diversification decisions is a phenomenon found in all household socioeconomic groups as defined by the three quintiles of household asset wealth endowments (Table 5) and farm size endowments (Table 6) in both countries. However, state dependency effects appear to be slightly stronger in most cases in groups of households with the highest land and household non-land asset endowments (Quintile 3) compared to medium and lowly endowed households (Quintiles 1 and 2) (Table 5 and Table 6).

Table 6: Crop diversification decisions in Malawi and Tanzania: Farm size endowment heterogeneities (Low-Medium, High)

	Malawi				**Tanzania**			
	Poisson models		**Tobit Models**		**Poisson models**		**Tobit Models**	
Variables	**DRE**	**CRE**	**DRE**	**CRE**	**DRE**	**CRE**	**DRE**	**CRE**
Low Farm size endowments (Quintile=1)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.090	0.064	0.004	0.005	0.013	0.031	0.010	0.020
(z-score)	(0.1183)	(0.1162)	(0.0442)	(0.0424)	(0.0802)	(0.0747)	(0.0212)	(0.0194)
Drought shock growing season 1-year lag	-0.049	-0.021	-0.010	0.006	0.055^{*}	0.041	0.011	0.010
(z-score)	(0.0373)	(0.0309)	(0.0142)	(0.0110)	(0.0319)	(0.0276)	(0.0085)	(0.0073)
Long-term season (Nov-April) rainfall	0.007^{***}	0.006^{***}	0.003^{***}	0.002^{***}	0.003^{**}	0.003^{***}	0.001^{***}	0.001^{**}
variability (1980-2018) (mm)	(0.0016)	(0.0007)	(0.0006)	(0.0003)	(0.0013)	(0.0010)	(0.0004)	(0.0003)
Lagged Crop diversification indices ($D_{i,t-n}$)								
Count (Simpson) index at baseline (D_0)	0.005		0.065^{*}		0.112^{***}		0.098^{***}	
	(0.0120)		(0.0352)		(0.0189)		(0.0321)	
Count (Simpson) index in previous panel	0.061^{***}		0.059		0.047^{***}		0.103^{***}	
round (D_{t-1})	(0.0102)		(0.0442)		(0.0169)		(0.0376)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	968	1295	968	1295	1337	1746	1337	1746
Medium Farm size endowments (Quintile=2)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.096	0.053	0.058	0.049	0.150^{**}	0.159^{**}	0.045	0.049^{**}
(z-score)	(0.0924)	(0.0939)	(0.0382)	(0.0379)	(0.0744)	(0.0619)	(0.0275)	(0.0205)
Drought shock growing season 1-year lag	0.003	0.004	0.008	0.009	0.012	0.008	-0.003	-0.001
(z-score)	(0.0325)	(0.0271)	(0.0140)	(0.0108)	(0.0326)	(0.0251)	(0.0114)	(0.0082)
Long-term season (Nov-April) rainfall	0.007^{***}	0.008^{***}	0.002^{***}	0.002^{***}	0.002	0.003^{***}	0.001^{**}	0.001^{***}
variability (1980-2018) (mm)	(0.0014)	(0.0007)	(0.0006)	(0.0003)	(0.0012)	(0.0009)	(0.0005)	(0.0003)
Lagged Crop diversification indices ($D_{i,t-n}$)								
Count (Simpson) index at baseline (D_0)	0.030^{***}		0.045		0.046^{***}		0.142^{***}	
	(0.0089)		(0.0337)		(0.0175)		(0.0468)	
Count (Simpson) index in previous panel	0.051^{***}		0.108^{***}		0.084^{***}		0.103^{*}	
round (D_{t-1})	(0.0078)		(0.0390)		(0.0175)		(0.0558)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	956	1297	956	1297	953	1634	953	1634
High Farm size endowments (Quintile=3)								
Rainfall shocks and variability								
Flood shock growing season 1-year lag	0.021	-0.038	0.041	0.018	0.092	0.065	0.015	0.013
(z-score)	(0.0679)	(0.0648)	(0.0266)	(0.0258)	(0.0716)	(0.0570)	(0.0282)	(0.0207)
Drought shock growing season 1-year lag	0.027	0.057^{**}	0.014	0.026^{***}	0.033	0.021	0.027^{**}	0.016^{**}
(z-score)	(0.0270)	(0.0227)	(0.0109)	(0.0088)	(0.0319)	(0.0218)	(0.0127)	(0.0080)
Long-term season (Nov-April) rainfall	0.008^{***}	0.007^{***}	0.001^{*}	0.001^{***}	0.002^{*}	0.005^{***}	0.001^{*}	0.002^{***}
variability (1980-2018) (mm)	(0.0016)	(0.0009)	(0.0005)	(0.0003)	(0.0012)	(0.0010)	(0.0004)	(0.0004)
Lagged Crop diversification indices ($D_{i,t-n}$)								
Count (Simpson) index at baseline (D_0)	0.038^{***}		0.113^{***}		0.083^{***}		0.137^{***}	
	(0.0086)		(0.0274)		(0.0147)		(0.0351)	
Count (Simpson) index in previous panel	0.014^{**}		0.044		0.031^{**}		0.158^{***}	
round (D_{t-1})	(0.0069)		(0.0294)		(0.0145)		(0.0352)	
Household & farm characteristics								
Observed household characteristics	Yes	No	Yes	No	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes	No	Yes	No	Yes
Year dummies & regional dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	989	1292	989	1292	1060	1645	1060	1645

*Notes: We run Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models and report Average partial effects (APE), The dependent variables for DRE and CRE Poisson and Tobit models are respectively the number of crops grown and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parentheses are robust standard errors.*

We also summarize the relationships between lagged crop diversification decisions and subsequent crop diversification indices using non-parametric local polynomial regressions in (Figure 6) and show that state dependency in crop diversification is common in all household socioeconomic strata and it is slightly stronger in relatively better-endowed households in terms of land and non-land assets (Figure 6).

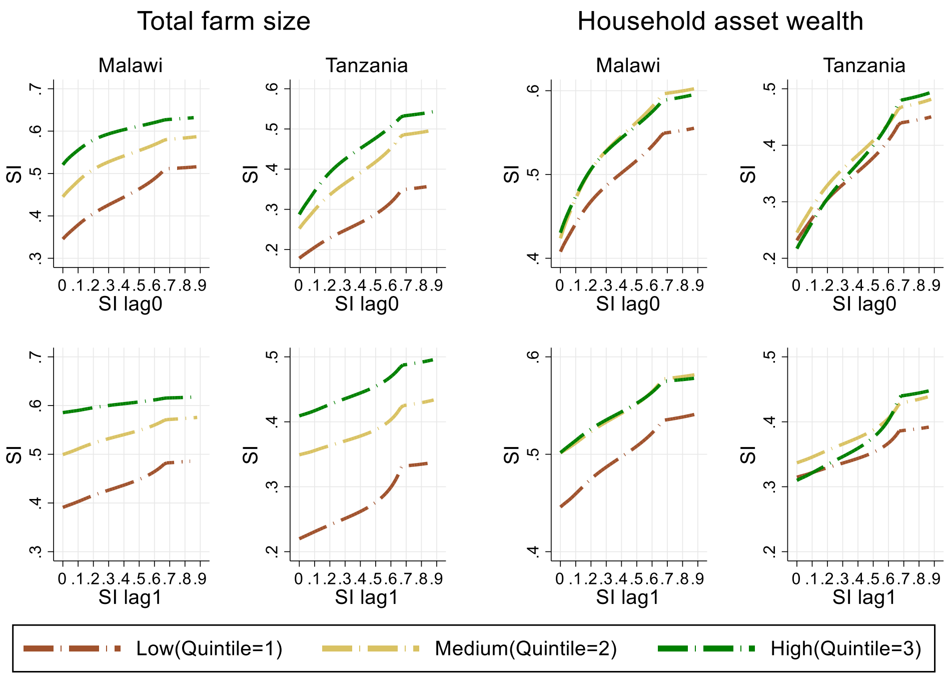

Figure 6: A summary of the relationships between lagged crop diversification decisions (Simpson index at baseline (SI lag 0), Simpson index in previous panel round (SI lag 1), and subsequent Simpson index of crop diversification by quintiles of farm size, and household asset wealth endowments. The figure plots non-parametric local polynomial regressions using the Simpson index (SI) of crop diversification as the dependent variable. The top panel of the graph shows the relationships between the SI and the SI index at baseline (SI lag0), while the bottom panel shows the relationship between the SI and SI in the previous panel round (SI lag1). The left and right panels in the graph show heterogeneities in the relationships by farm size and household asset wealth endowment quintiles respectively.

Results from heterogeneity analysis support our hypothesis that better resource endowed smallholder farmers are more likely to capitalize on the experience gained from past diversification and enhance subsequent crop diversification to meet multiple objectives including buffering their crop portfolios to future climate risks, unlike their poorer counterparts.

5. Discussion

Our paper focuses on smallholder farming in Malawi and Tanzania and evaluated: (i) whether covariate rainfall shocks (drought and flood) and long-term rainfall variability significantly influence crop diversification in smallholder farming? (ii) whether lagged crop diversification decisions influence later diversification decisions (are state-dependent), and (iii) whether the impact of shocks and lagged crop diversification on subsequent diversification differ for farmers in different strata of land and non-land asset endowments. Our paper highlights a few key findings for discussion, which we discuss below:

5.1. Impact of lagged rainfall shocks and long-term rainfall variability

First, we establish that lagged rainfall shocks, particularly drought shocks and long-term rainfall uncertainty (variability) enhance crop diversification in both Malawi and Tanzania. This result implies that smallholder farmers in Malawi and Tanzania respond to rainfall shocks by intensifying on-farm crop diversification. Our results align with the state-contingent theory of technology adoption under risk, which reflects that farming households learn from shocks and may be more willing to adopt risk substituting strategies as an adaptative mechanism (Quiggin and Chambers 2006; Holden and Quiggin 2017). An exogenous increase in the probability of having less favorable conditions as drought shocks and highly variable rainfall thus trigger the adoption of a more diversified cropping portfolio. These results are in line with literature that alludes to the fact that rural households switch from their business as usual practices to practices that increase their mutual insurance to shocks to better cope with shocks (Takasaki 2011; Angelsen and Dokken 2018). Intensifying crop production is hence a strategy that helps farmers to buffer climate risks. Our findings also corroborate previous studies that found exposure to climate shocks to increase crop diversification(Huang et al. 2014; Mulwa and Visser 2020; Matsuura 2021; Swinger 2022), and the general literature that attribute climate risk as one of the key determinants of crop diversification (see, for example, Alobo Loison (2015) ,Tacconi et al. (2022), and Acevedo et al. (2020). Crop diversification is important in smallholder farming as it helps them increase their chances of dealing with the uncertainty created by rainfall shocks and provides alternative means of generating food and income with increasing uncertainty. The importance of crop diversification as an adaptation mechanism to rainfall variability and drought shocks over time poses interesting implications for seed security under stress. Smallholder farmers will need access to diverse and well-adapted crop seed and planting materials overtime to help them adapt to recurrent rainfall shocks.

5.2. Impact of lagged crop diversification decisions

Second, we also establish that crop diversification decisions are state-dependent. State-dependency in crop diversification decisions implies past crop diversification enhances later diversification. Household initial conditions associated with crop diversification, including transaction costs, hence facilitate crop diversification in subsequent years. Seed and other input markets in SSA are characterized by imperfections making markets access by households not to be uniform, as households may face different transaction costs (Renkow et al. 2004; Barrett 2008; Kassie et al. 2013). In crop diversification, such transaction costs may include all costs incurred in acquiring crop seed and complementary inputs required to implement a diversified cropping portfolio. For instance, in acquiring crop seed through formal or informal sources, such transaction costs may include the costs of searching and getting information on production and consumption traits of the seed of different crops, costs of searching and locating them, and negotiation costs (Badstue 2004; Salazar and Winters 2012). In such circumstances, where access to inputs is complicated by high and non-linear transaction costs, knowledge and experience in crop markets matter as it helps in marginally reducing transaction costs in access to seed from available channels over time. Farmers with knowledge and experience in implementing diversified cropping portfolios gained from past experiences and engagements with their social networks hence have an elevated advantage in successfully implementing diversified cropping portfolios in subsequent years. State-dependency in crop diversification decisions may also come from the benefits of on-farm crop diversification towards the conservation of plant genetic resources. Crop diversification on the farm support *in situ* agrobiodiversity conservation (Bellon 1996; Love and Spaner 2007; Bezabih 2008) which is an essential long term source of seed and planting material by farmers. The bulk of smallholder farmers in developing regions, including Malawi and Tanzania, get most of their crop seed from informal seed sources, including own-farm seed saving practices (Bellon et al. 2006; Coomes et al. 2015). This notion is further supported by literature that supports the fact that: (a) resource-poor farmers may grow certain crops or crop varieties to prevent their loss in the future, and (b) may maintain high crop diversity to cope and adapt to marginal environments. Hence, household crop diversification decisions in the past help in reducing transaction costs incurred in sourcing seeds in the future, which enhances later crop diversification.

5.3. Heterogeneity effects

Third, from heterogeneity analysis we gather revealing evidence to suggest that households better endowed with capital endowments (particularly land and household assets) are: (i) highly diversified, (ii) able to capitalize on the experience gained from past diversification to intensify crop diversification in subsequent years, and (iii) more likely to intensify crop diversification following drought shock exposure when compared to their opposite poorer counterparts. The findings imply that land and non-land asset (household wealth) endowments help farmers implement diversified cropping portfolios to help them deal with rainfall shocks. Household resource endowments, markets, and infrastructure that facilitate access and use of these capital endowments help farmers implement agricultural risk management strategies such as crop diversification (Dercon 2005; Winters et al. 2009; Alobo Loison 2015). In addition, household asset endowments act as informal insurance for rural households against shocks; hence, farmers better endowed with assets are more resilient (Dercon 2004; IPCC 2014; Angelsen and Dokken 2018). As a result, they can intensify crop diversification to adapt to future shock exposure, unlike their poorer counterparts. Given the importance of crop diversification in supporting both seed and food security objectives over time (Bellon 1996; Love and Spaner 2007; Di Falco et al. 2010; Asfaw et al. 2019; Bozzola and Smale 2020), our results imply that poorer farmers are more likely to become seed and food insecure with recurrent rainfall shock exposure compared to their opposite counterparts.

6. Conclusions and policy implications

Covariate rainfall shocks affect agricultural production, threatening the livelihoods of many people dependent on agriculture for survival in SSA. Crop diversification is an important strategy that smallholder farmers can embrace to improve resilience against such covariate risk. This paper focuses on the evolution of smallholder farmers' crop diversification decisions, and their responses to short-term rainfall shocks, and long-term rainfall variability using balanced household panel data from Malawi and Tanzania. We specifically used four (three) rounds of balanced household panel data built from the Malawi (Tanzania) Living Standards Measurement Survey-Integrated Surveys on Agriculture (LSMS-ISA) combined with historical monthly weather data to fulfill our research objectives. We study crop diversification at the household level as a state-contingent risk decision, given that households make their costly investment decision before the state of nature is revealed. We test two main hypotheses: (a) Recent (past) exposure to drought or flood shocks and long-term rainfall variability increase crop diversification, (b) Crop diversification decisions are state-dependent (lagged crop diversification decisions strongly and positively explain later crop diversification decisions). Additionally, we test two sub-hypotheses that households better endowed with assets (land, and household assets) are more likely to: (i) intensify crop diversification to help them deal with rainfall shocks (i.e., drought shocks), and (ii) capitalize on the experience gained from past diversification decisions and use it to diversify production overtime to deal with recurrent climate risk, unlike their poorer counterparts. We analyze balanced household panel data using correlated random effects and dynamic random effects panel Poisson and Tobit models that control for unobserved heterogeneity in household crop diversification decisions plus initial conditions that may influence crop diversification across space and time. Our findings support a few policy-relevant conclusions:

i. First, smallholder farmers in Malawi and Tanzania respond to recent drought shocks and long-term rainfall variability by intensifying on-farm crop diversification. Smallholder farmers hence need access to diverse seed and planting materials over time to help them adapt to recurrent rainfall shocks.
ii. Second, crop diversification decisions are state-dependent, implying past crop diversification enhances later diversification. Knowledge and experience gained from implementing crop diversification in the past, and contributions of past crop diversification to on-farm agrobiodiversity (e.g., on-farm seed saving) reduce transaction costs in acquiring seed over time, which supports diversification in subsequent years.
iii. Land and non-land assets (household wealth) endowments help farmers implement diversified cropping portfolios over time to help them deal with rainfall shocks. Precisely, it is the relatively better-off farmers with sufficient land and non-land assets who are more likely to achieve crop diversification post-drought shocks exposure as an adaptation to future expected shocks.

Our findings support policies that ensure and promote access to a diversity of affordable, well-adapted crop seeds that meet farmers' needs and preferences in Malawi and Tanzania to improve resilience under rainfall uncertainty and shocks. Farmers' informal seed systems, including such channels and social networks and local markets, continues to supply the bulk of the seeds and planting material to farmers in SSA and must therefore not be overlooked in policies and programmes aimed at improving access to seed. From the formal seed sector, policies that target improving the supply of affordable, diverse good quality seed, that meet farmers' needs and preferences are needed. On the demand side, working on capacitating smallholder farmers to enhance their skills to achieve successful diversification and conserve locally adapted crop diversity on-farm (*in situ*) will help complement supply-side efforts. For instance, well-designed pro-poor extension services that can (i) capacitate farmers in improving on-farm seed saving as a strategy to ensure future access to seed, and (ii) give them information on how to implement different crop combinations to buffer against different climate risks (e.g., droughts or floods) at the lowest possible costs. Policies that successfully promote such diversity in approaches to seed security will reduce transaction costs and improve overall access to seed. Such will play an important role in ensuring that farmers, including the disadvantaged, can successfully implement a diversified cropping portfolio to adapt to recurrent rainfall variability and drought shocks.

Acknowledgments

This research work is an output of a research project, "*Access to seeds: from Emergencies to Seed System development (ACCESS),"* which is fully supported by the Research Council of Norway, project number (RCN 288493). The authors are also grateful to the World Bank for making the Livelihood Standard Measurement survey data for Tanzania and Malawi available for this research.

References

Acevedo M et al. (2020) A scoping review of adoption of climate-resilient crops by small-scale producers in low-and middle-income countries Nature plants 6:1231-1241

Alobo Loison S (2015) Rural Livelihood Diversification in Sub-Saharan Africa: A Literature Review The Journal of Development Studies 51:1125-1138 doi:10.1080/00220388.2015.1046445

Angelsen A, Dokken T (2018) Climate exposure, vulnerability and environmental reliance: a cross-section analysis of structural and stochastic poverty Environment and Development Economics 23:257-278 doi:10.1017/S1355770X18000013

Asfaw S, Scognamillo A, Caprera GD, Sitko N, Ignaciuk A (2019) Heterogeneous impact of livelihood diversification on household welfare: Cross-country evidence from Sub-Saharan Africa World Development 117:278-295 doi:https://doi.org/10.1016/j.worlddev.2019.01.017

Badstue LB (2004) Identifying the factors that influence small-scale farmers' transaction costs in relation to seed acquisition.

Barrett CB (2008) Smallholder market participation: Concepts and evidence from eastern and southern Africa Food policy 33:299-317

Behrman JR (2000) Basic economic models and econometric tools. In: Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study.

Bellon MR (1996) The dynamics of crop infraspecific diversity: A conceptual framework at the farmer level Economic botany 50:26-39

Bellon MR, Hellin J (2011) Planting hybrids, keeping landraces: Agricultural modernization and tradition among small-scale maize farmers in Chiapas, Mexico World Development 39:1434-1443

Bellon MR, Adato M, Becerril J, Mindek D (2006) Poor farmers' perceived benefits from different types of maize germplasm: the case of creolization in lowland tropical Mexico World development 34:113-129

Bellon MR, Kotu BH, Azzarri C, Caracciolo F (2020) To diversify or not to diversify, that is the question. Pursuing agricultural development for smallholder farmers in marginal areas of Ghana World Development 125:104682 doi:https://doi.org/10.1016/j.worlddev.2019.104682

Bezabih M (2008) Agrobiodiversity conservation under an imperfect seed system: the role of community seed banking schemes Agricultural Economics 38:77-87

Binswanger HP, Rosenzweig MR (1986) Behavioural and material determinants of production relations in agriculture The Journal of Development Studies 22:503-539

Bora K (2022) Rainfall shocks and fertilizer use: a district level study of India Environment and Development Economics:1-22

Bozzola M, Smale M (2020) The welfare effects of crop biodiversity as an adaptation to climate shocks in Kenya World Development 135:105065
Carney D (1998) Implementing the sustainable rural livelihoods approach Sustainable rural livelihoods: What contribution can we make 3:27
Caviglia-Harris JL (2004) Household production and forest clearing: the role of farming in the development of the Amazon Environment and Development Economics 9:181-202
Chakraborty S, Newton AC (2011) Climate change, plant diseases and food security: an overview Plant pathology 60:2-14
Chamberlain G (1982) Multivariate regression models for panel data Journal of econometrics 18:5-46
Chambers RG, Quiggin J (2000) Uncertainty, production, choice, and agency: the state-contingent approach. Cambridge University Press,
Coase RH (1937) The nature of the firm economica 4:386-405
Coase RH (1960) The problem of social cost. In: Classic papers in natural resource economics. Springer, pp 87-137
Coomes OT et al. (2015) Farmer seed networks make a limited contribution to agriculture? Four common misconceptions Food Policy 56:41-50
De Janvry A, Fafchamps M, Sadoulet E (1991) Peasant household behaviour with missing markets: some paradoxes explained The Economic Journal 101:1400-1417
Dercon S (2002) Income risk, coping strategies, and safety nets The World Bank Research Observer 17:141-166
Dercon S (2004) Insurance against poverty. Oxford University Press. OUP Oxford, Oxford
Dercon S (2005) Risk, poverty and vulnerability in Africa Journal of African Economies 14:483-488
Dercon S, Christiaensen L (2011) Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia Journal of development economics 96:159-173
Di Falco S, Chavas JP (2009) On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia American Journal of Agricultural Economics 91:599-611
Di Falco S, Bezabih M, Yesuf M (2010) Seeds for livelihood: crop biodiversity and food production in Ethiopia Ecological Economics 69:1695-1702
Ellis F (2000) Rural livelihoods and diversity in developing countries. Oxford university press,
Fafchamps M (2004) Market institutions in sub-Saharan Africa. Cambridge, MA: MIT Press,
FAO (2018) Reviewing and enhancing the existing seed security conceptual framework. http://www.fao.org/in-action/food-security-capacity-building/project-components/seeds/seed-security-conceptual-framework/en/.
FAO, ECHA (2015) Building Capacity for Seed Security Assessments: Household Seed Security Concepts and Indicators, Discussion Paper. FAO, and European Commission Humanitarian Aid. .
Feliciano D (2019) A review on the contribution of crop diversification to Sustainable Development Goal 1 "No poverty" in different world regions Sustainable Development 27:795-808
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas International journal of climatology 37:4302-4315
Filmer D, Pritchett LH (2001a) Estimating wealth effects without expenditure data–or tears: an application to educational enrollments in states of India Demography 38
Filmer D, Pritchett LH (2001b) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India Demography 38:115-132
Filmer D, Pritchett LH (2001c) Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. Population Association of America,
Gebru M, Holden ST, Tilahun M (2019) Tenants' land access in the rental market: evidence from northern Ethiopia Agricultural Economics 50:291-302
Gebru M, Holden ST, Alfnes F (2021) Adoption analysis of agricultural technologies in the semiarid northern Ethiopia: a panel data analysis Agricultural and Food Economics 9:1-16
Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset International journal of climatology 34:623-642
Holden ST, Quiggin J (2017) Climate risk and state-contingent technology adoption: shocks, drought tolerance and preferences European Review of Agricultural Economics 44:285-308
Holden ST, Otsuka K, Place FM (2010) The emergence of land markets in Africa: Impacts on poverty, equity, and efficiency. Routledge,
Huang J-k, Jiang J, Wang J-x, Hou L-l (2014) Crop diversification in coping with extreme weather events in China Journal of Integrative Agriculture 13:677-686
IPCC (2014) Climate Change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of working group 2 to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, New York

IPCC (2022) Climate Change 2022: Impacts, adaptation, and vulnerability. . Cambridge University Press, New York
Jarvis DI et al. (2008) A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities Proceedings of the National Academy of Sciences 105:5326-5331
Kassie G, MacRobert J, Setimela P, Erenstein O, Mwangi W, Shiferaw B (2013) Political and economic features of the maize seed industry in southern Africa Agrekon 52:104-127
Katengeza SP, Holden ST, Lunduka RW (2019a) Adoption of drought tolerant maize varieties under rainfall stress in Malawi Journal of Agricultural Economics 70:198-214 doi: https://doi.org/10.1111/1477-9552.12283
Katengeza SP, Holden ST, Fisher M (2019b) Use of Integrated Soil Fertility Management Technologies in Malawi: Impact of Dry Spells Exposure Ecological Economics 156:134-152 doi:https://doi.org/10.1016/j.ecolecon.2018.09.018
Kubik Z, Maurel M (2016) Weather shocks, agricultural production and migration: Evidence from Tanzania The Journal of Development Studies 52:665-680
Labeyrie V et al. (2021) The role of crop diversity in climate change adaptation: insights from local observations to inform decision making in agriculture Current Opinion in Environmental Sustainability 51:15-23
Lin BB (2011) Resilience in Agriculture through Crop Diversification: Adaptive Management for Environmental Change BioScience 61:183-193 doi:10.1525/bio.2011.61.3.4
Love B, Spaner D (2007) Agrobiodiversity: Its value, measurement, and conservation in the context of sustainable agriculture Journal of Sustainable Agriculture 31:53-82
Makate C, Wang R, Makate M, Mango N (2016) Crop diversification and livelihoods of smallholder farmers in Zimbabwe: adaptive management for environmental change SpringerPlus 5:1135 doi:10.1186/s40064-016-2802-4
Markelova H, Meinzen-Dick R, Hellin J, Dohrn S (2009) Collective action for smallholder market access Food policy 34:1-7
Masarie KA, Tans PP (1995) Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record Journal of Geophysical Research: Atmospheres 100:11593-11610
Matsuura M (2021) Climate Change and Crop Diversification: Evidence from Bangladesh Available at SSRN 3884599
McGuire S, Sperling L (2013) Making seed systems more resilient to stress Global Environmental Change 23:644–653 doi:http://dx.doi.org/10.1016/j.gloenvcha.2013.02.001
McKenzie DJ (2005) Measuring inequality with asset indicators Journal of population economics 18:229-260
Michler JD, Baylis K, Arends-Kuenning M, Mazvimavi K (2019) Conservation agriculture and climate resilience Journal of environmental economics and management 93:148-169
Morton JF (2007) The Impact of Climate Change on Smallholder and Subsistence Agriculture Proceedings of the National Academy of Sciences of the United States of America 104:19680-19685
Mulwa CK, Visser M (2020) Farm diversification as an adaptation strategy to climatic shocks and implications for food security in northern Namibia World Development 129:104906
Mundlak Y (1978) On the Pooling of Time Series and Cross Section Data Econometrica 46:69-85 doi:10.2307/1913646
Nordhagen S, Pascual U (2013) The impact of climate shocks on seed purchase decisions in Malawi: implications for climate change adaptation World Development 43:238-251
North (1987) Transaction Costs. In: John Eatwell, Murray Milgate, Newman. P (eds) The New Palgrave: A Dictionary of Economics, vol 4. Macmillan, London,
Quiggin J, Chambers RG (2006) The state-contingent approach to production under uncertainty Australian Journal of Agricultural and Resource Economics 50:153-169
Renard D, Tilman D (2019) National food production stabilized by crop diversity Nature 571:257-260
Renkow M, Hallstrom DG, Karanja DD (2004) Rural infrastructure, transactions costs and market participation in Kenya Journal of development economics 73:349-367
Ricker-Gilbert J, Chamberlin J (2018) Transaction costs, land rental markets, and their impact on youth access to agriculture in Tanzania Land Economics 94:541-555
Rose E (2001) Ex ante and ex post labor supply response to risk in a low-income area Journal of development economics 64:371-388
Salazar L, Winters P (2012) The impact of seed market access and transaction costs on potato biodiversity and yields in Bolivia Environment and Development Economics 17:633-661
Scoones I (1998) Sustainable rural livelihoods: a framework for analysis
Simpson EH (1949) Measurement of diversity nature 163:688-688
Singh I, Squire L, Strauss J (1986) A survey of agricultural household models: Recent findings and policy implications The World Bank Economic Review 1:149-179

Smale M, Meng E, Brennan JP, Hu R (2003) Determinants of spatial diversity in modern wheat: examples from Australia and China Agricultural Economics 28:13-26
Swinger B (2022) Are Risk Attitudes and Rainfall Variability Determinants of Diversification? Evidence from Rondonia, Brazil
Tacconi F, Waha K, Ojeda JJ, Leith P (2022) Drivers and constraints of on-farm diversity. A review Agronomy for Sustainable Development 42:2 doi:10.1007/s13593-021-00736-6
Takasaki Y (2011) Do the commons help augment mutual insurance among the poor? World Development 39:429-438
Tione SE, Holden ST (2021a) Non-convex transaction costs and land rental market participation in Malawi Land Economics:062620-060092R
Tione SE, Holden ST (2021b) Can rainfall shocks enhance access to rented land? Evidence from Malawi Agricultural Economics
Waha K et al. (2022) The benefits and trade-offs of agricultural diversity for food security in low- and middle-income countries: A review of existing knowledge and evidence Global Food Security 33:100645 doi:https://doi.org/10.1016/j.gfs.2022.100645
Westengen OT, Ring KH, Berg PR, Brysting AK (2014) Modern maize varieties going local in the semi-arid zone in Tanzania BMC evolutionary biology 14:1
Winters P et al. (2009) Assets, activities and rural income generation: Evidence from a multicountry analysis World Development 37:1435-1452
Wood D (1997) The conservation of agrobiodiversity on-farm: questioning the emerging paradigm Biodiversity & Conservation 6:109-129
Wooldridge JM (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity Journal of applied econometrics 20:39-54
Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press,

Appendix: Evolution of farm-level crop diversification and response to rainfall shocks in smallholder farming: Evidence from Malawi and Tanzania

Clifton Makate* , Arild Angelsen, Stein Terje Holden, & Ola Tveitereid Westengen

A. Descriptive Statistics

- **Summary statistics of other control variables used in analysis**

Table A: Descriptive statistics of control variables used in the analysis

				Malawi			**Tanzania**
	2010	2013	2016	2019	2009	2011	2013
VARIABLES	mean	mean	mean	mean	mean	mean	mean
Female head(1=yes)	0.23	0.26	0.29	0.29	0.23	0.23	0.24
The household head is single(1=yes)	0.23	0.23	0.33	0.31	0.24	0.23	0.25
Age of household head(years)	43.55	46.06	48.05	50.35	48.63	50.61	47.77
Education level attained by the head [at least JCE-Malawi (1=yes), higher than D7-Tanzania]	0.34	0.34	0.27	0.31	0.78	0.79	0.79
Household age dependency ratio	114.99	112.58	101.31	94.18	101.36	96.91	92.88
Distance to markets (km)-ADMARC-Malawi, main urban market-Tanzania)	7.39	7.55	7.65	7.38	84.88	84.48	84.66
Long-term average rainfall (mm)	902.51	902.51	902.51	902.51	864.44	864.44	864.44
Long-term average maximum temperature(deg)	27.88	27.88	27.89	27.88	28.67	28.67	28.67
Observations	971	971	971	971	1675	1675	1675

Rainfall variables are shown for the main rainy season (November to April); Statistics are not weighted; source (own calculation from LSMS-ISA data for Malawi and Tanzania).

- **Main crops grown in panel**

Table B:Main crops grown by farmers in panel Malawi 2010-2019

	2010		2013		2016		2019		Pooled	
	Prop(%)	Area(ha)	Prop(%)	Area(ha)	Prop(%)	Area(ha)	Prop(%)	Area(ha)	Prop(%)	Area(ha)
Maize	0.977	0.568	0.956	0.549	0.966	0.586	0.964	0.538	0.966	0.560
Groundnut	0.355	0.110	0.412	0.149	0.264	0.106	0.398	0.137	0.357	0.126
Pigeon pea	0.224	0.119	0.306	0.168	0.287	0.143	0.367	0.166	0.296	0.149
Pumpkin[1] (Nhakwani)	0.086	0.037	0.204	0.092	0.178	0.081	0.510	0.223	0.244	0.108
Sorghum	0.115	0.066	0.133	0.068	0.173	0.091	0.152	0.060	0.143	0.071
Tobacco	0.172	0.065	0.124	0.048	0.125	0.066	0.103	0.049	0.131	0.057
Common bean	0.068	0.031	0.109	0.053	0.094	0.041	0.166	0.069	0.109	0.049
Soybean	0.072	0.024	0.105	0.039	0.101	0.036	0.146	0.050	0.106	0.037
Sweet potato	0.051	0.013	0.039	0.011	0.059	0.016	0.104	0.024	0.063	0.016
Rice	0.034	0.009	0.037	0.007	0.027	0.011	0.042	0.018	0.035	0.011
Finger millet	0.009	0.005	0.010	0.004	0.018	0.006	0.016	0.005	0.013	0.005
Pearl millet	0.016	0.012	0.019	0.016	0.023	0.017	0.029	0.021	0.021	0.017
Cowpeas	0.010	0.002	0.077	0.040	0.029	0.011	0.052	0.022	0.042	0.019
Cotton	0.013	0.006	0.051	0.025	0.024	0.011	0.021	0.011	0.027	0.013
Sunflower	0.014	0.010	0.016	0.011	0.019	0.007	0.015	0.006	0.016	0.008
Observations	971	971	971	971	971	971	971	971	3884	3884

Notes: Statistics are not weighted; source (own calculation from LSMS-ISA data for Malawi).

[1] Pumpkin leaves are given the local name Nhakwani in Malawi

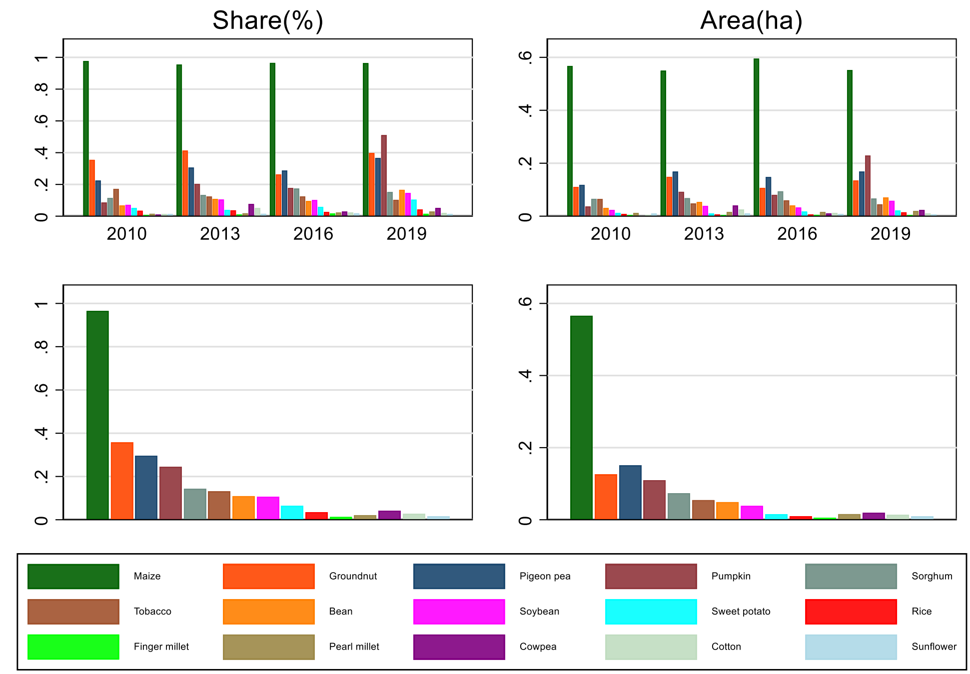

Figure A: main crops grown in Malawi panel by proportion of households growing crop (%), and average area devoted to crop(ha) in pooled samples and by panel year

Table C: Main crops grown by farmers in panel Tanzania 2008-2013

	2008		2011		2013		Pooled	
	Prop(%)	Area(ha)	Prop(%)	Area(ha)	Prop(%)	Area(ha)	Prop(%)	Area(ha)
Maize	0.635	0.876	0.623	0.900	0.651	0.920	0.636	0.899
Common bean	0.225	0.249	0.212	0.241	0.219	0.268	0.219	0.252
Rice	0.200	0.162	0.211	0.165	0.208	0.148	0.206	0.158
Groundnut	0.143	0.191	0.119	0.172	0.132	0.210	0.132	0.191
Sorghum	0.122	0.149	0.110	0.153	0.087	0.108	0.107	0.137
Sweet potato	0.088	0.088	0.072	0.083	0.076	0.107	0.079	0.093
Sunflower	0.072	0.102	0.050	0.085	0.076	0.114	0.066	0.101
Pigeon pea	0.059	0.069	0.056	0.071	0.069	0.079	0.061	0.073
Cowpeas	0.059	0.070	0.066	0.077	0.064	0.069	0.063	0.072
Cassava	0.098	0.071	0.024	0.028	0.006	0.003	0.043	0.034
Sesame	0.034	0.044	0.034	0.057	0.036	0.042	0.035	0.047
Pearl millet	0.027	0.030	0.020	0.030	0.023	0.030	0.023	0.030
Finger millet	0.019	0.033	0.014	0.032	0.007	0.008	0.013	0.025
Cotton	0.029	0.075	0.020	0.054	0.032	0.102	0.027	0.077
Bambara nut	0.038	0.047	0.021	0.035	0.021	0.043	0.026	0.042
Tobacco	0.012	0.025	0.014	0.049	0.016	0.050	0.014	0.041
Observations	1675	1675	1675	1675	1675	1675	5025	5025

Notes: Statistics are not weighted; source (own calculation from LSMS-ISA data for Tanzania).

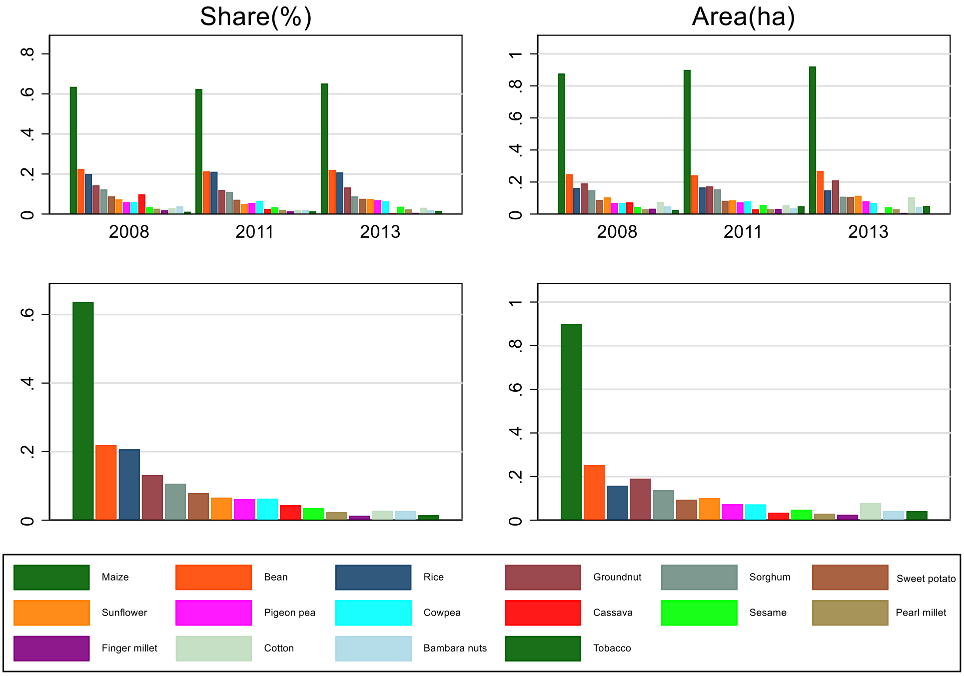

Figure B: main crops grown in Tanzania panel by proportion of households growing crop (%), and average area devoted to crop(ha) in pooled samples and by panel year.

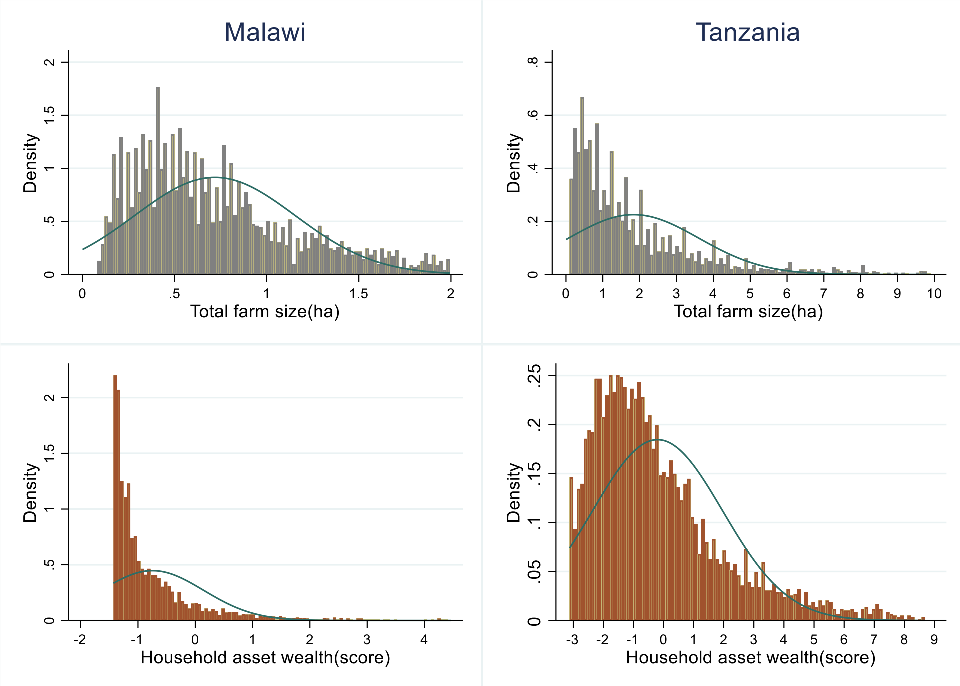

Figure C: Distribution of farm size and household asset wealth endowments in the pooled sample of households analyzed in respective countries

B. Tables showing coefficients and full set of control variables

Table D: Crop diversification decisions in Malawi: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models reporting Coefficients

	Poisson models				Tobit Models			
	DRE		CRE		DRE		CRE	
Variables	b	se	b	se	b	se	b	se
Rainfall shocks and variability								
Flood shock growing season 1-year lag (z-score)	0.049	(0.0478)	0.003	(0.0462)	0.034	(0.0237)	0.016	(0.0237)
Drought shock growing season 1-year lag (z-score)	0.004	(0.0176)	0.025*	(0.0146)	0.004	(0.0088)	0.019***	(0.0071)
Long-term early season (Nov-Jan) rainfall variability (1980-2018) (mm)	0.008***	(0.0009)	0.007***	(0.0005)	0.002***	(0.0004)	0.002***	(0.0002)
Lagged Crop diversification indices								
Count (Simpson) index at baseline (D_0)	0.038***	(0.0061)			0.105***	(0.0251)		
Count (Simpson) index in previous panel round (D_{t-1})	0.023***	(0.0050)			0.078**	(0.0358)		
Household & farm characteristics								
farmsize_gps(Farm size in ha)	0.414***	(0.0216)			0.194***	(0.0489)		
hh_laborunits (Household labor units)	0.008	(0.0060)			-0.003	(0.0028)		
household_wealth (Household wealth index)	-0.003	(0.0078)			-0.002	(0.0033)		
sex_hhh_female (female household head)	0.003	(0.0309)			-0.003	(0.0138)		
Single (Single household head)	-0.029	(0.0300)			-0.016	(0.0139)		
age_hhh (Age of household head)	-0.000	(0.0008)			0.000	(0.0004)		
edu_JCE_atleast (Education attained at least JCE)	-0.019	(0.0239)			-0.029***	(0.0109)		
hh_depend_ratio (Household age dependency ratio)	0.0001	(0.0001)			-0.0001	(0.0001)		
distance_admarc (distance to ADMARC)	-0.004	(0.0024)			-0.0001	(0.0011)		
Year dummies & regional dummies								
years4	0.243***	(0.0236)	0.378***	(0.0371)	0.076***	(0.0117)	0.132***	(0.0188)
years3	-0.029	(0.0264)	0.106***	(0.0377)	-0.107***	(0.0125)	-0.050***	(0.0180)
years2			0.136***	(0.0351)			0.052***	(0.0176)
reg_northern	0.037	(0.0624)	0.051	(0.0495)	-0.019	(0.0260)	0.004	(0.0208)
reg_southern	-0.002	(0.0406)	0.117***	(0.0318)	-0.014	(0.0171)	0.002	(0.0137)
Mean of observed characteristics								
mnhh_laborunits			0.004	(0.0101)			-0.004	(0.0045)
mnfarmsize_gps			0.517***	(0.0307)			0.190***	(0.0137)
mnhousehold_wealth			-0.007	(0.0098)			-0.012***	(0.0042)
mnsex_hhh_female			-0.017	(0.0498)			0.022	(0.0218)
mnSingle			-0.023	(0.0564)			-0.035	(0.0248)
mnage_hhh			0.000	(0.0010)			0.001*	(0.0005)
mnedu_JCE_atleast			-0.089***	(0.0345)			-0.052***	(0.0152)
mnhh_depend_ratio			0.000	(0.0002)			-0.000	(0.0001)
mndistance_admarc			-0.012***	(0.0026)			-0.004***	(0.0011)
Deviation from means of observed characteristics								
devhh_laborunits			0.008	(0.0073)			-0.002	(0.0037)
devfarmsize			0.426***	(0.0264)			0.134***	(0.0137)
devhousehold_wealth			0.008	(0.0121)			0.013**	(0.0058)
devfemale			0.003	(0.0436)			-0.019	(0.0209)
devSingle			-0.045	(0.0345)			-0.015	(0.0170)
devage_hhh			0.001	(0.0016)			-0.001	(0.0007)
devedu_JCE_atleast			-0.016	(0.0292)			-0.019	(0.0143)
devhh_depend_ratio			-0.000	(0.0001)			0.000	(0.0001)
devdistance_admarc			-0.015***	(0.0052)			-0.001	(0.0026)
_cons	0.288***	(0.1117)	-0.221**	(0.0961)			0.013	(0.0427)
lnalpha	-3.047***	(0.1381)	-2.677***	(0.0848)				
sigma_u					0.050**	(0.0215)	0.098***	(0.0066)
sigma_e					0.250***	(0.0055)	0.254***	(0.0038)
Observations	2913		3884		2913		3884	

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively number of crops grown, and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parenthesis are robust standard errors.*

Table E: Crop diversification decisions in Tanzania: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models reporting Coefficients

	Poisson models				**Tobit Models**			
	DRE		CRE		DRE		CRE	
Variables	b	se	b	se	b	se	b	se
Rainfall shocks and variability								
Flood shock growing season 1-year lag (z-score)	0.063	(0.0452)	0.046	(0.0361)	0.028	(0.0287)	0.030	(0.0221)
Drought shock growing season 1-year lag (z-score)	0.064***	(0.0185)	0.038***	(0.0138)	0.039***	(0.0115)	0.027***	(0.0084)
Long-term early season (Nov-Jan) rainfall variability (1980-2018) (mm)	0.002**	(0.0008)	0.003***	(0.0007)	0.002***	(0.0005)	0.002***	(0.0004)
Lagged Crop diversification indices								
Count (Simpson) index at baseline (D_0)	0.131***	(0.0108)			0.397***	(0.0548)		
Count (Simpson) index in previous panel round (D_{t-1})	0.024**	(0.0101)			0.095	(0.0665)		
Household & farm characteristics								
farmsize_gps (Farm size in ha)	0.016***	(0.0035)			0.011***	(0.0025)		
hh_laborunits (Household labor units)	0.012	(0.0076)			-0.002	(0.0051)		
household_wealth (Household wealth index)	0.265***	(0.0657)			0.150***	(0.0454)		
sex_hhh_female (female household head)	-0.002	(0.0483)			-0.004	(0.0307)		
Single (Single household head)	-0.093*	(0.0487)			-0.047	(0.0306)		
age_hhh (Age of household head)	0.002**	(0.0008)			0.001	(0.0005)		
edu_atleastD7 (Education attained at least D7)	0.024	(0.0327)			0.004	(0.0209)		
hh_depend_ratio (Household age dependency ratio)	-0.000	(0.0002)			-0.000	(0.0001)		
distancemain_districtHQ (distance to main urban market)	0.001***	(0.0003)			0.001***	(0.0002)		
Year dummies & regional dummies								
years3	-0.053**	(0.0268)	-0.083***	(0.0235)	-0.021	(0.0171)	-0.081***	(0.0145)
years2			-0.058**	(0.0228)			-0.076***	(0.0141)
Northern region(1=yes)	-0.000	(0.0776)	-0.010	(0.0791)	-0.001	(0.0504)	0.026	(0.0471)
Coast region (1=yes)	-0.011	(0.0772)	-0.008	(0.0787)	-0.036	(0.0502)	-0.005	(0.0468)
Central region(1=yes)	0.115	(0.0798)	0.055	(0.0814)	0.101*	(0.0520)	0.075	(0.0484)
Lake region(1=yes)	-0.001	(0.0812)	0.014	(0.0828)	0.009	(0.0527)	0.059	(0.0491)
Southern Highlands(1=yes)	0.148*	(0.0817)	0.149*	(0.0830)	0.111**	(0.0535)	0.137***	(0.0496)
West region (1=yes)	0.160*	(0.0869)	0.021	(0.0889)	0.105*	(0.0565)	0.046	(0.0527)
Deviation from means of observed characteristics								
devfarmsize			0.014***	(0.0044)			0.009***	(0.0030)
devhh_laborunits			0.033***	(0.0115)			0.010	(0.0072)
devZnorm_Agricassetindex			-0.051	(0.0714)			-0.081*	(0.0466)
devfemale			-0.127*	(0.0757)			-0.068	(0.0456)
devmarried_single			-0.009	(0.0595)			-0.060*	(0.0357)
devage_hhh			0.003**	(0.0013)			0.001	(0.0008)
devedu_atleastD7			0.033	(0.0420)			0.026	(0.0259)
devhh_depend_ratio			-0.000	(0.0002)			-0.000	(0.0001)
devdistancemain_districtHQ			-0.001	(0.0012)			-0.001	(0.0008)
Mean of observed characteristics								
mnfarmsize			0.040***	(0.0061)			0.022***	(0.0034)
mnhh_laborunits			-0.001	(0.0101)			-0.010*	(0.0059)
mnZnorm_Agricassetindex			0.362***	(0.0883)			0.264***	(0.0534)
mnfemale			0.096*	(0.0578)			0.053	(0.0332)
mnmarried_single			-0.154**	(0.0615)			-0.058	(0.0353)
mnage_hhh			0.000	(0.0010)			-0.000	(0.0006)
mnedu_atleastD7			-0.039	(0.0414)			-0.025	(0.0245)
mnhh_depend_ratio			0.000	(0.0002)			0.000	(0.0001)
mndistancemain_districtHQ			0.001**	(0.0003)			0.001***	(0.0002)
_cons	0.130	(0.1207)	0.449***	(0.1265)	-0.038	(0.0773)	0.037	(0.0744)
lnalpha	-2.285***	(0.1371)	-1.630***	(0.0614)				
sigma_u					0.225***	(0.0238)	0.253***	(0.0090)
sigma_e					0.365***	(0.0125)	0.359***	(0.0059)
Observations	3350		5025		3350		5025	

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively number of crops grown, and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parenthesis are robust standard errors.*

C. Attrition probit models

Table F:Probit estimation of attrition bias in samples of respective countries.

	Malawi	Tanzania
VARIABLES	*Drop out in 2013(1=yes)*	*Drop out in 2011(1=yes)*
Female household head(1=yes)	-0.2383	0.2506*
	(0.1535)	(0.1024)
Age of household head(years)	-0.0083*	-0.0063*
	(0.0042)	(0.0031)
Household size (count)	-0.1091	-0.0987*
	(0.0577)	(0.0408)
Household labor units	0.0063	0.1133
	(0.1026)	(0.0718)
Farm size(ha)	-0.8385***	-0.0498**
	(0.2089)	(0.0159)
Household wealth index (PCA)	0.0773***	0.0411*
	(0.0208)	(0.0180)
Distance to nearest market (Km)	-0.0092	-0.0015
	(0.0119)	(0.0021)
Number of plots	-0.2052*	-0.3322***
	(0.0824)	(0.0566)
_cons	0.2023	-0.3559
	(0.2433)	(0.1908)
LR chi2(8)	120.59	114.61
Prob > chi2	0.0000	0.0000
Observations	1 144	2 063

Normal standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; The dependent variable is a dummy for household dropping out in follow up sample from baseline sample. The baseline sample for Malawi is 2010 survey and 2008 survey for Tanzania.

D. Robustness checks: Main results where we test and control for possible attrition bias by including the Inverse mills ratio (IMR) from attrition probit models as an additional explanatory variable

Table G: Crop diversification decisions in Malawi: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models with Inverse mills ratio from attrition probit models reporting Coefficients

	Poisson models				**Tobit Models**			
	DRE		CRE		DRE		CRE	
Variables	b	se	b	se	b	se	b	se
Rainfall shocks and variability								
Flood shock growing season 1-year lag (z-score)	0.049	(0.0478)	0.003	(0.0462)	0.034	(0.0237)	0.016	(0.0237)
Drought shock growing season 1-year lag (z-score)	0.003	(0.0176)	0.025*	(0.0146)	0.003	(0.0088)	0.019***	(0.0071)
Long-term early season (Nov-Jan) rainfall variability (1980-2018) (mm)	0.008***	(0.0009)	0.007***	(0.0005)	0.002***	(0.0004)	0.002***	(0.0002)
Lagged Crop diversification indices								
Count (Simpson) index at baseline (D_0)	0.038***	(0.0061)			0.107***	(0.0252)		
Count (Simpson) index in previous panel round (D_{t-1})	0.023***	(0.0050)			0.077**	(0.0358)		
Household & farm characteristics								
farmsize_gps(Farm size in ha)	0.414***	(0.0216)			0.127***	(0.0103)		
hh_laborunits (Household labor units)	0.008	(0.0060)			-0.003	(0.0028)		
household_wealth (Household wealth index)	-0.003	(0.0078)			-0.002	(0.0033)		
sex_hhh_female (female household head)	0.004	(0.0309)			-0.002	(0.0138)		
Single (Single household head)	-0.029	(0.0300)			-0.016	(0.0139)		
age_hhh (Age of household head)	-0.0001	(0.0008)			0.0001	(0.0004)		
edu_JCE_atleast (Education attained at least JCE)	-0.018	(0.0240)			-0.028***	(0.0109)		
hh_depend_ratio (Household age dependency ratio)	0.0001	(0.0001)			-0.0001	(0.0001)		
distance_admarc (distance to ADMARC)	-0.004	(0.0024)			-0.0001	(0.0011)		
Year dummies & regional dummies								
years4	0.244***	(0.0236)	0.378***	(0.0371)	0.076***	(0.0117)	0.132***	(0.0188)
years3	-0.029	(0.0264)	0.106***	(0.0377)	-0.107***	(0.0125)	-0.050***	(0.0180)
years2			0.136***	(0.0351)			0.052***	(0.0176)
reg_northern	0.036	(0.0624)	0.050	(0.0495)	-0.019	(0.0260)	0.004	(0.0208)
reg_southern	-0.004	(0.0408)	0.118***	(0.0321)	-0.015	(0.0172)	0.003	(0.0138)
Mean of observed characteristics								
mnhh_laborunits			0.005	(0.0101)			-0.004	(0.0045)
mnfarmsize_gps			0.516***	(0.0307)			0.190***	(0.0137)
mnhousehold_wealth			-0.006	(0.0099)			-0.012***	(0.0042)
mnsex_hhh_female			-0.018	(0.0498)			0.021	(0.0218)
mnSingle			-0.023	(0.0564)			-0.035	(0.0248)
mnage_hhh			0.000	(0.0010)			0.001*	(0.0005)
mnedu_JCE_atleast			-0.090***	(0.0345)			-0.052***	(0.0152)
mnhh_depend_ratio			0.000	(0.0002)			-0.000	(0.0001)
mndistance_admarc			-0.012***	(0.0026)			-0.004***	(0.0011)
Deviation from means of observed characteristics								
devhh_laborunits			0.008	(0.0073)			-0.002	(0.0037)
devfarmsize			0.426***	(0.0264)			0.134***	(0.0137)
devhousehold_wealth			0.008	(0.0121)			0.013**	(0.0058)
devfemale			0.003	(0.0436)			-0.019	(0.0209)
devSingle			-0.045	(0.0345)			-0.015	(0.0170)
devage_hhh			0.001	(0.0016)			-0.001	(0.0007)
devedu_JCE_atleast			-0.016	(0.0292)			-0.019	(0.0143)
devhh_depend_ratio			-0.000	(0.0001)			0.000	(0.0001)
devdistance_admarc			-0.015***	(0.0052)			-0.001	(0.0026)
Inverse mills Ratio (IMR) from attrition probit model	-0.013	(0.0176)	0.006	(0.0178)	-0.006	(0.0073)	0.005	(0.0077)
_cons	0.319***	(0.1196)	-0.236**	(0.1059)	0.208***	(0.0521)	0.0001	(0.0468)
lnalpha	-3.047***	(0.1381)	-2.677***	(0.0848)				
sigma_u					0.050**	(0.0214)	0.098***	(0.0066)
sigma_e					0.250***	(0.0055)	0.254***	(0.0038)
Observations	2913		3884		2913		3884	

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively number of crops grown, and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parenthesis are robust standard errors.*

Table H: Crop diversification decisions in Tanzania: Dynamic Random Effects (DRE) and Correlated Random Effect (CRE) Poisson and Tobit models with Inverse mills ratio from attrition probit models reporting Coefficients

	Poisson models				**Tobit Models**			
	DRE		CRE		DRE		CRE	
Variables	b	se	b	se	b	se	b	se
Rainfall shocks and variability								
Flood shock growing season 1-year lag (z-score)	0.074	(0.0452)	0.072**	(0.0359)	0.031	(0.0287)	0.037*	(0.0221)
Drought shock growing season 1-year lag (z-score)	0.061***	(0.0185)	0.041***	(0.0137)	0.039***	(0.0115)	0.027***	(0.0084)
Long-term early season (Nov-Jan) rainfall variability (1980-2018) (mm)	0.002***	(0.0008)	0.003***	(0.0006)	0.002***	(0.0005)	0.002***	(0.0004)
Lagged Crop diversification indices								
Count (Simpson) index at baseline (D_0)	0.118***	(0.0110)			0.387***	(0.0548)		
Count (Simpson) index in previous panel round (D_{t-1})	0.021**	(0.0101)			0.094	(0.0664)		
Household & farm characteristics								
farmsize_gps (Farm size in ha)	0.011***	(0.0036)			0.010***	(0.0026)		
hh_laborunits (Household labor units)	0.006	(0.0077)			-0.003	(0.0052)		
household_wealth (Household wealth index)	0.267***	(0.0656)			0.149***	(0.0454)		
sex_hhh_female (female household head)	0.020	(0.0484)			0.001	(0.0308)		
Single (Single household head)	-0.083*	(0.0486)			-0.045	(0.0306)		
age_hhh (Age of household head)	0.002**	(0.0008)			0.001	(0.0005)		
edu_atleastD7 (Education attained at least D7)	0.018	(0.0326)			0.003	(0.0209)		
hh_depend_ratio (Household age dependency ratio)	-0.000	(0.0002)			-0.000	(0.0001)		
distancemain_districtHQ (distance to main urban market)	0.001***	(0.0003)			0.001***	(0.0002)		
Year dummies & regional dummies								
years3	-0.059**	(0.0269)	-0.083***	(0.0235)	-0.022	(0.0171)	-0.081***	(0.0145)
years2			-0.051**	(0.0227)			-0.075***	(0.0141)
Northern region(1=yes)	0.012	(0.0776)	-0.001	(0.0750)	0.000	(0.0504)	0.023	(0.0467)
Coast region (1=yes)	-0.007	(0.0772)	-0.031	(0.0746)	-0.037	(0.0502)	-0.016	(0.0464)
Central region(1=yes)	0.119	(0.0797)	0.032	(0.0773)	0.100*	(0.0520)	0.064	(0.0479)
Lake region(1=yes)	0.008	(0.0811)	0.015	(0.0785)	0.010	(0.0526)	0.053	(0.0487)
Southern Highlands(1=yes)	0.164**	(0.0817)	0.162**	(0.0787)	0.113**	(0.0535)	0.136***	(0.0491)
West region (1=yes)	0.179**	(0.0868)	0.068	(0.0843)	0.108*	(0.0565)	0.054	(0.0522)
Deviation from means of observed characteristics								
Devfarmsize			0.014***	(0.0045)			0.009***	(0.0030)
devhh_laborunits			0.031***	(0.0114)			0.010	(0.0072)
devZnorm_Agricassetindex			-0.048	(0.0712)			-0.079*	(0.0466)
devfemale			-0.126*	(0.0761)			-0.068	(0.0456)
devmarried_single			-0.007	(0.0598)			-0.060*	(0.0357)
devage_hhh			0.003**	(0.0013)			0.001	(0.0008)
devedu_atleastD7			0.032	(0.0420)			0.026	(0.0259)
devhh_depend_ratio			-0.000	(0.0002)			-0.000	(0.0001)
devdistancemain_districtHQ			-0.000	(0.0011)			-0.001	(0.0008)
Mean of observed characteristics								
Mnfarmsize			0.000	(0.0060)			0.012***	(0.0038)
mnhh_laborunits			-0.031***	(0.0097)			-0.018***	(0.0060)
mnZnorm_Agricassetindex			0.474***	(0.0831)			0.288***	(0.0530)
mnfemale			0.175***	(0.0546)			0.075**	(0.0330)
mnmarried_single			-0.115**	(0.0578)			-0.049	(0.0350)
mnage_hhh			-0.001	(0.0010)			-0.001	(0.0006)
mnedu_atleastD7			-0.043	(0.0389)			-0.026	(0.0242)
mnhh_depend_ratio			-0.000*	(0.0002)			-0.000	(0.0001)
mndistancemain_districtHQ			0.001***	(0.0003)			0.001***	(0.0002)
Inverse mills Ratio (IMR) from attrition probit model	0.134***	(0.0276)	0.415***	(0.0288)	0.030*	(0.0177)	0.110***	(0.0182)
_cons	-0.082	(0.1283)	-0.181	(0.1272)	-0.089	(0.0827)	-0.124	(0.0784)
lnalpha	-2.314***	(0.1396)	-1.860***	(0.0683)				
sigma_u					0.225***	(0.0238)	0.248***	(0.0089)
sigma_e					0.364***	(0.0124)	0.359***	(0.0059)
Observations	3350		5025		3350		5025	

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively number of crops grown, and the Simpson index. The asterisk represents ***p<0.001, **<0.05, *<0.1, shown in parenthesis are robust standard errors.*

Table I: Crop diversification decisions in Tanzania: Dynamic Random Effects (DRE) and Correlated Random Effect (*CRE) Poisson and Tobit models reporting Average partial effects (APE)-Without attrition adjustment*

	Poisson models		Tobit Models	
Variables	**DRE**	**CRE**	**DRE**	**CRE**
Rainfall shocks and variability				
Flood shock growing season 1-year lag (z-score)	0.063	0.046	0.014	0.015
	(0.0452)	(0.0361)	(0.0141)	(0.0111)
Drought shock growing season 1-year lag (z-score)	0.064***	0.038***	0.019***	0.013***
	(0.0185)	(0.0138)	(0.0057)	(0.0042)
Long-term early season (Nov-Jan) rainfall variability (1980-2018) (mm)	0.002**	0.003***	0.001***	0.001***
	(0.0008)	(0.0007)	(0.0003)	(0.0002)
Lagged Crop diversification indices ($D_{i,t-n}$)				
Count (Simpson) index at baseline (D_0)	0.131***		0.196***	
	(0.0108)		(0.0266)	
Count (Simpson) index in previous panel round (D_{t-1})	0.024**		0.047	
	(0.0101)		(0.0328)	
Household & farm characteristics (X)				
Observed household characteristics	Yes	No	Yes	No
Mean of Observed household characteristics	No	Yes	No	Yes
Deviation from means of Observed household characteristics	No	Yes	No	Yes
Year dummies & regional dummies (α)				
2013 (1=yes)	-0.053**	-0.083***	-0.010	-0.041***
	(0.0268)	(0.0235)	(0.0084)	(0.0073)
2011 (1=yes)		-0.058**		-0.038***
		(0.0228)		(0.0071)
Northern region(1=yes)	-0.000	-0.010	-0.001	0.013
	(0.0776)	(0.0791)	(0.0248)	(0.0236)
Coast region (1=yes)	-0.011	-0.008	-0.018	-0.003
	(0.0772)	(0.0787)	(0.0247)	(0.0235)
Central region(1=yes)	0.115	0.055	0.049*	0.038
	(0.0798)	(0.0814)	(0.0256)	(0.0243)
Lake region(1=yes)	-0.001	0.014	0.005	0.030
	(0.0812)	(0.0828)	(0.0259)	(0.0247)
Southern Highlands(1=yes)	0.148*	0.149*	0.055**	0.069***
	(0.0817)	(0.0830)	(0.0264)	(0.0249)
West region (1=yes)	0.160*	0.021	0.052*	0.023
	(0.0869)	(0.0889)	(0.0278)	(0.0264)
Constant	0.130	0.449***	-0.038	0.037
	(0.1207)	(0.1265)	(0.0773)	(0.0744)
lnalpha/ sigma_u	-2.285***	-1.630***	0.225***	0.253***
	(0.1371)	(0.0614)	(0.0238)	(0.0090)
sigma_e			0.365***	0.359***
			(0.0125)	(0.0059)
Panel households	1675	1675	1675	1675
Observations	3350	5025	3350	5025

*Notes: The dependent variables for DRE and CRE Poisson and Tobit models are respectively number of crops grown, and the Simpson index. The asterisk represent ***p<0.001, **<0.05, *<0.1, shown in parenthesis are robust standard errors; Household and farm characteristics included as controls include farm size(ha), household labor units, asset wealth index(PCA), household age dependency ratio(%), distance to agricultural markets(km), age(years) of household head, dummies for female and single headed households, and a dummy for household head attaining at least standard 7(D7). The reference for regions is Zanzibar zone. Year 2008 and 2011 are reference dummies in CRE and DRE models, respectively.*

Paper IV

Smallholder access to purchased seeds in the presence of pervasive market imperfections and rainfall shocks: Panel Data Evidence from Malawi and Ethiopia

Clifton Makate[1*] , Arild Angelsen[1], Stein Terje Holden[1], & Ola Tveitereid Westengen[2]

[1*]School of Economics and Business, Norwegian University of Life Sciences, P.O.Box 5003, 1432 Ås, Norway; clifton.makate@nmbu.no (C.M); arild.angelsen@nmbu.no (A.A); stein.holden@nmbu.no (S.T.H)
[2]Department of International Environment and Development Studies, Noragric, Norwegian University of Life Sciences, Box 5003, 1432 Ås, Norway; ola.westengen@nmbu.no (O.T.W.)
* Corresponding author: clifton.makate@nmbu.no (C.M)

Abstract
Seed purchasing enables farmers to respond to adverse events that may cause chronic and temporary seed insecurity by allowing them to exploit opportunities associated with accessing new seeds. However, as with other inputs, seed purchasing is complicated by pervasive market imperfections and climate risk common in Sub Sahara Africa. This study uses balanced household panel data for Malawi (2010-2018) and Ethiopia (2012-2016), and applies dynamic random effects Probit and Tobit models to assess how seed purchase decisions are affected by earlier participation in the market, lagged rainfall shocks, and historical climate variables. Our findings show that there are non-linear effects of lagged seed purchase decisions on subsequent decisions with strong initial effects (weakening over time). For instance, initial maize seed purchase decisions are associated with between 11-13% (1kg) and 21-27% (2kgs) higher probability (intensity/household) of purchase in later rounds in Malawi and Ethiopia, respectively. Seed purchase decisions also respond to climate variability and shocks. For instance, lagged drought shocks enhance subsequent maize purchase decisions in Malawi and Ethiopia. Historical average rainfall and temperature enhance maize seed purchase decisions in both countries. Overall, results point to state dependency on the demand side of the seed market, leading to selective access to purchased seeds. Also, seed purchase in smallholder farming is a liquidity and risk-dependent input choice. To enhance access to seed through purchase and support adaptation to rainfall shocks, policy efforts need to continue targeting reducing transaction costs and other barriers to entry into seed markets.

Keywords: non-linear transaction costs; household seed security; dynamic random effect models; smallholder farmers; rainfall shocks; Malawi & Ethiopia.

1. Introduction

Access to a diversity of good quality seeds is crucial for smallholder farmers' food production, nutrition, and resilience in the face of climate change and natural disasters. Smallholder farmers in developing countries access seeds through formal and informal seed systems (FAO 1998; Sperling et al. 2008). Historically, most smallholder farmers save seeds from their previous harvests, a strategy that reduces the costs associated with purchasing seeds and provides seeds of guaranteed quality and well adapted to their local agroecology (Tripp 2006; Nordhagen and Pascual 2013). However, to meet diverse needs and challenges, smallholder farmers also source seeds outside the farm (Bellon et al. 2006; Coomes et al. 2015). Access to seed off-farm is important in complementing farmer-saved seed and enhancing overall household seed security. Seed security exists when both men and women within farming households have ready access to sufficient quantities of quality seeds and planting materials of preferred crop varieties, adapted to their local agroecological conditions and socioeconomic needs, at planting times in both good and bad seasons (FAO 1998; FAO and ECHA 2015). Access to seeds off-farm becomes even more critical when farmers want to access new seeds and want to grow new crops, when farmers' stock of farmer saved seed has depleted (e.g., through destruction from pests and disasters or family consumption in periods of food scarcity), or when the quality of seed stored has degenerated (Almekinders et al. 1994; Almekinders et al. 2007; Nordhagen and Pascual 2013). While farmers may access new seed free of charge through emergency aid and or social networks, the bulk of new seed is paid for in cash, either at local markets or from agro-dealers (McGuire and Sperling 2016; Sperling 2020).

As with other farming inputs, access to new seeds through the market is complicated by the pervasive imperfections that characterize many markets in the developing world and Sub Sahara Africa (SSA) in particular. Missing information on commodity prices and technologies, credit constraints, high transaction costs, and poor infrastructure make it difficult for smallholder farmers to fully engage in input and output markets (Fafchamps 2004; Dorward et al. 2005; Shiferaw et al. 2008; Markelova et al. 2009). For instance, smallholder farmers' access to new seeds through local and regional/national markets is complicated by transaction costs associated with participating in those markets. With imperfect factor markets that are not well integrated, developed, and spatially dispersed, smallholder farmers face dynamically variable transaction costs (Binswanger and Rosenzweig 1986; Key et al. 2000; Renkow et al. 2004; Barrett 2008; Holden et al. 2010; Ricker-Gilbert and Chamberlin 2018). Following Coase (1937) and North (1987), transaction costs are the costs incurred in making a market transaction, excluding the actual price paid for the commodity. These include costs associated with: (i) searching and attracting potential trading partners, including pre-sale inspection, (ii) negotiation, contracting, and fulfilment costs, and (iii) monitoring and implementation costs(Coase 1937; North 1987). Such costs can significantly influence decisions by households on whether to participate or not to participate in the market. This is because transaction costs raise the price effectively paid by buyers and lower the price effectively received by sellers of a good creating a price range within which some households may find it unprofitable either to sell or buy (De Janvry et al. 1991; Key et al. 2000). In seed markets and on the demand side, such transaction costs may include the costs of searching and obtaining information on production and consumption traits of the seed of different crops and or varieties, costs of searching and locating them, and negotiation costs.

To overcome challenges posed by imperfect factor markets and maintain their position in the market, smallholder farmers invest their time in establishing localized information networks (Fafchamps 2004) and engaging in collective action (Markelova et al. 2009). Over time, such efforts help farmers reduce transaction costs and enhance their linkage to factor markets. This study focuses on smallholder farmers' seed purchase decisions in Malawi and Ethiopia and investigates the extent to which access to purchased seeds is constrained (or facilitated) by state dependency and other factors. State dependency in markets implies that market participants capitalize on their experience and established networks gained through repeated engagements in the markets to identify trading partners, which is not the case with new entrants without such experience and networks in factor markets (Fafchamps 2004; Gebru et al. 2019; Tione and Holden 2021).

Given the prevalence of pervasive and non-linear transaction costs in input markets in SSA (Key et al. 2000; Renkow et al. 2004; Barrett 2008; Holden et al. 2010; Ricker-Gilbert and Chamberlin 2018; Tione and Holden 2021) and subsequent low access to seed through formal markets (Tripp 2006; McGuire and Sperling 2013; Nordhagen and Pascual 2013), investigating the extent of state dependency in seed markets is important. Therefore, this paper investigates the extent to which smallholder farming households' access to purchased seeds is constrained (or facilitated) by state dependency (past market access), farmer characteristics, community characteristics, lagged rainfall shocks, and long-term average climate (rainfall and temperature). Empirical evidence from such a study will help inform policies that aim to reduce barriers to entry in seed markets (through purchase) to enhance access to new seeds to complement farmers' saved seed (and seed from other sources) and enhance overall seed security under shocks for better livelihoods in smallholder farming.

We compare Malawi and Ethiopia- two countries with contrasting features in terms of the policy framework governing the development of seed systems that farmers use. Key differences in policies and institutions governing formal seed systems in Malawi and Ethiopia lie in the roles played by the government and the private sector seed value chains (Langyintuo et al. 2010; Kassie et al. 2013; Erenstein and Kassie 2018; Westengen et al. 2019). In Malawi, the seed industry is characterized by the dominance of the government as buyer and distributor of seed and a high market share and power of few private seed companies (Kassie et al. 2013). On the contrary, the government dominates Ethiopia's formal seed system in all functions and for most crops, with the private sector having a minimal role. In both countries, seed policies and regulations have evolved, with the efforts directed towards the growth of the formal systems. However, in Malawi, policy efforts have mainly targeted the facilitation and growth of the formal sector while Ethiopia currently adopts a pluralistic approach that aims to target the growth of formal, intermediate, and informal sectors (Westengen et al. 2019; Mulesa et al. 2021). The differences in policy frameworks governing seed systems in the two countries could offer different constraints and opportunities to reduce transaction costs in accessing seeds through purchasing.

We use four (three) panel rounds of the Malawi (Ethiopia) Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) household data combined with historical monthly weather data from WorldClim (Masarie and Tans 1995; Fick and Hijmans 2017) to construct balanced household panel data. We rely on the theory of dynamic non-linear transaction costs in factor markets (Holden et al. 2007; Holden et al. 2010) and estimate dynamic Probit and Tobit random-effects models (Wooldridge 2005) to assess the extent to which non-linear transaction costs in seed markets, climate shocks, market access, and other factors influence access to off-farm seed while controlling for unobserved household heterogeneity that might influence participation in the seed market by smallholder farmers.

The rest of this article is organized as follows: the next section describes the theoretical framework and specifies the study hypotheses. Section 3 outlines the empirical estimation approach and data, while section 4 presents the results. Finally, section 5 discusses the results, while section 6 concludes the article.

2. Theoretical framework

Theory and evidence point to poorly developed factor markets in some parts of SSA. The implication is that market access by farming households is not uniform, as they may face different transaction costs to participation (Binswanger and Rosenzweig 1986; Key et al. 2000; Renkow et al. 2004; Barrett 2008). In addition, geographic markets are spatially dispersed and not well integrated into the global economy because of differences in costs of commerce and the disparities in the degree of competition among marketing intermediaries (Fackler and Goodwin 2001; Barrett 2008). With such imperfections, market participants may face different participation transaction costs, which may change over time (Holden et al. 2007; Holden et al. 2010; Tione and Holden 2021). In SSA, such costs are high, and they emerge from policies, institutions, and other socioeconomic factors that contribute to high information asymmetries and differential access and use of productive resources by households (Key et al. 2000; Renkow et al. 2004; Barrett 2008; Holden et al. 2010; Ricker-Gilbert and Chamberlin 2018; Tione and Holden 2021).

Transaction costs related to market engagement affect both the demand side (e.g., acquiring inputs from the market) and the supply side(e.g., delivering farm produce to the market). Transaction costs on the demand-side include expenditures incurred when conducting market transactions for inputs other than the price, including costs associated with search, negotiation, supervision, and bargaining. In contrast, farm-to-market transaction costs include the costs associated with the trading output produced from the farm. This paper focuses on transaction costs associated with access to seed through purchase from the market. In terms of seed access through purchase, transaction costs incurred may include the costs of searching and getting information on production and consumption traits of farmers' preferred seed for their different crops and or varieties, costs of searching and locating them, and costs of negotiating and making the transactions, excluding the price paid for the seeds. Farmers require new seeds every year to fulfill their production activities, and they can access them through their seed savings, relief, subsidies (e.g., coupons), or purchases. Saving own seed is a common practice by smallholder farmers in SSA with low transaction costs (Tripp 2006), but sometimes faces storage and seed quality-related challenges. Access to relief or subsidized inputs is another important seed source in SSA, particularly following the recent revitalization of government subsidy programs (Jayne and Rashid 2013). However, such programs target specific farmers based on underlying objectives and rarely meet the farmer's demand for inputs. Besides, farmers also incur transaction costs in accessing subsidized inputs. Seed purchases allow the farmers to access new seeds and supplement other seed sources. Given the poorly developed markets common in some parts of SSA, access to purchased seed may be restricted by high transaction costs associated with participation by farmers as buyers in seed markets. Overall, information asymmetries, limited knowledge by farmers, resource constraints, and uncertainties related to future weather add to imperfect information and transaction costs that influence access to seed by farmers (Binswanger and Rosenzweig 1986; Crawford et al. 2003; Barrett 2008).

Survival and progression of societies in the presence of pervasively imperfect factor markets require production relations[1] that allow farming households to effectively carry out their current and inter-temporal decisions (Binswanger and Rosenzweig 1986). For households to achieve (i) high incomes and consumption and (ii) even out consumption over time by avoiding risk and disasters and making provisions for dealing with the consequences of unavoidable and unforeseen risks and disasters, production relations should adapt to current and inter-temporal problems by the people (Binswanger and Rosenzweig 1986). In the context of seed markets, farmers will need to adapt to the high transaction costs and other factors that limit their access to seed through the market. For example, farming households may engage in collective action and build their social networks over time to overcome or reduce transaction costs in buying seed from the market. Upon entering the seed market for the first time, farming households may invest in establishing networks of information and social capital that may help them face lower transaction costs in subsequent years (Binswanger and Rosenzweig 1986; Key et al. 2000; Fafchamps 2004; Barrett 2008). Therefore, past trade experience in the seed market may affect current seed market access and intensity of participation. We, hence, expect to find state dependency when analyzing farm household panel data capturing seed purchasing decisions over time.

In line with previous studies that have applied dynamic transaction costs models in studying mainly household land rental market decisions (Holden et al. 2007; Holden et al. 2010; Gebru et al. 2019; Tione and Holden 2021), we study dynamic seed purchase decisions in smallholder farming in Malawi and Ethiopia. Following dynamic transaction costs models, household inter-temporal decisions to purchase seed may be expressed as in Eqn1:

$$\bar{P}_t^H = \sum_S \bar{P}_t^{HS} \left(c_t^{HS} \left[c_0 + c_t^{HS} \left\{ \bar{E}_t^H, \bar{P}_{t-n}^{HS}, R_v, \mathcal{R}_{t-1}, \mathbb{C}_v, M_{vt}, \int_{t-e}^{t} G dt; v_t^w, v_t^f \right\} \right] \right) \qquad \text{[Eqn1]}$$

The dynamic model specified in Eqn1 states that a household's access to purchased seeds (P) at a time (t) is the sum of access to seeds from seed suppliers in the seed market (S), represented as $(\sum_S \bar{P}_t^{HS})$. This access is itself a function of transaction costs (c) that consist of two components: an initial fixed cost component (c_0), and a non-linear variable transaction cost (c_t^{HS}). The variable non-linear transaction cost components (c_t^{HS}) depend on both observable and unobservable factors, which include the household's resource endowments (land, labor, and household assets) $(\bar{E}_t^H)$, previous participation in the seed market $(\bar{P}_{t-n}^{HS})$ that help accumulate knowledge and experience over time. These non-linear transaction costs are not directly observable but can be identified by investigating the influence of households' previous participation in the seed market on later participation decisions using panel data on seed purchase decisions. We capture previous participation in the seed market $(\bar{P}_{t-n}^{HS})$ using lagged market participation variables that capture both participation and intensity of participation, including initial survey year participation variables and participation variables for the previous survey round(t-1).

In addition, R_v, $\mathcal{R}_{t-1}$, $\mathbb{C}_v$, respectively, captures long-term average rainfall, lagged rainfall shocks (1-year lag drought and flood shock), and long-term average temperature (38-year average)) for the main crop growing season. Recurrent erratic rains, and weather shocks that lead to crop failure, may disrupt farmer stocks of their own saved seed or make it hard to set aside seed from harvest due to urgent consumption needs, forcing farmers to source seed from elsewhere through trade (Bellon et al. 2011; Nordhagen and Pascual 2013). However, for

[1] Production relations according to Binswanger and Rosenzweig (1986) refer to the relations of people to factors of production, and corresponding relations of people among each other as factor owner and renters (e.g., as tenants, landlords, employers, workers, debtors, creditors).

rural smallholder farming households operating under uncertain production environments with imperfect credit and insurance markets, recurrent rainfall variability may also impose liquidity constraints, limiting technology adoption and input use decisions such as seed purchase. Therefore, response to rainfall shocks is complex, as households may switch from selling food (relaxed liquidity constraints) in years with good rainfall and becoming net buyers in years with poor rainfall (tighter liquidity constraints). Seed purchase is a liquidity-dependent risky input (determined both by the level of liquidity constraints and the degree of uncertainty in the production environment) which implies that it may directly respond to measures of rainfall variability and shocks.

The component (M_{vt}) captures community market access. Households in communities with better access to market and market infrastructure are likely to face lower transaction costs in accessing seed and may have higher chances of participating in the seed market through purchase. The component ($\int_{t-e}^{t} Gdt$) captures the dynamic effects of policy changes that may influence transaction costs in the seed market over time. More so, the spatial nature of the seed market implies that access to purchased seed is location specific and conditional on household characteristics (v_t^w), and community characteristics (v_t^f) including agroecological conditions, population pressure, and market access. Based on this theoretical model, we seek to test a few hypotheses:

(H1). There is persistent state dependency in the smallholder farmer seed purchases causing selective access to purchased seeds over time. We hence expect to find lagged seed purchase variables (purchase and extent of purchase) to explain latter participation and extent of participation strongly and positively in the seed market.

(H2). Long-term average rainfall in the crop growing season positively affects seed purchase decisions.

(H3): Lagged rainfall shocks positively influence seed purchase decisions.

(H4). The likelihood and intensity of seed purchasing (participation and extent) increase with market access within the community and household wealth endowments.

3. Data and Estimation strategy

3.1. Data

We rely on household survey data from Malawi and Ethiopia, available through the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA), to study dynamic seed purchase decisions. The LSMS-ISA data, commonly known as the Ethiopia Socioeconomic Survey (ESS) and Integrated Household Survey (IHS) in Ethiopia and Malawi, respectively collect comprehensive information on agricultural activities and household living conditions in respective countries. This study uses data from rural households engaged in agricultural activity with complete and usable information on seed purchasing.

We constructed a three-year, balanced household panel for Ethiopia of 2 398 rural households interviewed successively in three-panel rounds (2011/12, 2013/14, 2015/16). The three-year household panel for Ethiopia started with 3 969 households, of which 3 466 (87%) were rural in 2011/12. We trace rural households successively interviewed in all three rounds, with consistent household identification information, those who engage in some agricultural activity, and usable information on seed use, particularly seed purchasing, to construct a balanced panel. Similarly, for Malawi, we rely on four rounds of Malawi LSMS-ISA data conducted in 2010, 2013, 2016, and 2019. The Malawi panel survey started with 1 619 households, with about 71% (1 144) rural households traced in four successive rounds. We constructed a four-year balanced panel data of 971 households for Malawi, which we analyze in this paper. The loss in households from the baseline to subsequent rounds could lead to attrition bias in estimation. We hence use probit models (one for each studied country) to assess and control for possible attrition bias in all our results. The probit models (Table D in supplementary material) use dummy variables(1=yes) for dropping out in the follow-up survey from the baseline surveys. We did not observe a significant attrition bias effect in our results. We, hence, present the results where we include the inverse mills ratio for testing and controlling for attrition bias in our analysis as part of the supplementary material (Tables F and G). We provide more detail on steps to test and control for potential attrition bias in later sections (estimation strategy).

The LSMS-ISA household survey data for Malawi and Ethiopia are supplemented with community-level information gathered through focus group interviews. The community-level information captures various information that defines the communities' access to basic services, infrastructure, and market access. We use such data to define the market access index (elaborated in the next section-estimation strategy). Besides the household and community information, we also gather historical climate (rainfall and temperature) data for clusters(villages) from where households were interviewed and use it to define long-term average climate (rainfall and temperature) and lagged rainfall shock (1-year lag) variables. We specifically use historical monthly weather data from WorldClim[2] (Masarie and Tans 1995; Fick and Hijmans 2017) to define (i) long-term average rainfall, (ii) 1-year lag rainfall shocks(1-year lag drought and flood shock), and (iii) long-term average temperature. We define a 1-year lag rainfall shock as a normalized deviation of rainfall received in the previous season (1-year lag) from the expected seasonal rainfall, as defined by its historical average. Accordingly, we define the 1-year lag of rainfall shock($\mathcal{R}_{t-1}$) as follows: $\mathcal{R}_{t-1}=\left[\frac{\mathrm{rain}_{\mathrm{vt-1}}-\overline{\mathrm{rain}_{\mathrm{v}}}}{\sigma_{\mathrm{rain}_{\mathrm{v}}}}\right]$, where $\mathcal{R}_{t-1}$ is a rainfall shock measure for a cluster(village) (v) in the year (t-1), and $rain_{vt-1}$ is the observed amount of rainfall in the previous season, $\overline{rain}_v$ is the historical average seasonal rainfall for the village(v) for the period (1980-2018), and, σ_{rain_v} is the standard deviation of rainfall during the same period. The resultant rainfall shock is a Z-score with negative (below average) and positive (above average) deviations. We split the variable into a drought shock (the absolute values of below-average deviations) and a flood shock(above average deviations) which measures the extent of below and above-average rainfall deviations from the expected mean (historical average). In addition to the 1-year lag rainfall shock ($\mathcal{R}_{t-1}$), and long-term average rainfall (R_v) we also include long-term average maximum temperature ($\mathbb{C}_v$). We incorporate long-term average maximum temperature in our analysis mainly to avoid potential omitted variable bias, given that crop production decisions respond both to rainfall and temperature. We present the distribution of the three climate variables we incorporate in our analysis in **Figure 1**.

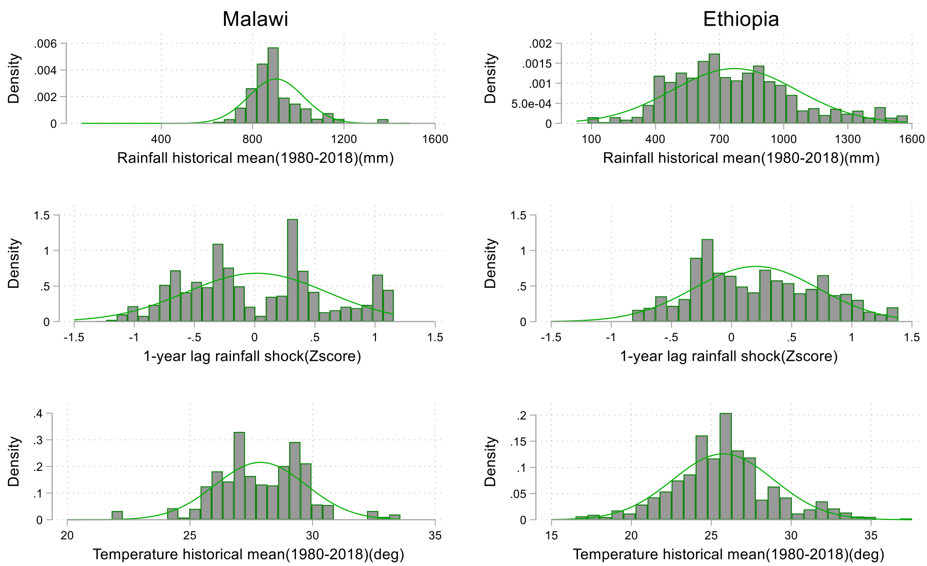

Figure 1: *Histograms showing the distribution of climate variables used in the analysis based on WorldClim data*

For our dynamic random-effects Probit and Tobit models for seed purchasing in Malawi and Ethiopia, we use the initial survey rounds in 2010 and 2012 as baseline survey rounds. Therefore, seed purchase variables (seed purchase and quantities purchased) in the baseline surveys are used as initial year participation variables and have the same values in all the successive survey rounds. Also, we used seed purchase variables for the previous survey round to define lagged seed purchase variables (seed purchase and quantity purchased). Therefore, for Malawi, we use seed purchase variables for 2016, 2013, and 2010 survey rounds as lagged seed

[2] https://www.worldclim.org/data/monthlywth.html

purchase variables for 2019, 2016, and 2013 respectively. Similarly, seed purchase variables in the 2014 and 2012 survey rounds are used for Ethiopia as lagged seed purchase variables for the 2016 and 2014 survey rounds, respectively. The final data sets we use for our dynamic random effect models for Malawi and Ethiopia comprise three and two-panel rounds, respectively, as the initial round(baseline) is lost because we do not have lagged seed purchase variables. We use these lagged seed purchase variables to test for dynamic state-dependent effects in the seed market in the studied countries. If we find initial and lagged seed purchase variables significantly enhancing later seed purchases, this will confirm the importance of non-linear transaction costs in the seed market, which highly influence access to seed through purchase.

3.2. Estimation strategy

We study dynamic seed purchase decisions using dynamic Probit and Tobit models (Wooldridge 2005). The dynamic Probit and Tobit models incorporate our key variables of interest (lagged market participation, long-term climate (rainfall and temperature), lagged rainfall shocks(1-year lag drought and flood shocks), and market access), from which we will test our hypotheses. Following Wooldridge (2005), we specify the dynamic Probit model for seed purchase as follows:

$$D\left(\pi_{jt}^{HS} = 1 \mid \pi_{j0}^{HS}, \pi_{j,t-n}^{HS}, \mathrm{R}_v, \mathcal{R}_{t-1}, \mathbb{C}_v, X_{jt}, b_j\right) = \Phi\left(\rho_1 \pi_{j0}^{HS} + \rho_2\, \pi_{j,t-n}^{HS} + \mathrm{R}_v \vartheta + \mathcal{R}_{t-1}\varphi + \mathbb{C}_v \omega + X_{jt}'\theta + b_j\right). \quad \text{[Eqn2]}$$

The dependent variable for this model (π_{jt}^{HS}) is a dummy variable measuring whether the household (j) purchased seed from the seed market (S) at a time (t). The subscript ($t-n$) communicates participation in the previous survey round (n). The equation is conditioned on several explanatory variables, including: (i) dummy variable for initial survey round seed market participation (π_{j0}^{HS}) which remains the same for subsequent survey rounds, (ii) dummy variable for participation in the seed markets in the previous survey round ($\pi_{j,t-n}^{HS}$), (iii) long-term average rainfall (R_v), (iv) lagged rainfall shocks(1-year lag drought and flood shock) ($\mathcal{R}_{t-1}$), (v) long-term average maximum temperature ($\mathbb{C}_v$),and other control variables (X_{jt}'). The statistical significance of ρ in Eqn2 assess whether there is state dependency in the seed market. The initial hypothesis is that there is no state dependency in the seed market (i.e., $\rho = 0$). Unobservable household heterogeneity is identified by (b_j) and is assumed to be additive in the standard normal cumulative distribution function (Φ) and is modeled on the initial conditions of the dependent variable (π_{j0}^{HS}) and the list of covariates(X_j) (Wooldridge 2005) as follows:

$$b_j = \delta_0 + \delta_1 \pi_{j0}^{HS} + \delta_2 X_j + \delta_j. \quad \text{[Eqn3]}$$

Where $\delta_j \sim Normal\left(0, \sigma_\delta^2\right)$ and is independent of $\left(\pi_{j0}^{HS} + X_j\right)$. δ_0 is a constant. The vector of control variables (X_{jt}') include the household wealth endowments (farm size(ha), household labor units(elaborated below), and asset wealth index (elaborated below)), household age dependency ratio, farm population pressure(consumer units/farm size), characteristics of the household head (age(years), education (at least secondary education(1=yes)), sex (1=female; 0 otherwise), marital status (1= single; 0 otherwise), and community characteristics (elaborated below). We define household male adult equivalent labor units where we assign 1, 0.8, and 0.5 to an adult male, adult female, and children between 5 and 15 years of age, respectively. Household members available within the household for at least a month within the panel year are counted as members in the LSMS-ISA data and are hence considered in computing labor units. We combine information on household ownership of durable non-land assets (e.g., agricultural equipment and machinery) and household dwelling characteristics common in each country to create the household asset wealth index, using Principal Components Analysis(PCA) (Filmer and Pritchett 2001). Community characteristics include market access index and community population pressure (elaborated below). We proxy market access within the community using a market access index generated using principal components analysis (PCA). We construct this index using captured proxies for market access within the two studied countries. In Malawi, the component variables used in constructing the index include: (a)community has a daily or weekly market, (b) community is near an urban center, (c) community has a permanent ADMARC center, (d) community has a farmer cooperative, (e) and community has a warehouse for storing produce before selling. In Ethiopia, the component variables used include: (a) community has a weekly agricultural market, (b) community is near an urban center, (c) community has private input dealers as sources of seed and other inputs (in addition to the government sources), and (d) community has farmer cooperatives working in the seed sector. The slight discrepancy in the component variables used in making the market access index is due to data availability. Community population pressure is measured by a ratio of the total number of people within the community (Pn) to the number of households within the community $(Ch)(Pn/Ch)$. In addition, we also include regional and survey year dummies to control for the variation in access to purchased seeds across space and time.

In addition to the dynamic probit model for seed purchase decisions, we also specify dynamic Tobit models to study the intensity of participation. The dynamic Tobit model controls for unobserved heterogeneity as with the dynamic probit model specified earlier, except that it uses the intensity of participation as a dependent variable, and it accounts for censoring in the intensity of participation decisions. Following Wooldridge (2005) and Wooldridge (2010), we specify the dynamic Tobit model for intensity of seed purchase as in Eqn4:

$$\bar{P}_{jt}^{HS} = max\left[0. X_{jt}'\theta, R_v\vartheta, \mathcal{R}_{t-1}\varphi, \mathbb{C}_v\omega, k\left(\bar{P}_{j,t-n}^{HS}, \pi_{j,t-n}^{HS}, \pi_{j0}^{HS}, \bar{P}_{j0}^{HS}\right)\rho + b_j + \varepsilon_{jt}\right]. \qquad \text{[Eqn4]}$$

For all ($t = 1, \dots \dots . . T$, and $j = 1, 2, 3, \dots . . , N$ households). The specification communicates that in the dynamic Tobit model intensity of seed purchase in kilograms/ha ($\bar{P}_{jt}^{HS}$) is regressed on the previous survey round seed purchase intensity ($\bar{P}_{j,t-n}^{HS}$), a dummy variable for the previous survey round seed purchase ($\pi_{j,t-n}^{HS}$), initial survey round seed purchase dummy(π_{j0}^{HS}), and intensity ($\bar{P}_{j0}^{HS}$), and other covariates($R_v, \mathcal{R}_{t-1}, \mathbb{C}_v, X_{jt}'$) as described prior. Where the idiosyncratic error term $\varepsilon_{jt} \sim (0, \sigma_\varepsilon^2)$, and is independent of ($R_v, \mathcal{R}_{t-1}, \mathbb{C}_v, X_{jt}', \bar{P}_{j,t-n}^{HS}, \pi_{j,t-n}^{HS}, \pi_{j0}^{HS}, \bar{P}_{j0}^{HS}$). The functional expression $k(.)$ allows the influence of the lagged seed purchase variables to be different depending on whether the previous response was a corner solution or not and the intensity of seed purchased in the previous survey round. In the dynamic Tobit model, the unobservable household effect is modeled on initial participation in the seed market, including the intensity of seed purchased and other covariates. We model these dynamic seed purchase decisions using balanced panel data described prior.

As a robustness check to our main results, we present results in the supplementary material (Tables F and G), where we test and control for potential attrition bias due to dropping out households in subsequent rounds. To handle the possible attribution bias effect, we follow the following steps: First, we estimate probit attrition models for respective countries with dummy variable(1=yes) for households not observed in the follow-up surveys (2013 for Malawi, and 2014 for Ethiopia), and zero otherwise, using household characteristics at baseline as explanatory variables. We present results from the attrition probit models as part of the supplementary material (Table E). From the attrition probit results, we see that some household characteristics were significant in explaining the probability of dropping out, indicating that attrition was not random, which could lead to bias. Second, we construct an Inverse mills ratio (IMR) from the attrition probit models. The IMR we construct becomes a time-invariant variable in our balanced panel data set as households retains the same value of IMR across panel rounds. Third, we use the constructed IMR to test and control for the potential attrition bias effect by including it as an additional explanatory variable in our dynamic random effect probit and Tobit models. The IMR was not significant in any of the models, which suggests that attrition does not significantly alter our results and conclusions. We present results from this exercise (where we test and control for potential attrition bias) in Tables E and F in the supplementary material.

4. Results

4.1. Descriptive statistics

The means and standard errors for variables used in the analysis are presented in Table1. We only interpret our main outcome variables for brevity, which we also present graphically in Table 1. First, we describe seed purchasing trends in general (All crops) and then describe trends for Maize seed purchase. Seed purchasing in Malawi has increased from 43% in 2010 to 53% in 2019. On the contrary, seed purchasing slightly decreased by about 5% in Ethiopia, from 54% in 2012 to 49% in 2016 (Table 1) and **Figure 2**.

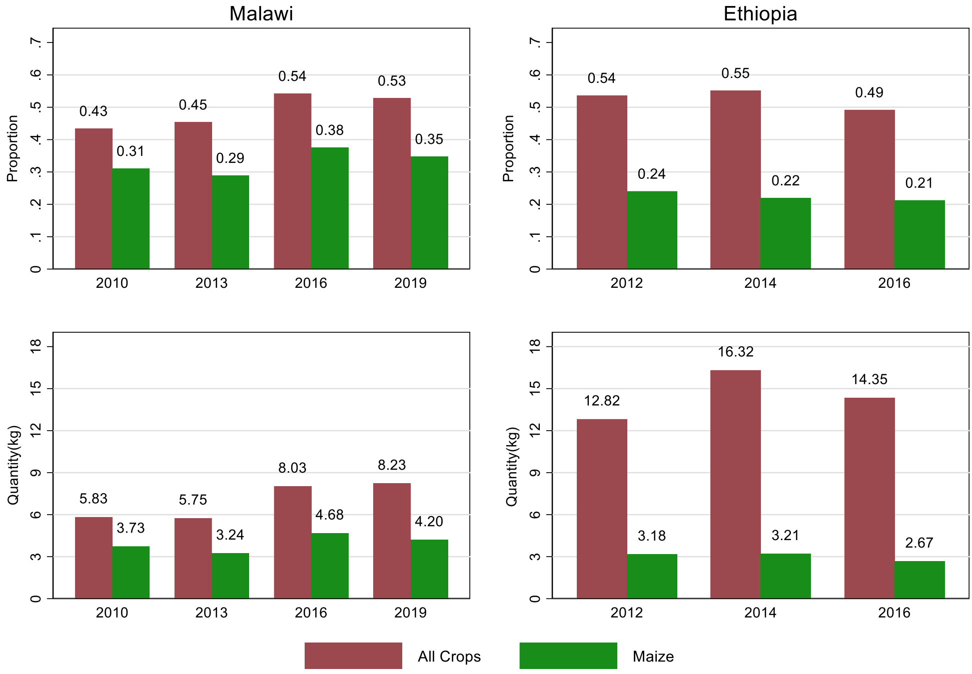

Figure 2: *Seed purchasing trends in the studied countries: The top panel figures show participation in the seed market (seed purchase) in general (for all crops grown) and specifically for maize. The bottom panel shows the intensity of participation in kgs/household purchased for the full sample of participants + potential participants*.

Likewise, the average quantities of seed purchased in Malawi have increased over time from about 5.8kgs in 2010 to 8.2ks in 2019. In the Ethiopian sample, on average, households purchased about 13kgs of seed in 2012, which rose by about 4kgs in 2014 before slightly reducing by 2kgs in 2016.

If we specifically focus on maize seed purchase decisions, we see that in Malawi, the proportion of farmers purchasing maize seeds lingers around 30% across survey rounds, with the lowest proportion being 29% in 2013 and the highest being 38% in 2016. In Ethiopia, the proportion of farmers purchasing maize seeds is slightly lower than those in Malawi, and they linger around 20%, with the highest proportion recorded in 2012(24%) and the lowest in 2016 (21%). Likewise, the average quantities of maize seed purchased in Malawi range between 3 and 5kgs across survey rounds (Table 1). In Ethiopia, the average quantity of maize seed purchased across rounds is about 3kgs per household.

The rest of the variables we use in the analysis, including lagged participation and intensity of participation variables, climate variables, household socioeconomics characteristics, characteristics of the household head, and community characteristics, are shown in Table 1. In addition to the descriptive statistics for control variables shown in Table 1, we also show how seed purchasers compare with non-purchasers in terms of their socioeconomic characteristics in the supplementary material (Table A-B).

Table 1: Descriptive statistics of variables used in the analysis

	Malawi								Ethiopia					
	2010		*2013*		*2016*		*2019*		*2012*		*2014*		*2016*	
VARIABLES	mean	se	mean	se	mean	se	mean	se	mean	se	mean	se	mean	se
Outcome variables (seed purchasing)														
Purchased seed in given survey year (1=yes)	0.43	0.016	0.45	0.016	0.54	0.016	0.53	0.016	0.54	0.013	0.55	0.013	0.49	0.013
Amount of seed purchased (kgs)	5.83	0.394	5.75	0.349	8.03	0.418	8.23	0.470	12.82	0.852	16.32	1.028	14.35	1.407
Purchased maize seed (1=yes)	0.31	0.015	0.29	0.015	0.38	0.016	0.35	0.015	0.24	0.012	0.22	0.013	0.21	0.012
Amount of maize seed purchased (kgs)	3.73	0.260	3.24	0.219	4.68	0.272	4.20	0.275	3.18	0.199	3.21	0.191	2.67	0.173
Lagged participation variables (All crops in general)														
Household purchased seed in the first panel round (1=yes)	0.43	0.016	0.43	0.016	0.43	0.016	0.43	0.43	0.54	0.013	0.54	0.013	0.54	0.013
Household purchased seed in the previous panel round (1=yes)	.	.	0.43	0.016	0.45	0.016	0.54	0.016	.	.	0.54	0.013	0.55	0.013
Amount of seed purchased seed in the first panel round (kg)	5.83	0.394	5.83	0.394	5.83	0.394	5.83	0.394	12.82	0.852	12.82	0.852	12.82	0.852
Amount of seed purchased seed in the previous panel round (kg)	.	.	5.83	0.394	5.75	0.349	8.03	0.418	.	.	12.82	0.852	16.32	1.028
Lagged participation variables (Maize)														
Household purchased maize seed in the first panel round (1=yes)	0.31	0.015	0.31	0.015	0.31	0.015	0.31	0.015	0.24	0.012	0.24	0.012	0.24	0.012
Household purchased maize seed in the previous panel round (1=yes)	.		0.31	0.015	0.29	0.015	0.38	0.016	.	.	0.24	0.012	0.22	0.013
Amount of maize seed purchased seed in the first panel round (kg)	3.73	0.260	3.73	0.260	3.73	0.260	3.73	0.260	3.18	0.199	3.18	0.199	3.18	0.199
Amount of maize seed purchased seed in the previous panel round (kg)	.	.	3.73	0.260	3.24	0.219	4.68	0.272	.	.	3.18	0.199	3.21	0.191
Climate variables														
Rainfall historical mean (1980-2018) in mm	902.55	1.926	902.55	1.926	902.55	1.926	902.55	1.926	770.78	3.438	770.78	3.438	770.78	3.438
1-year lag rainfall shock (Z-score)	0.75	0.010	-0.33	0.017	-0.32	0.007	-0.03	0.014	-0.08	0.009	0.35	0.011	0.35	0.010
Temperature historical mean (1980-2018) in degrees celsius	27.88	0.030	27.88	0.030	27.88	0.030	27.88	0.030	25.75	0.037	25.75	0.037	25.75	0.037
Other explanatory variables														
Farm size(ha)	0.73	0.016	0.74	0.017	0.78	0.018	0.77	0.017	1.20	0.042	1.38	0.078	1.36	0.044
Household labor units (adult equivalent)	2.88	0.042	3.24	0.044	3.18	0.083	3.30	0.045	3.88	0.031	3.60	0.031	3.91	0.035
Share of male labor (male labor units/total household labor units)	0.39	0.007	0.38	0.006	0.42	0.010	0.38	0.007	0.43	0.004	0.41	0.004	0.41	0.004
Household asset wealth index(normalized)	0.13	0.006	0.13	0.006	0.12	0.007	0.14	0.007	0.23	0.003	0.21	0.004	0.20	0.004
Farm population pressure ratio (Household size/farm size)	11.77	0.414	12.80	0.424	12.72	0.458	12.50	0.462	10.88	0.237	11.41	0.253	11.90	0.270
Community population pressure ratio (No. of people in community/No. of households in community)	5.48	0.116	5.48	0.116	8.03	0.312	7.34	0.180	5.43	0.036	5.36	0.037	5.49	0.048
Community market access index(normalized)	0.26	0.009	0.26	0.009	0.25	0.010	0.26	0.009	0.66	0.004	0.65	0.004	0.66	0.004
Female decision maker (1=yes)	0.23	0.014	0.26	0.014	0.29	0.015	0.29	0.015	0.18	0.008	0.19	0.008	0.20	0.008
Household head is single(1=yes)	0.23	0.013	0.23	0.014	0.33	0.015	0.31	0.015	0.15	0.007	0.18	0.008	0.20	0.008
Household dependency ratio ((number of dependents/ economically active members*100)	114.99	2.819	112.58	2.600	101.31	2.534	94.18	2.569	70.26	1.024	108.56	1.718	107.20	1.658
Age of household head(years)	43.55	0.508	46.06	0.500	48.05	0.450	50.35	0.462	44.54	0.303	46.17	0.296	48.12	0.299
Education level attained at least 12th grade (Ethiopia) , at least JCE(Malawi)	0.34	0.015	0.34	0.015	0.27	0.014	0.31	0.015	0.69	0.009	0.68	0.010	0.66	0.010
Observations	971		971		971		971		2398		2398		2398	

Notes: Descriptive statistics are not weighted; For all "normalized" variables we use unit-based normalization method as follows: Normalized V=(V-minimum(V))/(maximum(V)-minimum(V)); where (V) is the original variable (the index), and (Normalized V) is the normalized version of variable which is bounded between 0(minimum), and 1(maximum). First panel round for Malawi (Ethiopia) is 2010(2012).

4.2. Main results

Findings from the dynamic Probit and Tobit models for seed purchasing decisions in general and maize seed with alternative specifications are presented in Table 2 and Table 3, respectively. We report average partial effects (APE) to help interpret the economic and not just the statistical significance of variables. We report results from three alternative model specifications for our dynamic Probit and Tobit models as follows: (a) Use_0 (Intensity_0) are parsimonious dynamic Probit (Tobit) specifications where we include lagged participation variables only as explanatory variables, (b) Use_1(Intensity_1) are specifications where we add household endowments, historical mean rainfall, 1-year lag rainfall shock, historical mean temperature, and community market access index to the parsimonious specifications in (a). lastly, Use_2(Intensity_2) are specifications that include variables in (b) plus other household and community controls. We include regional and survey year dummies in all the three alternative specifications (a, b, and c). The full spectrum of variables used for each model specification is shown in the tables of results (Table 2 and Table 3). In addition, we provide results reporting corresponding coefficients for presented APEs in the supplementary material (Tables C and D). We present and interpret the results from our key variables, as clarified earlier.

4.2.1.Impact of lagged seed purchase variables on current seed purchase decisions

We start with results from the model of general input purchasing (All crops model) for both Malawi and Ethiopia, as presented in Table 2, and then move to the model for Maize seed purchase decisions. In Malawi, we learn that dummy variables for purchasing seed in the first and previous rounds significantly enhance the probability and intensity of purchasing seed in all our model specifications (Table 2). Precisely, households who purchased seed in the first round and previous round, respectively, had a 7-10% and 6 % higher probability of purchasing seed in the analyzed sample. More so, purchasing seed in the first and previous round in Malawi is associated with a marginal increase in quantities of seed purchased by 0.9-1.2 kgs and about 1 kg, respectively (Table 2). Similarly, in Ethiopia, we find that dummy variables for purchasing seed in the first round and previous survey round enhanced both the probability and intensity of seed purchasing in the studied sample. We, however, find previous seed purchasing decisions to have a somewhat greater impact in driving seed purchasing in Ethiopia compared to Malawi. For instance, the seed purchase dummy in the first survey round increased the probability of purchasing seed by between 16-and 22% in Ethiopia, which is more than double that we found in Malawi (7-10%). In addition, dummies for participation in the seed market (as buyers) in the first survey round and previous survey round are associated with marginal increases in the quantity of seeds purchased by between 4.5-5.9 and 3.2-3.8 kgs, respectively, in Ethiopia (Table 2). In addition, In Malawi, we see that a 1 kg of seed purchased in the first round is associated with a 0.04 kg increase in the quantity of seed purchased in the analyzed sample (Intensity_0). In Ethiopia, we also find that the initial year quantity of seed purchased is associated with a marginal increase in the average quantity purchased by 0.12 to 0.15 kgs (Table 2).

Table 2: Dynamic random effects Probit and Tobit models for seed purchase decisions in Malawi and Ethiopia. Reported are average partial effects

	Malawi ("All crops model")						*Ethiopia ("All crops model")*					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
VARIABLES	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2
Household purchased seed in the first panel round (1=yes)	0.1045***	0.0924***	0.0683***	1.0537**	1.2283***	0.8625**	0.2223***	0.1681***	0.1583***	5.9251***	4.8193***	4.4721***
	(0.02305)	(0.02286)	(0.02246)	(0.42721)	(0.41303)	(0.40623)	(0.02401)	(0.02369)	(0.02345)	(1.17442)	(1.14901)	(1.13255)
Household purchased seed in the previous panel round (1=yes)	0.0624**	0.0652**	0.0652**	0.9768**	1.0209**	0.9436**	0.0418	0.0510	0.0517*	3.7968***	3.2212***	3.1879***
	(0.02994)	(0.02989)	(0.02970)	(0.40427)	(0.40206)	(0.39950)	(0.03174)	(0.03128)	(0.03091)	(1.18472)	(1.17667)	(1.16985)
Amount of seed purchased in the first panel round (kg)				0.0397**	0.0159	0.0133				0.1487***	0.1247***	0.1157***
				(0.01680)	(0.01622)	(0.01589)				(0.01737)	(0.01712)	(0.01685)
Amount of seed purchased in the previous panel round (kg)				0.0032	0.0036	0.0032				0.0131	0.0204	0.0251
				(0.01727)	(0.01712)	(0.01701)				(0.01730)	(0.01733)	(0.01718)
Farm size(ha)		0.0724***	0.0746***		2.6176***	2.9130***		0.0059*	0.0038		0.3328***	0.2792***
		(0.01823)	(0.02355)		(0.30083)	(0.38528)		(0.00312)	(0.00274)		(0.10641)	(0.10626)
Household labor units		0.0093*	0.0086		0.1597*	0.1005		0.0125***	0.0083*		0.9992***	0.7808***
		(0.00512)	(0.00559)		(0.08208)	(0.08977)		(0.00432)	(0.00482)		(0.23856)	(0.26607)
Rainfall historical mean (1980-2018) in mm(log)		0.0284	0.0483		1.0238	1.6214		0.1603***	0.1560***		4.4868***	4.1005***
		(0.11879)	(0.11653)		(1.96515)	(1.94112)		(0.01957)	(0.01949)		(1.12549)	(1.11607)
1-year lag negative rainfall deviation		0.0927**	0.1027***		1.6823***	1.9189***		-0.1062**	-0.1317***		-2.8371	-4.1784*
		(0.04009)	(0.03981)		(0.63617)	(0.63497)		(0.04474)	(0.04468)		(2.49183)	(2.50199)
1-year lag positive rainfall deviation		0.0152	0.0112		1.5696	1.6918*		-0.0594***	-0.0598***		-1.3667	-1.2390
		(0.06092)	(0.06133)		(0.97686)	(0.98564)		(0.01505)	(0.01502)		(0.85564)	(0.85623)
Temperature historical mean (1980-2018) in degrees celsius		0.0294***	0.0312***		0.3927***	0.4244***		-0.0197***	-0.0200***		-1.0113***	-1.0539***
		(0.00701)	(0.00688)		(0.11649)	(0.11550)		(0.00272)	(0.00269)		(0.15560)	(0.15388)
Community market access index(normalized)		0.0095	0.0019		0.2012*	0.0604		0.0095*	0.0090*		0.5210*	0.4454
		(0.00645)	(0.00651)		(0.10580)	(0.10697)		(0.00505)	(0.00503)		(0.28370)	(0.28321)
Community population pressure ratio			0.0018			0.0063			0.0080**			0.5226***
			(0.00136)			(0.02126)			(0.00323)			(0.18166)
Farm population pressure			0.0005			0.0251*			-0.0001			-0.0203**
			(0.00086)			(0.01426)			(0.00004)			(0.00877)
Household asset wealth index(normalized)			0.2272***			4.1379***			0.0466***			2.7491***
			(0.04982)			(0.78431)			(0.00724)			(0.42122)
Female decision maker (1=yes)			-0.0602**			-0.6632			-0.0118			-2.4132*
			(0.02568)			(0.43264)			(0.02582)			(1.43071)
Household head is single(1=yes)			0.0129			-0.2879			0.0382			3.7722***
			(0.02604)			(0.42914)			(0.02645)			(1.46205)
Age of household head(years)			-0.0030***			-0.0472***			-0.0013**			-0.1051***
			(0.00070)			(0.01197)			(0.00054)			(0.03028)
Education level attained at least 12th grade (Ethiopia) , at least JCE(Malawi)			0.0066			0.1820			-0.0028			0.1751
			(0.02032)			(0.33709)			(0.01571)			(0.85712)
Household dependency ratio			0.0001			0.0022			0.0002**			0.0058
			(0.00012)			(0.00204)			(0.00009)			(0.00467)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses. Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

Table 3 reports APEs from dynamic Probit and Tobit random effects models for Maize seed purchase decisions in Malawi and Ethiopia. From the results, we learn that, like what we found with general seed purchase decisions, lagged maize seed purchase variables significantly enhance current maize seed purchasing in both Malawi and Ethiopia. In Malawi and Ethiopia, the dummy for maize seed purchase in the first survey round is associated with an 11-13% and 21-27% higher probability of purchasing maize seed in later rounds (Table 3). In addition, the dummy for purchasing maize seed in the first survey round is associated with a 1 to 1.3kgs and a 1.9 to 2.4 kgs increase in quantities of Maize seed purchased in later rounds in Malawi and Ethiopia respectively (Table 3). Also, a marginal 1kg increase in the quantity of maize seed purchased in the first survey round is associated with a 0.04 and between 0.07 to 0.09 kgs of maize seed purchased in the following survey rounds in Malawi and Ethiopia, respectively (Table 3).

Overall, we gather evidence that there is persistent state dependency on the demand side of the seed market in general and particularly in the maize seed market, causing selective access to purchased seed in Malawi and Ethiopia. Also, it is mainly initial participation and extent of participation that explain participation in subsequent years compared to participation in previous surveys. The state dependency is more pronounced in Ethiopia than in Malawi. Seed purchase from available markets gives an advantage to smallholder farmers with experience and established networks compared to new entrants. We, hence, could not reject our first hypothesis. We discuss this main result in the following sections.

4.2.2. Climate variability and seed purchase decisions

Here we report results on the link between historical climate (rainfall and temperature) and lagged rainfall shocks (1-year lag drought and flood shock) on seed purchase decisions in studied countries. We found long-term climate (rainfall and temperature) and the lagged rainfall shock (1-year lag) to explain seed purchase decisions in studied countries. In Ethiopia, we found a 1% increase in historical average rainfall associated with about a 0.002 unit increase in the probability of purchasing seed and a 4-5kg increase in the intensity of purchase (Table **2**). In Malawi, the link between historical rainfall and seed purchasing (in general) is also positive (Table **2**). For maize seed purchase decisions, a marginal (1%) increase in historical mean rainfall is associated with a 0.002-0.003 probability increase in maize seed purchase and about a 3kg increase in the intensity of purchase in Malawi (Table **3**). Similarly, in Ethiopia, a marginal (1%) increase in historical average rainfall is associated with a 0.001 units and 2kg increase in the probability and intensity of maize seed purchases, respectively (Table **3**).

Results linking lagged rainfall shocks to seed purchase decisions reveal that in Malawi, a marginal increase in the drought shock (absolute value of below average rainfall deviations) enhances the probability and intensity of seed purchases (in general) by about 9-10% and 2kgs, respectively (Table **2**). On the contrary, in Ethiopia, a marginal increase in the drought shock variable is associated with an 11-13% reduction in the probability of seed purchases (Table **2**). However, for maize seed purchase decisions, we found that a marginal increase in the 1-year lag drought shock variable in Ethiopia and Malawi marginally enhances maize seed purchases (Table **3**). In Malawi, the drought shock enhances the chances of maize seed purchases by 4%, while in Ethiopia, it enhances the probability of maize seed purchases by between 8-10%. The flood shock variable is associated with a reduction in seed purchase decisions in Ethiopia (Table **2** and Table **3**).

Our results also show that in Malawi, a marginal (1 degree) increase in historical average temperature is associated with a 3% and 0.4kgs increase in the probability and intensity of seed purchases in Malawi (Table **2**). In Ethiopia, general seed purchase decisions decline with increasing historical average temperature (Table **2**). For maize, we find the probability and intensity of maize seed purchases to increase with the historical average temperature in both countries. For instance, in Malawi (Ethiopia), the probability and intensity of maize seed purchases were found to increase by 2% (0.2%) and 0.2kgs(0.06kgs), respectively (Table **3**).

In addition to the results presented in the tables, we plot average partial effects showing the relationships between rainfall shock variables (positive and negative rainfall deviations and historical mean rainfall), and maize seed purchase decisions in general (in pooled samples), and in sub-groups of farmers: (a) between initial market participants (with knowledge and experience from past engagements) vs. non-initial participants (without experience), (b) farmers with relatively small vs. larger farm sizes based on total farm size, (c) Rich vs. poor farmers based on their non-land asset wealth endowments. The results are summarized in the supplementary material (Figures A-D). Insights from the plots (i.e., on the effects of negative rainfall deviations on Maize seed purchase decisions) reveal that smallholder farmers with knowledge and experience of the maize seed markets (gained from previous market engagements) and those with high land and non-land asset wealth endowments in both Malawi and Ethiopia have an elevated advantage in purchasing maize seeds post-drought shock exposure as adaptation compared to their poorer counterparts (see supplementary material Figures A-D attached with submission).

Overall, (i) Historical mean rainfall enhances seed purchase decisions in both Malawi and Ethiopia, (ii) rainfall variability (drought or flood shocks) generally enhances (discourages) general seed purchase decisions in Malawi (Ethiopia), (iii) lagged drought shocks enhances maize seed purchases in both countries, and (iii) historical mean temperature enhances maize seed purchase decisions in both Malawi and Ethiopia. Also, wealthier farmers and those with experience gained from past market engagements are more likely to purchase maize seeds post negative rainfall deviation exposure than their counterparts. Therefore, the hypothesis that lagged rainfall shocks do not encourage seed purchasing was rejected for Maize seed purchase decisions in Malawi and Ethiopia. Also, we could not reject our hypothesis that historical mean rainfall enhances seed purchase in subsequent seasons in Malawi and Ethiopia. Also, maize seed purchase decisions increase with increasing historical mean temperature. We discuss some of these key results in the next sections(discussion).

Table 3: Dynamic random effects Probit and Tobit models for Maize seed purchase decisions in Malawi and Ethiopia. Reported are average partial effects

	Malawi ("Maize model")						***Ethiopia ("Maize model")***					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
VARIABLES	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2
Household purchased maize seed in the first panel round (1=yes)	0.1310***	0.1277***	0.1065***	1.2009***	1.2495***	0.9798***	0.2664***	0.2296***	0.2199***	2.4207***	2.1470***	1.9841***
	(0.02278)	(0.02283)	(0.02264)	(0.37884)	(0.37519)	(0.36961)	(0.02024)	(0.02177)	(0.02240)	(0.26911)	(0.26194)	(0.26226)
Household purchased maize seed in the previous panel round (1=yes)	0.0247	0.0244	0.0216	0.0213	0.1364	0.0915	0.0204	0.0357	0.0439	0.2225	0.2353	0.3936
	(0.02806)	(0.02802)	(0.02778)	(0.36500)	(0.36699)	(0.36346)	(0.02430)	(0.02607)	(0.02685)	(0.23493)	(0.23691)	(0.25062)
Amount of maize seed purchased in the first panel round (kg)				0.0398*	0.0322	0.0297				0.0929***	0.0858***	0.0752***
				(0.02074)	(0.02044)	(0.02006)				(0.01428)	(0.01416)	(0.01427)
Amount of maize seed purchased in the previous panel round (kg)				0.0336	0.0282	0.0237				0.0016	0.0044	0.0118
				(0.02070)	(0.02050)	(0.02025)				(0.01278)	(0.01306)	(0.01389)
Farm size(ha)		0.0206	0.0410*		0.9965***	1.3223***		0.0019	0.0013		0.0277*	0.0237
		(0.01749)	(0.02245)		(0.22299)	(0.28537)		(0.00179)	(0.00170)		(0.01551)	(0.01561)
Household labor units		0.0060	0.0042		0.0833	0.0331		0.0080**	0.0044		0.1130**	0.0537
		(0.00480)	(0.00526)		(0.06089)	(0.06670)		(0.00350)	(0.00389)		(0.04398)	(0.04751)
Rainfall historical mean (1980-2018) in mm(log)		0.2268**	0.2470**		2.7227*	3.1735**		0.1466***	0.1446***		1.7029***	1.6674***
		(0.11545)	(0.11399)		(1.47900)	(1.46413)		(0.01702)	(0.01695)		(0.22169)	(0.21608)
1-year lag negative rainfall deviation		0.0359*	0.0396**		0.2719	0.3100		0.0952***	0.0790**		0.7592*	0.5713
		(0.01859)	(0.01877)		(0.23693)	(0.23861)		(0.03415)	(0.03460)		(0.41697)	(0.42178)
1-year lag positive rainfall deviation		0.0529	0.0557		0.7938	0.7484		-0.0571***	-0.0571***		-0.7269***	-0.7137***
		(0.05177)	(0.05208)		(0.66662)	(0.66894)		(0.01347)	(0.01350)		(0.16170)	(0.16266)
Temperature historical mean (1980-2018) in degrees celsius		0.0171**	0.0171**		0.1742**	0.1941**		0.0021	0.0023		0.0592*	0.0596**
		(0.00685)	(0.00678)		(0.08767)	(0.08711)		(0.00232)	(0.00230)		(0.03043)	(0.02953)
Community market access index(normalized)		0.0037	-0.0022		0.0713	-0.0296		0.0063	0.0059		0.0382	0.0270
		(0.00617)	(0.00625)		(0.07892)	(0.07983)		(0.00401)	(0.00401)		(0.04934)	(0.04906)
Community population pressure ratio			-0.0011			-0.0086			0.0050*			0.0494
			(0.00128)			(0.01604)			(0.00272)			(0.03443)
farm population pressure			0.0014*			0.0232**			-0.0002			-0.0039**
			(0.00081)			(0.01039)			(0.00013)			(0.00183)
Household asset wealth index(normalized)			0.1793***			2.9343***			0.0276***			0.3903***
			(0.04651)			(0.58260)			(0.00607)			(0.07746)
Female decision maker (1=yes)			-0.0112			-0.1456			-0.0121			-0.1654
			(0.02520)			(0.32301)			(0.02146)			(0.26545)
Household head is single(1=yes)			-0.0105			-0.2808			0.0407*			0.3704
			(0.02495)			(0.31963)			(0.02157)			(0.26408)
Age of household head(years)			-0.0027***			-0.0292***			-0.0001			-0.0003
			(0.00070)			(0.00918)			(0.00045)			(0.00552)
Education level attained at least 12th grade (Ethiopia) , at least JCE(Malawi)			0.0071			0.2233			0.0072			-0.0524
			(0.01974)			(0.25182)			(0.01243)			(0.15232)
Household dependency ratio			0.0002			0.0019			0.0001			0.0011
			(0.00012)			(0.00152)			(0.00007)			(0.00082)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses; Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

4.2.3. Other correlates of seed purchase decisions

In addition to our results on key variables of interest, we also found market access and household endowments significantly explain household general seed purchasing decisions and maize seed purchase decisions in Malawi and Ethiopia (Table 2 and Table 3).In Malawi and Ethiopia, households with better market access are more likely to purchase seeds. A marginal increase in the community market access index is associated with a 0.20 kgs increase in the quantity of seeds purchased in Malawi (Table **2**). Likewise, a marginal increase in the market access index is associated with a 1% higher probability of purchasing seed and a 0.5kg higher quantity of seed purchased in Ethiopia (Table **2**). Overall, we establish evidence that better market access correlates to higher access to seeds through purchase in Malawi and Ethiopia. We hence could not reject our hypothesis, which states that seed purchasing increases with improved market access.

We also attempt to report associations between household resource endowments and seed purchase decisions. We found that a 1 ha increase in farm size is associated with a 7% increase in the probability of purchasing seed and about a 3kg increase in purchase intensity in Malawi (Table 2). Similarly, in Ethiopia, a 1 ha increase in farm size is associated with a 0.6% increase in the probability of purchasing seed and a 0.3 kgs increase in purchase intensity. Household labor units are also positively associated with seed purchasing decisions. A unit increase in labor units is associated with a 1% increase in the probability of purchasing seeds in Both Malawi and Ethiopia. Also, a unit increase in household labor units is associated with a 0.2 and about 1 kg increase in the intensity of seed purchase in Malawi and Ethiopia, respectively (Table 2). For maize seed purchase decisions, we also found farm size to positively explain the probability and intensity of maize purchase in both Malawi and Ethiopia and that household labor units enhance both probability and intensity of maize seed purchase in Ethiopia (Table **3**). The household asset wealth index also positively correlates with seed purchase decisions in general (Table **2**) and maize seed purchase decisions (Table **3**) in Malawi and Ethiopia. Overall, and as expected, households better endowed with land, labor, and household assets are more likely to purchase seed from available markets than their poorer counterparts.

5. Discussion

Our findings show that there is state dependency on the demand side of the seed market in general and for maize seed, causing selective access in both Malawi and Ethiopia. The implication of the result is that ceteris paribus, smallholder farmers with experience and established networks have an advantage in the seed market compared to new entrants. This finding is in line with previous studies and the theory of non-convex transactions costs in factors markets that alludes to the fact that where markets are imperfect, participants are likely to face pervasive and dynamically variable non-linear transactions, which lead to selective access (Holden et al. 2007; Holden et al. 2010; Gebru et al. 2019; Tione and Holden 2021). Hence, entry into the seed market and establishing information networks is a sunk cost that potential market participants must overcome and later use to make future transactions (Fafchamps 2004; Barrett 2008).

However, the most significant challenge for potential market participants is getting over the first hurdle of entering the market. Upon entry, participation (and extent of participation) in subsequent years is a factor of initial market investments that marginally reduce overtime across space (Holden et al. 2007; Gebru et al. 2019; Tione and Holden 2021). This notion possibly explains why we found initial participation variables to explain participation in subsequent years more compared to participation in previous surveys. This notion further confirms the importance of building useful networks and gathering experience once households enter the seed market for the first time, reducing constraints in seed access through purchase in subsequent years. This notion is also in line with literature that alludes that smallholder farming households engage in collective action and invest in building their social networks to overcome or reduce widespread and pervasive transaction costs in accessing input markets in SSA (Key et al. 2000; Fafchamps 2004; Barrett 2008).

The observation of relatively more potent state dependency effects that lead to selective access to seed through purchase in Ethiopia than Malawi may be partly explained by key differences that characterize seed systems in the two countries. Such factors may include different policies governing access to seed through the formal systems and differences in the development of formal seed systems. The use of the formal seed system in Malawi, particularly for maize seed, has developed much faster in the recent past than in Ethiopia (Jayne and Rashid 2013; Kassie et al. 2013; Sheahan and Barrett 2017), which could explain the slight contrast in our findings. For example, government input support programs such as the Farm Input Support Program (FISP) in Malawi have had a greater impact on improving input market development and improved availability, awareness, and access to seed through the formal system in Malawi compared to other countries such as Ethiopia (Jayne and Rashid 2013; Sheahan and Barrett 2017) which could explain the difference. Also, the political economy of seed in Ethiopia and Malawi differ particularly in the roles played by public and private entities in the formal system (Langyintuo et al. 2010; Erenstein and Kassie 2018; Westengen et al. 2019). For instance, in Malawi, formal seed systems are dominated by both public and private players in the seed value chain (Kassie et al. 2013), while in Ethiopia, only the public entities dominate much of the functions of the system with little room for the private sector in practice a phenomenon that has been to linked to higher transaction costs in seed value chains and low access to improved seeds (Husmann 2015; Mekonnen et al. 2021). Given that access to seed through purchase is a crucial factor behind household seed security (Sperling 2002; Nordhagen and Pascual 2013; Sperling 2020), the existence of non-linear transaction costs which constrain access to seed in Malawi and Ethiopia contribute to household seed insecurity among other factors.

Overall, the state-dependency we found in seed purchase decisions shows the importance of accumulating market experience, established information networks, and market linkages in facilitating access to purchased seeds over time. However, some other factors, for instance, demand-side ignorance and stubborn preferences for sourcing seeds, could also contribute to state dependency.

We also gather evidence that seed purchase decisions respond to rainfall variability (lagged rainfall shocks) in both countries. Rainfall variability (drought or flood shocks) generally enhances (discourages) seed purchase decisions (averaged for all crops) in Malawi (Ethiopia), but lagged drought shocks (below average rainfall deviations) encourage maize seed purchasing decisions in both Malawi and Ethiopia. For maize seed purchase decisions, the effects of lagged rainfall shocks (e.g., negative rainfall deviation) are greater for farmers with market experience and those with relatively high wealth endowments compared to their counterparts. Also, historical average rainfall and temperature enhance maize seed purchase decisions in both countries. Recurrent rainfall variability, which may lead to crop failure, may disrupt farmer stocks of their own saved seed or make it hard to set aside seed from harvest due to urgent consumption needs, driving farmers to source seed from elsewhere through trade (Bellon et al. 2011; Nordhagen and Pascual 2013). Also, recurrent exposure to rainfall variability may induce learning on the benefits of different seed options, promoting farmers to choose seed options such as purchases that help them deal with future shocks (Holden and Quiggin 2017). This view possibly explains why we found the 1-year lag of drought shocks to enhance general seed purchasing decisions in Malawi and maize seed purchases in Malawi and Ethiopia. The Maize crop is highly sensitive to rainfall shocks, particularly drought shocks (Katengeza et al. 2019; McCann 2009), possibly explaining why maize seed purchase decisions respond to drought shocks in both countries.

However, it is also possible that rainfall shocks, through their effects on household economies, lead to less resource allocation for purchasing seeds which could partly explain the findings in Ethiopia, where lagged rainfall variability (drought and flood shocks) reduces general seed purchase decisions in general (all crops model). This idea is partly because exposure to shocks may increase hunger and poverty and discourage the adoption of beneficial technologies (Dercon 2005; Yesuf and Bluffstone 2009). Also, farmer perceptions of the benefits of different seed options (purchased seeds vs. farmers save seed) with increased rainfall uncertainty may explain the contrasting effect of rainfall variability. Also, the contrast in seed systems and the availability of subsidized seed inputs in the two countries could explain the disparity in the results. Access to the Farm Input Subsidy Program (FISP) in Malawi has improved awareness, availability, and access to improved seeds and has also enhanced input market development in Malawi (Jayne and Rashid 2013) which is not the case in Ethiopia. For example, a study in Malawi by Katengeza et al. (2019) found that exposure to lagged rainfall shocks (drought shocks) combined with the provision of subsidized seeds after shock exposure leads to higher uptake of improved (drought-tolerant) maize varieties. On the contrary, in Ethiopia, some studies (e.g., Alem et al. (2010)) have found lagged rainfall variability to discourage the use of productivity-improving risky inputs such as fertilizers. Hence, the key differences in seed systems features, farmer perceptions, differences in exposure and access to seeds off-farm, and the level of transaction costs in accessing seeds from available markets could explain the contrast in the effects of lagged rainfall variability (drought or flood shocks) on general seed purchase decisions (defined for al crops) in the studied countries.

The findings that higher historical average rainfall enhances seed purchasing in both Malawi and Ethiopia could reflect on the marginal benefits of rainfall amount received on crop harvest, which may ease liquidity constraints faced by households and hence their ability to source seeds through purchase. In rural contexts like Malawi and Ethiopia, where households have low-income levels and input, and credit markets are imperfect, abundant rainfall may be associated with increased crop harvest and household disposable incomes, which relax liquidity constraints in the adoption of agricultural technologies (Alem et al. 2010; Dercon and Christiaensen 2011; Falco et al. 2014). Overall, results from this paper portray that seed purchasing practices respond to measures of rainfall variability and experience and knowledge of the seed market, and high wealth endowments enhance their response to negative shocks (e.g., drought shocks).

We also gather evidence that other factors, including market access and household resource endowments, positively correlate with seed purchasing. Households well-endowed with resources such as labor, land, and other durable household assets and with better market access are well known to face lower constraints to technology adoption(Crawford et al. 2003; Croppenstedt et al. 2003; Barrett 2008; Winters et al. 2009; Jagwe et al. 2010), and this also stands true for accessing purchased seeds. This view is in line with theory and evidence that states that the choice of technologies by smallholder farmers is a function of many factors, including resource endowments(land, labor, assets), markets, institutions, and infrastructure that facilitates access and use of these resource endowments and markets (Crawford et al. 2003; Croppenstedt et al. 2003; Winters et al. 2009).

6. Conclusions and policy implications

We study the evolution of the seed market in Malawi and Ethiopia, focusing specifically on smallholder seed purchasing decisions over time. By investigating the influence of the household's previous participation decisions in the seed market (through purchase) on later participation decisions, we gather evidence that pervasive non-linear transaction costs characterize access to off-farm seed through the market. These non-linear transaction costs emanate from policies, institutions, and social factors that determine the degree of information asymmetries in farmers' access and use of seed and productive resources. As a result, these transaction costs constrain access to seeds through the markets, reducing household seed security over time. However, the problem is likely to reduce over time for households that can break the first hurdle of entering the market because of established social networks, experience, and market linkages that may marginally reduce transaction costs and improve subsequent access to purchased seeds.

Seed purchasing also responds to historical mean rainfall, lagged rainfall shocks, market access, and household resource endowments. The decisions to purchase and intensity of purchase seed are positively affected by historical mean rainfall in both Malawi and Ethiopia. Also, lagged drought shocks enhance general seed purchasing decisions in Malawi and maize seed purchases in both Malawi and Ethiopia. Maize farmers in Malawi and Ethiopia with market experience (gained from previous market engagements) and high asset wealth endowments have an elevated advantage in purchasing maize seeds post-drought shock exposure as an adaptation strategy compared to their opposite counterparts. High rainfall levels(historically) may increase harvest and household disposable income, thereby reducing liquidity constraints faced by households allowing them to access seed off-farm through purchase. Also, lagged rainfall shocks may increase liquidity constraints and the risk and uncertainty in the production environment, influencing seed purchasing by smallholder farmers. Improved market access and high asset endowments enhance seed purchasing in smallholder farming.

Given the importance of seed purchasing in enhancing seed diversity and improving household seed security, policy efforts may target reducing transaction costs and other entry barriers into formal and informal seed markets to improve access to seeds. For instance, continual development and upgrading of road infrastructure and agricultural support services such as rural financing and extension are some worthwhile interventions. All such efforts that reduce transaction costs will increase the effective demand for purchased seed and enhance farmers' seed security over time. Improved road infrastructure and access to market information would also facilitate output market participation and hence provide farmers with income from selling surplus produce. In addition, Investments in rural financing and insurance will ease constraints that farmers face when they try to access seeds off-farm through purchasing to enhance the adaptation of their cropping activities to recurrent rainfall shocks.

Acknowledgments

This research work is an output of a research project, "*Access to seeds: from Emergencies to Seed System development (ACCESS),"* which is fully supported by the Research Council of Norway, project number (RCN 288493), and the Norwegian University of Life Sciences (NMBU). The authors are also grateful to the World Bank (and partners) for making the Living Standards Measurement Study - Integrated Surveys on Agriculture (*LSMS-ISA*) data for Malawi and Ethiopia available for this research.

References

Alem Y, Bezabih M, Kassie M, Zikhali P (2010) Does fertilizer use respond to rainfall variability? Panel data evidence from Ethiopia Agricultural economics 41:165-175
Almekinders CJ, Louwaars NP, De Bruijn G (1994) Local seed systems and their importance for an improved seed supply in developing countries Euphytica 78:207-216
Almekinders CJ, Thiele G, Danial DL (2007) Can cultivars from participatory plant breeding improve seed provision to small-scale farmers? Euphytica 153:363-372
Barrett CB (2008) Smallholder market participation: Concepts and evidence from eastern and southern Africa Food policy 33:299-317
Bellon MR, Hodson D, Hellin J (2011) Assessing the vulnerability of traditional maize seed systems in Mexico to climate change Proceedings of the National Academy of Sciences 108:13432-13437
Bellon MR, Adato M, Becerril J, Mindek D (2006) Poor farmers' perceived benefits from different types of maize germplasm: the case of creolization in lowland tropical Mexico World development 34:113-129
Binswanger HP, Rosenzweig MR (1986) Behavioural and material determinants of production relations in agriculture The Journal of Development Studies 22:503-539
Coase RH (1937) The nature of the firm economica 4:386-405
Coomes OT et al. (2015) Farmer seed networks make a limited contribution to agriculture? Four common misconceptions Food Policy 56:41-50
Crawford E, Kelly V, Jayne TS, Howard J (2003) Input use and market development in Sub-Saharan Africa: an overview Food Policy 28:277-292
Croppenstedt A, Demeke M, Meschi MM (2003) Technology adoption in the presence of constraints: the case of fertilizer demand in Ethiopia Review of Development Economics 7:58-70
De Janvry A, Fafchamps M, Sadoulet E (1991) Peasant household behaviour with missing markets: some paradoxes explained The Economic Journal 101:1400-1417
Dercon S (2005) Risk, poverty and vulnerability in Africa Journal of African Economies 14:483-488
Dercon S, Christiaensen L (2011) Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia Journal of Development Economics 96:159-173 doi:https://doi.org/10.1016/j.jdeveco.2010.08.003
Dorward A, Kydd J, Morrison J, Poulton C (2005) Institutions, markets and economic co-ordination: Linking development policy to theory and praxis Development and change 36:1-25
Erenstein O, Kassie GT (2018) Seeding eastern Africa's maize revolution in the post-structural adjustment era: a review and comparative analysis of the formal maize seed sector International Food and Agribusiness Management Review 21:39-52
Fackler PL, Goodwin BK (2001) Spatial price analysis Handbook of agricultural economics 1:971-1024
Fafchamps M (2004) Market institutions in sub-Saharan Africa. Cambridge, MA: MIT Press,

Falco SD, Adinolfi F, Bozzola M, Capitanio F (2014) Crop insurance as a strategy for adapting to climate change Journal of Agricultural Economics 65:485-504
FAO Developing seed security strategies and programmes for food security in developing countries. In, Rome, Italy, 1998. Proceedings of the international workshop on Seed Security for Food security, 30 November to 1 December 1997. FAO,
FAO, ECHA (2015) Building Capacity for Seed Security Assessments: Household Seed Security Concepts and Indicators, Discussion Paper. FAO, and European Commission Humanitarian Aid. .
Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas International journal of climatology 37:4302-4315
Filmer D, Pritchett LH (2001) Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India Demography 38:115-132
Gebru M, Holden ST, Tilahun M (2019) Tenants' land access in the rental market: evidence from northern Ethiopia Agricultural Economics 50:291-302
Holden ST, Quiggin J (2017) Climate risk and state-contingent technology adoption: shocks, drought tolerance and preferences European Review of Agricultural Economics 44:285-308
Holden ST, Deininger K, Ghebru HH (2007) Impact of land certification on land rental market participation in Tigray region, Northern Ethiopia Northern Ethiopia (October 1, 2007)
Holden ST, Otsuka K, Place FM (2010) The emergence of land markets in Africa: Impacts on poverty, equity, and efficiency. Routledge,
Husmann C (2015) Transaction costs on the Ethiopian formal seed market and innovations for encouraging private sector investments Quarterly Journal of International Agriculture 54:59-76
Jagwe JN, Machethe CL, Ouma E (2010) Transaction costs and smallholder farmers' participation in banana markets in the Great Lakes Region of Burundi, Rwanda and the Democratic Republic of Congo
Jayne TS, Rashid S (2013) Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence Agricultural economics 44:547-562
Kassie G, MacRobert J, Setimela P, Erenstein O, Mwangi W, Shiferaw B (2013) Political and economic features of the maize seed industry in southern Africa Agrekon 52:104-127
Katengeza SP, Holden ST, Lunduka RW (2019) Adoption of drought tolerant maize varieties under rainfall stress in Malawi Journal of Agricultural Economics 70:198-214 doi: https://doi.org/10.1111/1477-9552.12283
Key N, Sadoulet E, Janvry AD (2000) Transactions costs and agricultural household supply response American journal of agricultural economics 82:245-259
Langyintuo AS, Mwangi W, Diallo AO, MacRobert J, Dixon J, Banziger M (2010) Challenges of the maize seed industry in eastern and southern Africa: A compelling case for private-public intervention to promote growth Food Policy 35:323-331 doi:10.1016/j.foodpol.2010.01.005
Markelova H, Meinzen-Dick R, Hellin J, Dohrn S (2009) Collective action for smallholder market access Food policy 34:1-7
Masarie KA, Tans PP (1995) Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record Journal of Geophysical Research: Atmospheres 100:11593-11610
McGuire S, Sperling L (2013) Making seed systems more resilient to stress Global Environmental Change 23:644-653
McGuire S, Sperling L (2016) Seed systems smallholder farmers use Food Security 8:179-195
Mekonnen DK, Abate GT, Yimam S, Benfica R, Spielman DJ, Place F (2021) The impact of Ethiopia's direct seed marketing approach on smallholders' access to seeds, productivity, and commercialization vol 1998. Intl Food Policy Res Inst,
Mulesa TH, Dalle SP, Makate C, Haug R, Westengen OT (2021) Pluralistic Seed System Development: A Path to Seed Security? Agronomy 11:372
Nordhagen S, Pascual U (2013) The impact of climate shocks on seed purchase decisions in Malawi: implications for climate change adaptation World Development 43:238-251

North (1987) Transaction Costs. In: John Eatwell, Murray Milgate, Newman. P (eds) The New Palgrave: A Dictionary of Economics, vol 4. Macmillan, London,
Renkow M, Hallstrom DG, Karanja DD (2004) Rural infrastructure, transactions costs and market participation in Kenya Journal of development economics 73:349-367
Ricker-Gilbert J, Chamberlin J (2018) Transaction costs, land rental markets, and their impact on youth access to agriculture in Tanzania Land Economics 94:541-555
Sheahan M, Barrett CB (2017) Ten striking facts about agricultural input use in Sub-Saharan Africa Food Policy 67:12-25
Shiferaw B, Obare G, Muricho G Rural market imperfections and the role of institutions in collective action to improve markets for the poor. In: Natural Resources Forum, 2008. vol 1. Wiley Online Library, pp 25-38
Sperling L (2002) Emergency seed aid in Kenya: some case study insights on lessons learned during the 1990s Disasters 26:329-342
Sperling L (2020) Seed security response during COVID-19: building on evidence and orienting to the future Food Security:1-5
Sperling L, Cooper HD, Remington T (2008) Moving towards more effective seed aid The Journal of Development Studies 44:586-612
Tione SE, Holden ST (2021) Non-convex transaction costs and land rental market participation in Malawi Land Economics:062620-060092R
Tripp R (2006) Strategies for seed system development in sub-Saharan Africa: A study of Kenya, Malawi, Zambia, and Zimbabwe. SAT eJournal 2:1-58
Westengen OT, Haug R, Guthiga P, Macharia EJFiSFS (2019) Governing seeds in East Africa in the face of climate change: Assessing political and social outcomes 3:53
Winters P et al. (2009) Assets, activities and rural income generation: Evidence from a multicountry analysis World Development 37:1435-1452
Wooldridge JM (2005) Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity Journal of applied econometrics 20:39-54
Wooldridge JM (2010) Econometric analysis of cross section and panel data
Yesuf M, Bluffstone RA (2009) Poverty, Risk Aversion, and Path Dependence in Low-Income Countries: Experimental Evidence from Ethiopia American Journal of Agricultural Economics 91:1022-1037 doi:10.1111/j.1467-8276.2009.01307.x %J American Journal of Agricultural Economics

Appendix: Smallholder access to purchased seeds in the presence of pervasive market imperfections and rainfall shocks: Panel Data Evidence from Malawi and Ethiopia

Clifton Makate* , Arild Angelsen, Stein Terje Holden, & Ola Tveitereid Westengen

A. Descriptive statistics: Comparison of seed purchasers and potential purchasers in their socioeconomic characteristics(t-test)

Table A: Comparison of market participants (P) with non-participants (non_P) (***across years***) in their socioeconomic characteristics and rainfall variability

VARIABLES	Malawi P	Malawi non_P	Malawi (p-value)	Ethiopia P	Ethiopia non_P	Ethiopia (p-value)
Rainfall historical mean (1980-2018) in mm	899.017	905.947	0.072*	835.201	702.884	0.000***
1-year lag rainfall shock (Z-score)	-0.000	0.034	0.068*	0.206	0.188	0.162
Temperature historical mean (1980-2018) in degrees Celsius	28.014	27.751	0.000***	24.951	26.663	0.000***
Farm size(ha)	0.788	0.721	0.000***	1.441	1.204	0.000***
Household labor units	3.286	3.018	0.000***	3.853	3.751	0.007***
Share of male labor (male labor units/total household labor units)	0.399	0.390	0.227	0.419	0.415	0.294
Household asset wealth index(normalized)	0.155	0.110	0.000***	0.223	0.206	0.000***
Community population pressure ratio (Number of people in community/number of households in community)	6.900	6.273	0.002***	5.537	5.297	0.000***
Community basic infrastructure index(normalized)	0.200	0.196	0.554	0.304	0.293	0.006***
Community market access index(normalized)	0.268	0.248	0.030**	0.682	0.631	0.000***
Female decision maker (1=yes)	0.234	0.300	0.000***	0.186	0.191	0.637
Household head is single(1=yes)	0.244	0.304	0.000***	0.166	0.181	0.080*
Household dependency ratio ((number of dependents/ economically active members*100)	107.917	103.697	0.111	95.186	95.399	0.905
Age of household head(years)	45.678	48.280	0.000***	45.789	46.751	0.006***
Education level attained at least 12th grade (Ethiopia), at least JCE(Malawi)	0.310	0.318	0.616	0.651	0.706	0.000***
Distance to market (ADMARC in Malawi) (nearest main markets in Ethiopia)	7.411	7.574	0.329	63.853	68.851	0.000***
Observations	1903	1981		3727	3467	

Notes: Statistics are not weighted: Number of panel households in Malawi and Ethiopia are 971 and 2398 respectively. The p-value is from a t-test mean comparison test; ***, **, and * communicates significance at 1, 5, and 10% respectively.

Table B: Comparison of market participants (P) with non-participants (non_P) ***at baseline*** in their socioeconomic characteristics and rainfall variability

VARIABLES	Malawi P	Malawi non_P	Malawi (p-value)	Ethiopia P	Ethiopia non_P	Ethiopia (p-value)
Rainfall historical mean (1980-2018) in mm	899.868	904.615	0.222	824.397	708.680	0.000***
1-year lag rainfall shock (Z-score)	0.023	0.012	0.566	0.206	0.208	0.850
Temperature historical mean (1980-2018) in degrees Celsius	27.985	27.799	0.002***	24.970	26.653	0.000***
Farm size(ha)	0.739	0.766	0.123	1.397	1.215	0.006***
Household labor units	3.263	3.062	0.000***	3.821	3.769	0.170
Share of male labor (male labor units/total household labor units)	0.400	0.391	0.222	0.421	0.412	0.042***
Household asset wealth index(normalized)	0.159	0.111	0.000***	0.224	0.205	0.000***
Community population pressure ratio (Number of people in community/number of households in community)	6.865	6.361	0.013**	5.537	5.295	0.000***
Community basic infrastructure index(normalized)	0.207	0.192	0.027**	0.301	0.293	0.041
Community market access index(normalized)	0.269	0.249	0.032**	0.680	0.628	0.000***
Female decision maker (1=yes)	0.243	0.286	0.002***	0.187	0.195	0.388
Household head is single(1=yes)	0.241	0.301	0.000***	0.163	0.188	0.004***
Household dependency ratio ((number of dependents/ economically active members*100)	109.213	103.114	0.022**	95.020	95.711	0.699
Age of household head(years)	45.225	48.374	0.000***	46.212	46.354	0.683
Education level attained at least 12th grade (Ethiopia), at least JCE(Malawi)	0.315	0.313	0.932	0.653	0.701	0.000***
Distance to market (ADMARC in Malawi) (nearest main markets in Ethiopia)	7.469	7.512	0.799	64.126	69.138	0.000***
Observations	1688	2196		3861	3333	

Notes: Statistics are not weighted: Number of panel households in Malawi and Ethiopia are 971 and 2398 respectively. The p-value is from a t-test mean comparison test; ***, **, and * communicates significance at 1, 5, and 10% respectively.

B. Full table of results showing coefficients

Table C: Dynamic random effects Probit and Tobit models for seed purchase (All crops) decisions in Malawi and Ethiopia-Coefficients

	Malawi ("All crops model")						*Ethiopia ("All crops model")*					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
	Use 0	Use 1	Use 2	Intensity 0	Intensity 1	Intensity 2	MUse 0	MUse 1	MUse 2	Intensity 0	Intensity 1	Intensity 2
Household purchased seed in the first panel round (1=yes)	0.2867***	0.2535***	0.1880***	2.9673**	3.4194***	2.3893**	0.7296***	0.5477***	0.5162***	19.5150***	15.8649***	14.6959***
	(0.06725)	(0.06608)	(0.06411)	(1.20251)	(1.14919)	(1.12510)	(0.10359)	(0.09377)	(0.09158)	(3.86280)	(3.78049)	(3.72134)
Household purchased seed in the previous panel round (1=yes)	0.1712**	0.1789**	0.1796**	2.7506**	2.8422**	2.6140**	0.1371	0.1662*	0.1687*	12.5053***	10.6039***	10.4760***
	(0.07917)	(0.07902)	(0.07893)	(1.13716)	(1.11801)	(1.10566)	(0.09960)	(0.09680)	(0.09571)	(3.89946)	(3.87072)	(3.84091)
Amount of seed purchased in the first panel round (kg)				0.1119**	0.0442	0.0368				0.4899***	0.4103***	0.3801***
				(0.04727)	(0.04516)	(0.04402)				(0.05700)	(0.05630)	(0.05534)
Amount of seed purchased in the previous panel round (kg)				0.0090	0.0101	0.0089				0.0433	0.0671	0.0824
				(0.04862)	(0.04766)	(0.04712)				(0.05697)	(0.05704)	(0.05641)
Farm size(ha)		0.1985***	0.2054***		7.2872***	8.0695***		0.0194*	0.0124		1.0956***	0.9174***
		(0.05075)	(0.06547)		(0.82774)	(1.05828)		(0.01024)	(0.00895)		(0.35011)	(0.34903)
Household labor units		0.0254*	0.0237		0.4447*	0.2783		0.0406***	0.0270*		3.2893***	2.5658***
		(0.01411)	(0.01544)		(0.22841)	(0.24865)		(0.01421)	(0.01580)		(0.78409)	(0.87371)
Rainfall historical mean (1980-2018) in mm		0.0778	0.1329		2.8501	4.4916		0.5223***	0.5087***		14.7703***	13.4749***
		(0.32595)	(0.32113)		(5.47077)	(5.37682)		(0.06988)	(0.06927)		(3.70693)	(3.67005)
1-year lag negative rainfall deviation		0.2543**	0.2828***		4.6835***	5.3157***		-0.3460**	-0.4294***		-9.3397	-13.7307*
		(0.10961)	(0.10937)		(1.76945)	(1.75707)		(0.14621)	(0.14623)		(8.20313)	(8.22175)
1-year lag positive rainfall deviation		0.0416	0.0309		4.3697	4.6864*		-0.1936***	-0.1951***		-4.4991	-4.0715
		(0.16710)	(0.16891)		(2.71842)	(2.72915)		(0.04941)	(0.04933)		(2.81753)	(2.81452)
Temperature historical mean (1980-2018) in degrees Celsius		0.0807***	0.0858***		1.0933***	1.1756***		-0.0642***	-0.0651***		-3.3292***	-3.4632***
		(0.01966)	(0.01944)		(0.32403)	(0.31964)		(0.00953)	(0.00946)		(0.51084)	(0.50429)
Community market access index(normalized)		0.0260	0.0053		0.5603*	0.1673		0.0311*	0.0292*		1.7150*	1.4635
		(0.01770)	(0.01794)		(0.29439)	(0.29629)		(0.01649)	(0.01648)		(0.93360)	(0.93046)
Community population pressure ratio			0.0049			0.0175			0.0260**			1.7173***
			(0.00376)			(0.05888)			(0.01058)			(0.59668)
Farm population pressure			0.0014			0.0694*			-0.0002			-0.0669**
			(0.00238)			(0.03949)			(0.00015)			(0.02881)
Household asset wealth index(normalized)			0.6257***			11.4627***			0.1520***			9.0338***
			(0.14028)			(2.16618)			(0.02435)			(1.38042)
Female decision maker (1=yes)			-0.1658**			-1.8371			-0.0386			-7.9303*
			(0.07113)			(1.19837)			(0.08417)			(4.70090)
Household head is single(1=yes)			0.0355			-0.7974			0.1245			12.3959***
			(0.07175)			(1.18866)			(0.08630)			(4.80239)
Age of household head(years)			-0.0082***			-0.1308***			-0.0041**			-0.3454***
			(0.00196)			(0.03312)			(0.00178)			(0.09942)
Education level attained at least 12th grade (Ethiopia) , at least JCE(Malawi)			0.0183			0.5041			-0.0092			0.5755
			(0.05597)			(0.93373)			(0.05122)			(2.81660)
Household dependency ratio			0.0003			0.0061			0.0006**			0.0190
			(0.00034)			(0.00566)			(0.00028)			(0.01535)
_cons	-0.2655***	-3.4184	-3.6556	-5.8609***	-65.1600*	-74.3160**	-0.8118***	-3.8846***	-4.7242***	-67.5250***	-151.8867***	-191.3923***
	(0.06110)	(2.26695)	(2.23726)	(1.10049)	(38.14837)	(37.58928)	(0.06789)	(0.59823)	(0.63162)	(3.94618)	(32.83518)	(34.28579)
lnsig2u/sigma_u	-1.9360***	-2.2523***	-2.6313***	7.8835***	6.5651***	5.9100***	-0.9348***	-1.3942***	-1.5387***	30.9427***	27.0850***	23.9161***
	(0.42925)	(0.55587)	(0.76632)	(0.97157)	(1.07070)	(1.12539)	(0.29947)	(0.40135)	(0.44298)	(4.05304)	(4.48361)	(4.91495)
sigma_e				19.0045***	18.7883***	18.6794***				64.0427***	64.7049***	65.0890***
				(0.48665)	(0.48309)	(0.47793)				(1.90603)	(1.91388)	(1.90542)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses; Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

Table D: Dynamic random effects Probit and Tobit models for Maize seed purchase decisions in Malawi and Ethiopia-Coefficients

	Malawi ("Maize model")						*Ethiopia ("Maize model")*					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
VARIABLES	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2	MUse_0	MUse_1	MUse_2	IntensityM_0	IntensityM_1	IntensityM_2
Household purchased maize seed in the first panel round (1=yes)	0.4115***	0.4024***	0.3377***	4.2482***	4.4037***	3.4417***	1.7222***	1.4203***	1.3287***	11.3744***	10.1378***	9.3945***
	(0.07792)	(0.07809)	(0.07647)	(1.33865)	(1.32063)	(1.29730)	(0.22786)	(0.21120)	(0.20588)	(1.24717)	(1.22391)	(1.23254)
Household purchased maize seed in the previous panel round (1=yes)	0.0776	0.0770	0.0684	0.0752	0.4806	0.3213	0.1321	0.2206	0.2652*	1.0456	1.1111	1.8637
	(0.08658)	(0.08671)	(0.08669)	(1.29119)	(1.29345)	(1.27672)	(0.15010)	(0.15001)	(0.14899)	(1.10534)	(1.11997)	(1.18831)
Amount of maize seed purchased in the first panel round (kg)				0.1408*	0.1134	0.1043				0.4366***	0.4053***	0.3561***
				(0.07329)	(0.07199)	(0.07044)				(0.06655)	(0.06644)	(0.06728)
Amount of maize seed purchased in the previous panel round (kg)				0.1187	0.0995	0.0832				0.0077	0.0210	0.0560
				(0.07320)	(0.07224)	(0.07112)				(0.06007)	(0.06167)	(0.06576)
Farm size(ha)		0.0649	0.1299*		3.5119***	4.6447***		0.0119	0.0081		0.1306*	0.1120
		(0.05523)	(0.07149)		(0.78244)	(0.99848)		(0.01113)	(0.01031)		(0.07323)	(0.07387)
Household labor units		0.0189	0.0132		0.2935	0.1163		0.0496**	0.0263		0.5334**	0.2541
		(0.01517)	(0.01670)		(0.21456)	(0.23430)		(0.02200)	(0.02367)		(0.20758)	(0.22493)
Rainfall historical mean (1980-2018) in mm		0.7146*	0.7829**		9.5956*	11.1469**		0.9071***	0.8736***		8.0412***	7.8952***
		(0.36574)	(0.36348)		(5.21139)	(5.14130)		(0.12406)	(0.12038)		(1.04671)	(1.02398)
1-year lag negative rainfall deviation		0.1132*	0.1256**		0.9581	1.0887		0.5888***	0.4777**		3.5849*	2.7049
		(0.05866)	(0.05961)		(0.83511)	(0.83821)		(0.21659)	(0.21398)		(1.96970)	(1.99756)
1-year lag positive rainfall deviation		0.1668	0.1764		2.7975	2.6288		-0.3534***	-0.3452***		-3.4325***	-3.3791***
		(0.16320)	(0.16511)		(2.34920)	(2.34957)		(0.08348)	(0.08180)		(0.76513)	(0.77164)
Temperature historical mean (1980-2018) in degrees celsius		0.0538**	0.0541**		0.6140**	0.6819**		0.0130	0.0139		0.2793*	0.2820**
		(0.02179)	(0.02171)		(0.30890)	(0.30587)		(0.01442)	(0.01396)		(0.14363)	(0.13977)
Community market access index(normalized)		0.0117	-0.0070		0.2514	-0.1038		0.0392	0.0356		0.1805	0.1278
		(0.01941)	(0.01981)		(0.27812)	(0.28041)		(0.02488)	(0.02439)		(0.23300)	(0.23230)
Community population pressure ratio			-0.0034			-0.0302			0.0302*			0.2337
			(0.00407)			(0.05633)			(0.01644)			(0.16304)
farm population pressure			0.0044*			0.0815**			-0.0009			-0.0187**
			(0.00259)			(0.03648)			(0.00081)			(0.00869)
Household asset wealth index(normalized)			0.5684***			10.3069***			0.1670***			1.8482***
			(0.15011)			(2.03887)			(0.03732)			(0.36714)
Female decision maker (1=yes)			-0.0354			-0.5115			-0.0731			-0.7832
			(0.07989)			(1.13457)			(0.12984)			(1.25698)
Household head is single(1=yes)			-0.0333			-0.9863			0.2462*			1.7538
			(0.07910)			(1.12259)			(0.13158)			(1.25074)
Age of household head(years)			-0.0084***			-0.1026***			-0.0008			-0.0015
			(0.00227)			(0.03225)			(0.00270)			(0.02612)
Education level attained at least 12th grade (Ethiopia), at least JCE(Malawi)			0.0226			0.7842			0.0432			-0.2483
			(0.06257)			(0.88434)			(0.07531)			(0.72126)
Household dependency ratio			0.0005			0.0068			0.0004			0.0052
			(0.00038)			(0.00533)			(0.00041)			(0.00386)
_cons	-0.6686***	-6.6471***	-6.8500***	-10.7860***	-90.5482**	-99.8609***	-2.1321***	-9.9182***	-10.8893***	-19.0810***	-92.5921***	-104.6188***
	(0.06408)	(2.55898)	(2.54864)	(1.02852)	(36.38192)	(35.98718)	(0.16656)	(1.13852)	(1.18267)	(1.11228)	(9.32755)	(9.68036)
Insig2u/sigma_u	-1.4014***	-1.4252***	-1.5352***	7.2804***	6.9187***	6.5964***	0.2971	-0.0492	-0.2002	10.5236***	9.7207***	8.7483***
	(0.30735)	(0.31344)	(0.33613)	(0.95712)	(0.97408)	(0.96224)	(0.24663)	(0.28202)	(0.30182)	(0.70873)	(0.72122)	(0.80718)
sigma_e				15.9198***	15.7852***	15.5935***				10.1624***	10.2759***	10.6945***
				(0.52204)	(0.51775)	(0.50671)				(0.44883)	(0.46275)	(0.50506)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses; Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

C. Probit model results for probability of dropping out from baseline to follow up survey

Table E:Probit estimation of attrition bias in samples of respective countries.

	Malawi	**Ethiopia**
VARIABLES	*Drop out in 2013(1=yes)*	*Drop out in 2014(1=yes)*
Female household head(1=yes)	-0.2383	0.2063*
	(0.1535)	(0.1037)
Age of household head(years)	-0.0083*	-0.0029
	(0.0042)	(0.0027)
Household size (count)	-0.1091	-0.1573**
	(0.0577)	(0.0607)
Household labor units	0.0063	0.0559
	(0.1026)	(0.0892)
Farm size(ha)	-0.8385***	-0.0657
	(0.2089)	(0.0486)
Household wealth index (PCA)	0.0773***	-0.0007
	(0.0208)	(0.0120)
Distance to nearest market (Km)	-0.0092	0.0004
	(0.0119)	(0.0008)
Number of plots	-0.2052*	-0.0820***
	(0.0824)	(0.0241)
_cons	0.2023	-0.6989***
	(0.2433)	(0.1849)
LR chi2(8)	120.59	89.06
Prob > chi2	0.0000	0.0000
Observations	1144	3466

Normal standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; The dependent variable is a dummy for household dropping out in follow up sample from baseline sample. The baseline sample for Malawi is 2010 survey and 2012 survey for Ethiopia.

D. Dynamic random effects models results(coefficients) where we control for potential attrition bias by including the Inverse mills ratio constructed from the probit models in(C)

Table F: Dynamic random effects Probit and Tobit models for seed purchase (All crops) decisions in Malawi and Ethiopia (Coefficients) with inverse mills ratio as an additional explanatory variable

	Malawi ("All crops model")						*Ethiopia ("All crops model")*					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
VARIABLES	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2	MUse_0	MUse_1	MUse_2	Intensity_0	Intensity_1	Intensity_2
Household purchased seed in the first panel round (1=yes)	0.3016***	0.2652***	0.2024***	3.5284***	3.9036***	2.8572**	0.7284***	0.5450***	0.5083***	19.6738***	15.8727***	14.5853***
	(0.07074)	(0.06956)	(0.06767)	(1.30123)	(1.24498)	(1.22154)	(0.10349)	(0.09365)	(0.09110)	(3.85754)	(3.78090)	(3.72035)
Household purchased seed in the previous panel round (1=yes)	0.1800**	0.1892**	0.1817**	2.3381*	2.4734**	2.0563*	0.1367	0.1688*	0.1739*	12.7173***	10.6187***	10.3897***
	(0.08376)	(0.08369)	(0.08397)	(1.23057)	(1.21162)	(1.19725)	(0.09956)	(0.09674)	(0.09539)	(3.90405)	(3.87175)	(3.84073)
Amount of seed purchased in the first panel round (kg)				0.0890*	0.0202	0.0086				0.4662***	0.4097***	0.3834***
				(0.04998)	(0.04788)	(0.04685)				(0.05720)	(0.05641)	(0.05546)
Amount of seed purchased in the previous panel round (kg)				0.0282	0.0265	0.0248				0.0477	0.0671	0.0831
				(0.05147)	(0.05047)	(0.04984)				(0.05714)	(0.05704)	(0.05638)
Farm size(ha)		0.1795***	0.2021***		7.3941***	8.4908***		0.0221**	0.0164*		1.0892***	0.9503***
		(0.05355)	(0.06960)		(0.90444)	(1.15793)		(0.01091)	(0.00988)		(0.35197)	(0.35035)
Household labor units		0.0298**	0.0293*		0.5257**	0.3605		0.0541***	0.0513***		3.1822***	3.1691***
		(0.01500)	(0.01653)		(0.25023)	(0.27357)		(0.01782)	(0.01873)		(0.98793)	(1.03824)
Rainfall historical mean (1980-2018) in mm		-0.0909	-0.1013		0.3758	1.0369		0.5165***	0.4991***		14.7920***	13.3676***
		(0.34684)	(0.34340)		(6.11662)	(6.03340)		(0.06975)	(0.06884)		(3.70905)	(3.66986)
1-year lag negative rainfall deviation		0.2742**	0.3090***		4.5756**	5.3255***		-0.3517**	-0.4420***		-9.2930	-14.0872*
		(0.11812)	(0.11794)		(1.94982)	(1.93131)		(0.14603)	(0.14577)		(8.20712)	(8.23012)
1-year lag positive rainfall deviation		0.0379	0.0243		4.1982	4.6074		-0.1934***	-0.1945***		-4.5056	-4.0248
		(0.17990)	(0.18214)		(2.98509)	(2.99130)		(0.04933)	(0.04918)		(2.81786)	(2.81469)
Temperature historical mean (1980-2018) in degrees celsius		0.0709***	0.0748***		1.0146***	1.0827***		-0.0652***	-0.0674***		-3.3244***	-3.5016***
		(0.01992)	(0.01978)		(0.34499)	(0.34220)		(0.00956)	(0.00950)		(0.51156)	(0.50531)
Community market access index(normalized)		0.0272	0.0060		0.6057*	0.1901		0.0329**	0.0331**		1.6989*	1.5631*
		(0.01849)	(0.01879)		(0.31831)	(0.31993)		(0.01653)	(0.01650)		(0.93794)	(0.93496)
Community population pressure ratio			0.0042			-0.0056			0.0266**			1.7269***
			(0.00386)			(0.06215)			(0.01055)			(0.59676)
Farm population pressure			0.0017			0.0766*			-0.0002			-0.0697**
			(0.00250)			(0.04253)			(0.00016)			(0.02897)
Household asset wealth index(normalized)			0.5973***			12.2610***			0.1570***			9.1831***
			(0.14690)			(2.36087)			(0.02441)			(1.38781)
Female decision maker (1=yes)			-0.2217***			-2.4807*			-0.0792			-8.8812*
			(0.07511)			(1.30910)			(0.08543)			(4.78101)
Household head is single(1=yes)			0.0961			0.0587			0.1193			12.2435**
			(0.07576)			(1.28899)			(0.08595)			(4.80243)
Age of household head(years)			-0.0084***			-0.1419***			-0.0037**			-0.3358***
			(0.00208)			(0.03656)			(0.00178)			(0.09973)
Education level attained at least 12th grade (Ethiopia) , at least JCE(Malawi)			0.0213			0.4810			-0.0180			0.3537
			(0.05912)			(1.01960)			(0.05111)			(2.82347)
Household dependency ratio			0.0005			0.0080			0.0007**			0.0215
			(0.00035)			(0.00608)			(0.00028)			(0.01551)
Inverse Mills Ratio (IMR)	0.0155	0.0046	0.0203	-0.0397	-0.2947	0.0163	0.0963	-0.0982	-0.1992**	10.6789***	0.7617	-4.8580
	(0.03872)	(0.03793)	(0.03729)	(0.70160)	(0.66872)	(0.65703)	(0.06300)	(0.07785)	(0.08198)	(3.39993)	(4.27567)	(4.51922)
_cons	-0.2838**	-2.1200	-1.9792	-5.3463**	-46.7130	-50.2395	-1.0179***	-3.6435***	-4.2822***	-90.3514***	-153.5018***	-182.2084***
	(0.11544)	(2.40806)	(2.39145)	(2.11231)	(42.58514)	(42.18421)	(0.15265)	(0.62353)	(0.64826)	(8.32451)	(34.06781)	(35.30597)
lnsig2u/sigma_u	-2.1167***	-2.4866***	-2.9088***	8.1908***	6.9469***	6.4642***	-0.9364***	-1.4153***	-1.5937***	30.3283***	27.0876***	23.7986***
	(0.52417)	(0.71898)	(1.04422)	(1.03854)	(1.13235)	(1.15465)	(0.29962)	(0.40776)	(0.46102)	(4.13169)	(4.48382)	(4.93350)
sigma_e				19.2959***	19.0468***	18.8662***				64.1576***	64.7034***	65.1213***
				(0.52715)	(0.52381)	(0.51593)				(1.91310)	(1.91406)	(1.90493)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses; Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

Table G: Dynamic random effects Probit and Tobit models for Maize seed purchase decisions in Malawi and Ethiopia (Coefficients) with inverse mills ratio as an additional explanatory variable

	Malawi ("Maize model")						*Ethiopia ("Maize model")*					
	Participation(1=yes)			**Intensity(kg)**			**Participation(1=yes)**			**Intensity(kg)**		
VARIABLES	Use_0	Use_1	Use_2	Intensity_0	Intensity_1	Intensity_2	MUse_0	MUse_1	MUse_2	Intensity_0	Intensity_1	Intensity_2
Household purchased maize seed in the first panel round (1=yes)	0.4300***	0.4166***	0.3500***	4.8068***	4.8639***	3.8720***	1.7159***	1.4219***	1.3267***	11.3772***	10.1387***	9.3955***
	(0.08257)	(0.08269)	(0.08110)	(1.42190)	(1.40339)	(1.38070)	(0.22779)	(0.21130)	(0.20567)	(1.24771)	(1.22424)	(1.23309)
Household purchased maize seed in the previous panel round (1=yes)	0.0804	0.0826	0.0678	-0.8843	-0.4574	-0.7173	0.1340	0.2198	0.2660*	1.1472	1.1568	1.8643
	(0.09155)	(0.09188)	(0.09215)	(1.37602)	(1.38038)	(1.36174)	(0.15023)	(0.15000)	(0.14887)	(1.11140)	(1.12275)	(1.18853)
Amount of maize seed purchased in the first panel round (kg)				0.1269	0.1035	0.0924				0.4177***	0.3992***	0.3560***
				(0.07733)	(0.07609)	(0.07462)				(0.06688)	(0.06664)	(0.06743)
Amount of maize seed purchased in the previous panel round (kg)				0.1866**	0.1641**	0.1492**				0.0073	0.0204	0.0560
				(0.07826)	(0.07722)	(0.07600)				(0.06018)	(0.06168)	(0.06581)
Farm size(ha)		0.0483	0.1155		3.3874***	4.6494***		0.0124	0.0091		0.1261*	0.1119
		(0.05876)	(0.07636)		(0.84212)	(1.07473)		(0.01130)	(0.01050)		(0.07338)	(0.07401)
Household labor units		0.0238	0.0200		0.3460	0.1811		0.0557**	0.0391		0.3979	0.2506
		(0.01616)	(0.01793)		(0.23104)	(0.25358)		(0.02733)	(0.02785)		(0.25600)	(0.26455)
Rainfall historical mean (1980-2018) in mm		0.3325	0.3511		4.0538	4.5774		0.9055***	0.8699***		8.0676***	7.8957***
		(0.28138)	(0.27983)		(4.09605)	(4.03295)		(0.12408)	(0.12015)		(1.04618)	(1.02419)
1-year lag negative rainfall deviation		0.1190*	0.1331**		1.0330	1.1949		0.5885***	0.4748**		3.5881*	2.7056
		(0.06198)	(0.06312)		(0.89126)	(0.89446)		(0.21666)	(0.21397)		(1.96833)	(1.99765)
1-year lag positive rainfall deviation		0.1394	0.1432		2.7532	2.5683		-0.3530***	-0.3441***		-3.4499***	-3.3797***
		(0.17150)	(0.17422)		(2.49230)	(2.49593)		(0.08347)	(0.08171)		(0.76566)	(0.77193)
Temperature historical mean (1980-2018) in degrees celsius		0.0460**	0.0450**		0.5206*	0.5508*		0.0126	0.0128		0.2868**	0.2823**
		(0.02107)	(0.02111)		(0.30382)	(0.30170)		(0.01445)	(0.01398)		(0.14377)	(0.14012)
Community market access index(normalized)		0.0112	-0.0085		0.1927	-0.1629		0.0401	0.0375		0.1592	0.1272
		(0.02035)	(0.02082)		(0.29533)	(0.29737)		(0.02499)	(0.02448)		(0.23397)	(0.23353)
Community population pressure ratio			-0.0035			-0.0377			0.0305*			0.2336
			(0.00422)			(0.05862)			(0.01643)			(0.16307)
farm population pressure			0.0040			0.0804**			-0.0010			-0.0187**
			(0.00273)			(0.03873)			(0.00081)			(0.00872)
Household asset wealth index(normalized)			0.5746***			10.3196***			0.1705***			1.8472***
			(0.15912)			(2.17270)			(0.03757)			(0.36943)
Female decision maker (1=yes)			-0.0915			-0.9822			-0.0928			-0.7778
			(0.08469)			(1.21516)			(0.13162)			(1.27479)
Household head is single(1=yes)			-0.0046			-0.6855			0.2412*			1.7553
			(0.08356)			(1.19558)			(0.13148)			(1.25217)
Age of household head(years)			-0.0090***			-0.1111***			-0.0006			-0.0015
			(0.00242)			(0.03484)			(0.00271)			(0.02624)
Education level attained at least 12th grade (Ethiopia), at least JCE(Malawi)			0.0380			0.8970			0.0378			-0.2468
			(0.06641)			(0.94825)			(0.07546)			(0.72375)
Household dependency ratio			0.0007*			0.0091			0.0005			0.0052
			(0.00039)			(0.00562)			(0.00041)			(0.00390)
Inverse Mills Ratio (IMR)	-0.0217	-0.0270	-0.0136	-0.2740	-0.3947	-0.1804	0.0954	-0.0451	-0.1085	1.8870**	1.0352	0.0304
	(0.04354)	(0.04371)	(0.04350)	(0.63949)	(0.63222)	(0.62460)	(0.10256)	(0.11972)	(0.12368)	(0.94250)	(1.14577)	(1.19600)
_cons	-0.5697***	-4.1565*	-4.0592*	-9.6154***	-54.6798*	-57.3639*	-2.3351***	-9.8166***	-10.6659***	-23.0421***	-94.8242***	-104.6754***
	(0.12688)	(2.13754)	(2.13638)	(1.92069)	(31.00663)	(30.69172)	(0.28217)	(1.16535)	(1.19781)	(2.30496)	(9.66305)	(9.93323)
lnsig2u/sigma_u	-1.4857***	-1.5181***	-1.6097***	7.1711***	6.8521***	6.5993***	0.2930	-0.0481	-0.2052	10.4361***	9.6955***	8.7486***
	(0.34288)	(0.35301)	(0.37503)	(1.03393)	(1.05339)	(1.03142)	(0.24729)	(0.28184)	(0.30238)	(0.71252)	(0.72264)	(0.80716)
sigma_e				15.9993***	15.8619***	15.6396***				10.1765***	10.2789***	10.6941***
				(0.55941)	(0.55580)	(0.54271)				(0.45095)	(0.46331)	(0.50522)
Regional & year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of panel households	971	971	971	971	971	971	2398	2398	2398	2398	2398	2398
Observations	2913	2913	2913	2913	2913	2913	4796	4796	4796	4796	4796	4796

Notes: ***, **, and * communicates significance at 1, 5, and 10% respectively; robust standard errors in parentheses; Participation and Intensity equations are respectively modeled by dynamic random effect Probit and Tobit models.

E. Plotting marginal effects (APEs) on the effects of rainfall variables (1-year lag negative rainfall deviation Zscore(Rshock_d1), 1-year lag positive rainfall deviation Zscore (Rshock_f1), and Rainfall historical mean (1980-2018) (lograin) on Maize seed purchase decisions:

- *In **Figure A** the margins plot shows the relationship between explanatory variables (Rshock_d1, Rshock_f1, and lograin), and the probability of purchasing maize seed (top panel), and the intensity of maize seed purchase (bottom panel).*
- *In **Figure B**, the margins show how seed purchase decisions change with rainfall variability variables between initial market participants (households that purchased maize seed in the first panel round) and non-initial market participants (those that did not purchase maize seeds in the first panel round). From this plot, we can determine whether initial market participants, because of their experience and knowledge from past engagements, have an elevated advantage in responding to rainfall shocks (by purchasing more seeds) than their opposite counterparts without experience.*
- *In **Figure C,** the margins compare the effects of rainfall variables on seed purchase decisions among the lowly endowed vs. highly endowed farm sizes. The small and large farm size variable is derived based on household total farm size.*
- *In **Figure D** the margins compare responses in household maize seed purchase decisions to rainfall variables between the rich and the poor as defined by the household's non-land assets (household wealth).*

Insights from the plots

- ✓ *The probability and intensity of Maize seed purchase in Malawi and Ethiopia increase with the previous season's negative rainfall shock (1-year lag negative rainfall deviation) and historical mean rainfall.*
- ✓ *Negative rainfall deviations and historical mean rainfall enhance seed purchase decisions to a greater extent in the group of farmers with knowledge and experience gained from previous market engagements (Initial market participants -those that purchased maize seed in the first panel round) compared to the non-participant group*
- ✓ *The effects of negative rainfall deviations on the intensity of maize seed purchase are greater in the group of farmers with larger farm sizes in Malawi. In Ethiopia, the effect of negative rainfall deviations on both the probability and intensity of Maize seed purchases is greater in farmers with relatively larger farm sizes (compared to their counterparts).*
- ✓ *The effects of rainfall shocks (negative and rainfall deviations) on the probability and intensity of Maize seed purchase decisions are positive and greater for wealthier farmers than their poorer counterparts.*
- ✓ *Overall, smallholder farmers with knowledge and experience of the maize seed markets (from previous engagements) and those with high land and non-land asset wealth endowments have an elevated advantage in purchasing Maize seeds post rainfall shock exposure as adaptation compared to their opposite counterparts.*

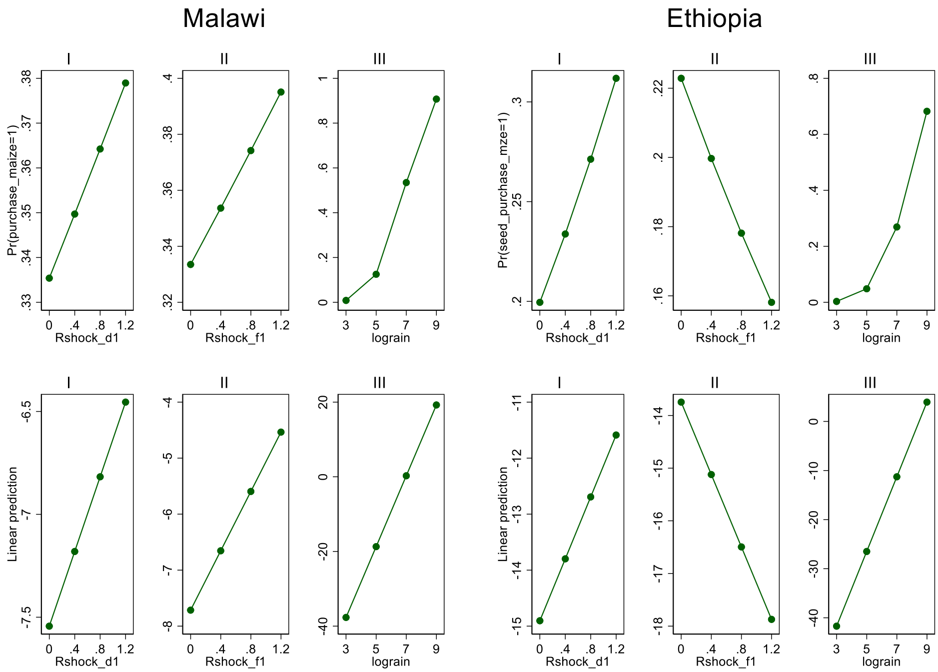

Figure A: A plot of margins showing the effect of rainfall variables (I) 1-year lag negative rainfall deviation(Rshock_d1), (II) 1-year lag positive rainfall deviation(Rshock_f1), and (III) historical mean rainfall in log form(lograin) on the probability(top panel) and the intensity(bottom panel) of Maize seed purchase decisions in Malawi (left panel), and Ethiopia(right panel).

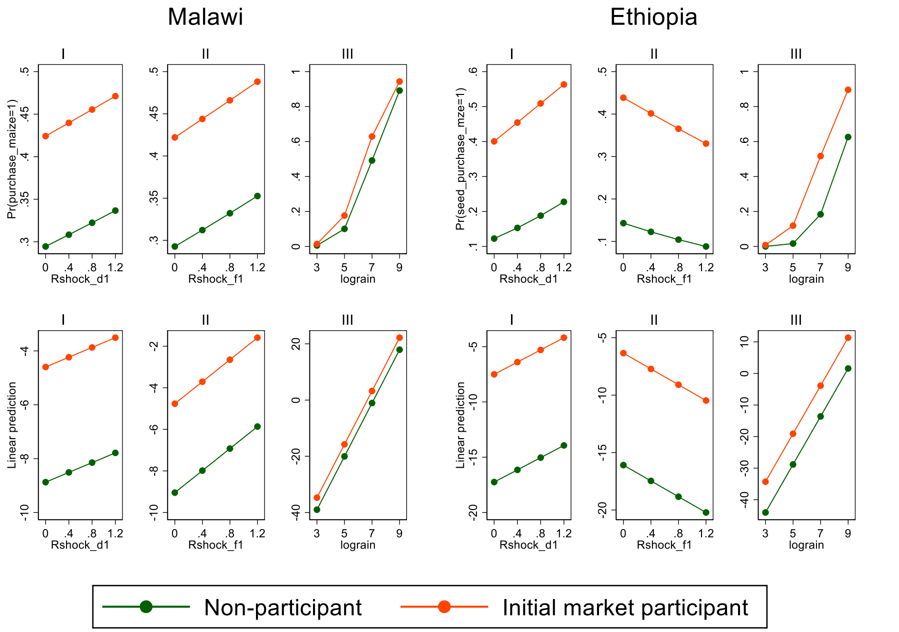

Figure B: A plot of margins showing the effect of rainfall variables (I) 1-year lag negative rainfall deviation(Rshock_d1), (II) 1-year lag positive rainfall deviation(Rshock_f1), and (III) historical mean rainfall in log form(lograin) on the probability(top panel) and the intensity(bottom panel) of Maize seed purchase decisions in Malawi (left panel), and Ethiopia(right panel) comparing initial market participants (households that purchased maize seed in the first panel round) and non-initial market participants (those that did not purchase maize seeds in the first panel round).

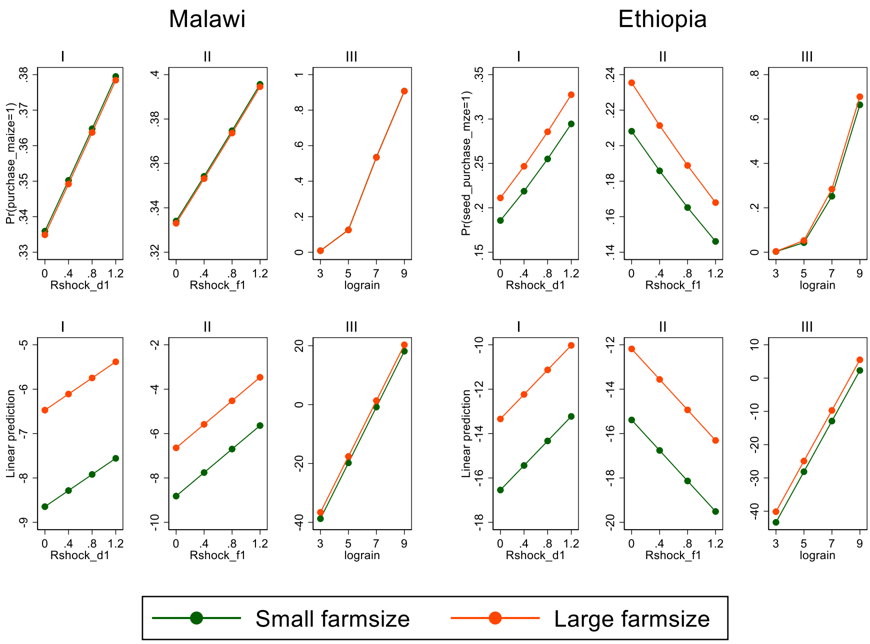

*Figure C**:** A plot of margins showing the effect of rainfall variables (I) 1-year lag negative rainfall deviation(Rshock_d1), (II) 1-year lag positive rainfall deviation(Rshock_f1), and (III) historical mean rainfall in log form(lograin) on the probability(top panel) and the intensity(bottom panel) of Maize seed purchase decisions in Malawi (left panel), and Ethiopia(right panel) comparing farmers in two quintiles of total farm size (Small farm size vs. large farm size).*

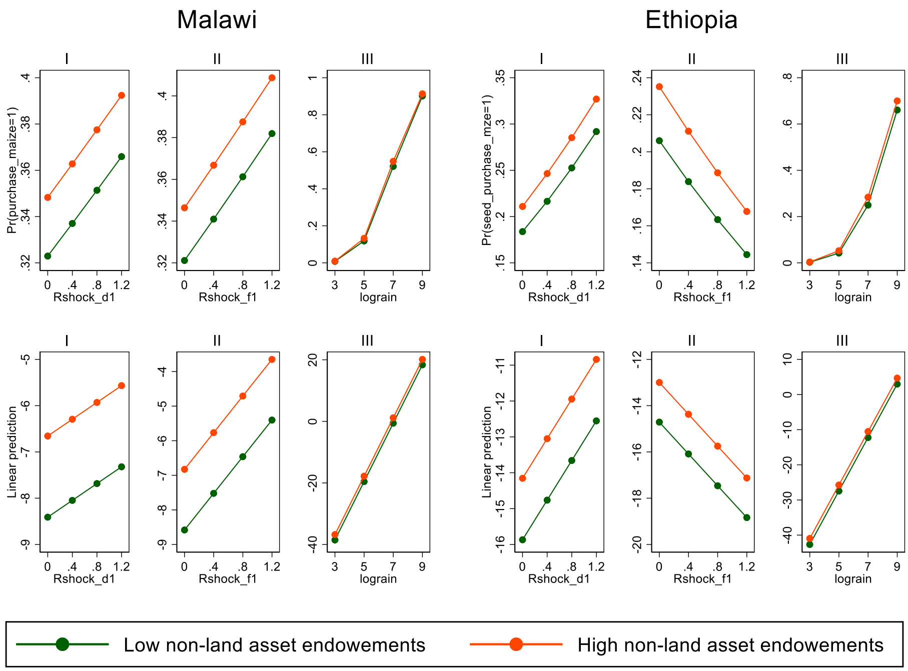

Figure D: A plot of margins showing the effect of rainfall variables (I) 1-year lag negative rainfall deviation(Rshock_d1), (II) 1-year lag positive rainfall deviation(Rshock_f1), and (III) historical mean rainfall in log form(lograin) on the probability(top panel) and the intensity(bottom panel) of Maize seed purchase decisions in Malawi (left panel), and Ethiopia(right panel) comparing farmers in two quintiles of non-land asset wealth endowments (Low vs. High).

Clifton Makate

School of Economics and Business, Norwegian University of Life Sciences (NMBU), P.O. Box 5003, 1432 Ås, Norway

Email(s):
ruumakate@live.com;
Clifton.makate@nmbu.no
Mobile: +47 9116 2671

Thesis number: 2022:65
ISSN: 1894-6402
ISBN: 978-82-575-2017-5

Clifton Makate was born in Mutare, Zimbabwe. He holds a BSc. in Agriculture (Agricultural Economics), an MSc. in Agricultural and Applied Economics, both obtained from the University of Zimbabwe in 2010 and 2013, respectively, and an MSc in Environmental Management and Sustainable Development from Tongji University, China, acquired in 2017.

This thesis comprises an introductory synthesis chapter and four individual but related empirical research papers. The thesis aims to investigate constraints and opportunities to household seed security in smallholder farming in the context of increasing climate variability, shocks, socioeconomic inequality, and pervasive transaction costs that characterize seed markets, focusing on three African countries: Tanzania, Malawi, and Ethiopia.

Paper I focuses on Ethiopia and household behavioral responses in their local and improved seed use and crop diversification decisions to recent exposure to covariate climate shocks and idiosyncratic household shocks. Results highlight a few interesting findings. First, lagged drought, temperature shocks, and historical mean rainfall enhance improved seed use. Second, lagged flood and temperature shocks and historical mean rainfall enhance crop diversification. Third, recurrent drought exposure significantly reduces overall agricultural activity. Fourth, idiosyncratic shocks minimally explain seed use and crop diversification decisions compared to covariate climate shocks. Finally, heterogeneity analysis reveals that drought shock exposure among farmers with small farm sizes and low asset endowments reduces improved seed use and diversification but increases local seed use.

Paper II focuses on seed purchase, an important dimension for understanding seed access. It explores the influence of previous exposure to drought shocks, gender, and wealth endowments on the likelihood and extent of purchasing seeds of key crops in Malawi, Tanzania, and Ethiopia. Results portray that, on average, lagged drought shock exposure increases seed purchasing for both improved and local seeds in Malawi and Tanzania while encouraging (discouraging) local (improved) seed purchases in Ethiopia. In all three countries, farmers better endowed with household assets increase seed purchasing, particularly for improved seeds, after a drought shock exposure. In addition, smaller farm sizes and low asset wealth endowments in all study countries are significant deterrents for buying seeds in the market, particularly improved seeds.

Paper III addresses the evolution of farm-level crop diversification and response to rainfall shocks and other factors in Malawi and Tanzania. Results reveal that smallholder farmers in Malawi and Tanzania respond to short-term drought shocks and long-term rainfall variability by intensifying on-farm crop diversification and that crop diversification decisions are state-dependent, implying past crop diversification enhances later diversification. Hence, *ceteris paribus* knowledge and experience from past diversification gradually reduce transaction costs in achieving subsequent crop diversification decisions. On the one hand, knowledge, linkages, and experience in formal seed markets marginally reduce transaction costs and enhance subsequent access to seeds from available markets. On the other hand, crop diversification supports *in-situ* agrobiodiversity conservation, which is also an essential source of seed and planting materials for farmers.

Paper IV analyses the dynamic nature of transaction costs in seed markets that can constrain seed access through purchases from available markets in Malawi and Ethiopia. Findings from the study reveal nonlinear effects of lagged seed purchase decisions on subsequent decisions with strong initial effects (weakening over time). In addition, the state dependency effects causing selective access to seeds over time are more pronounced in Ethiopia than in Malawi. Seed purchase from available markets hence gives an advantage to smallholder farmers with experience compared to new entrants. Further, results reveal that seed purchase decisions also respond to climate variability and shocks. Overall, results from the paper point to state dependency on the demand side of the seed market, leading to selective access to purchased seeds over time. Also, seed purchase in smallholder farming is a liquidity and risk-dependent input choice.

Main advisor: Prof. Arild Angelsen

ISBN: 978-82-575-2017-5
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

115305 / ANDVORDGRAFISK.NO