博士論文

Farmers' Livelihood Strategies and their Heterogeneity by Economic Levels: Economic Framework and Empirical Evidence from Tanzania

（農家の生活戦略と経済水準による異質性：
経済学的枠組みとタンザニアの事例）

藤本　丈史

(Abstract)

In Sub-Saharan Africa (SSA), farmers are known to use various strategies to secure their livelihoods. For example, crop diversification, in which farmers grow a wide range of crops with different environmental adaptabilities, offers the endurance from drought and flood and prevents outbreaks of crop diseases and pests. Income diversification, in which income sources are diversified over on-farm and non-farm income, enables farmers to compensate for the income loss caused by climatic and economic shocks. Previous studies have explored the effectiveness for these livelihood strategies to reduce poverty among smallholders (Michler and Josephson 2017).

Empirical studies have observed the heterogeneity of livelihood strategies by farmers' economic levels (Tesfaye and Tirivayi 2020; Makate et al. 2022; Makate et al. 2023). For example, it has been widely observed that crop diversification improves farmers' welfare among the poorest the most (Asfaw et al. 2018; Asfaw et al. 2019). In contrast, poor farmers sometimes struggle to grow multiple crops because crop diversification requires agricultural resources that poor farmers lack (e.g., land and seeds) (Weinberger and Lumpkin 2007; Birthal et al. 2013). Farmers' livelihood strategies have different roles depending on their economic levels and sometimes cannot perform the role of poverty reduction owing to the lack of resources. To utilize livelihood strategies as policy tools, we must deepen our understanding on where the heterogeneity between poor and non-poor farmers originates.

To address this problem, this thesis first constructs an economic framework to analyze choices and welfare of poor households at subsistence levels. Then, I apply this framework to explain sources of heterogeneity by households' economic levels in two livelihood strategies: crop diversification and the cultivation in a secondary cropping season. Using the Tanzanian National Panel Survey (NPS), this thesis shows that empirical findings are consistent with theoretical predictions derived using my framework.

The second chapter of this thesis develops a microeconomic framework to characterize choices and welfare of poor households at subsistence levels. I introduce a generalized form of the subsistence constraint, $f(x) \geq R$, into the standard consumer problem, i.e. the utility maximization under the budget constraint. For applications to wide research topics in development economics, this thesis maintains the generality of f when deriving microeconomic properties. In this thesis, I restrict the domain of utility functions to consumption bundles satisfying the subsistence constraint. I first show that the subsistence constraint is binding ($f(x) = R$) when the MRS does not coincide with a price ratio at a solution, which is practically useful to obtain an analytical solution. I also show that a price ratio lies between two marginal rates of substitution for a utility function u and subsistence function f at a solution. Finally, I propose a hypothesis for subsequent chapters on farmers' livelihood strategies. That is, farmers' production decisions and welfare impacts are heterogeneous between poor and

non-poor farmers, based on their different purposes: to preserve the subsistence requirement or to improve the quality of livelihoods.

Based on this microeconomic framework, the third chapter of this thesis attempts to explain that poor and non-poor farmers adopt different strategies of crop diversification. Crop diversification has received attention as a risk-reducing strategy for smallholders. Introducing the subsistence constraint into a farmer's utility maximization framework under uncertainty, this thesis illuminates three potential factors behind heterogeneous behaviors on crop diversification between poor and non-poor farmers: subsistence requirement, balanced dietary intakes, and stabilization of market income. Using the Tanzanian NPS, I then examine whether past experiences of shocks affect the adoption of crop diversification differently between poor and non-poor farmers. I rely on a threshold model to estimate heterogeneous impacts between poor and non-poor farmers. I find that poor farmers below the threshold adopt crop diversification in response to drought/flood and a rise in food price. Conversely, the loss of livestock (productive assets for tillage) traps poor farmers in low crop diversity. In contrast, non-poor farmers above the threshold adopt crop diversification in response to a fall in crop price for sale and a rise in input price for purchase.

The fourth chapter of this thesis sheds light on welfare impacts of farmers' cultivation in the short rainy season, which is a secondary cropping period in East Africa. In Tanzania, most farmers cultivate in the long rainy season from March to May and harvest crops in July and August. The short rainy season begins in September/October and continues through December/January. Farmers who cultivate in the short rainy season produce additional food until the start of the main cropping season, which may enable them to address the lack of food available during the lean season. Referring to the extended microeconomic framework in Chapter 2, I first conceptualize how additional agricultural income obtained in the short rainy season improve farmers' welfare in the lean season. Using six waves from 2008/09 to 2020/21 in the Tanzanian NPS, I then examine whether the cultivation in the short rainy season improves farmers' food consumption. I find that resource-poor farmers with small farm size gain more food consumption per acre of planted land in the short rainy season than resource-rich farmers with large farm size do. Moreover, farmers who do not have off-farm jobs obtain more calories by cultivating in the short rainy season, whereas farmers who have off-farm jobs realize more dietary diversity by doing so.

Welfare improvements by the cultivation in the short rainy season are not limited to food security. Farmers who cultivate in the short rainy season obtain additional cash income by selling harvested crops, which may enable farmers to solve wide ranges of economic issues in their livelihoods. The fifth chapter of this thesis attempts to clarify that the double-cropping in the short rainy season after the main one improves the schooling of children among Tanzanian farmers. My dataset comes from the six waves from 2008/09 to 2020/21 of the Tanzanian NPS. I find that the double-cropping increases the probability of enrollment in primary and lower-secondary schools by children, in particular by girls. The double-cropping also increases the number of boys who continue to attend schools without dropping out. Moreover, these improvements of the children's education are stronger for asset-poor farmers than for asset-rich farmers, but exist over a wide range of asset levels.

The findings in the preceding three chapters support the hypothesis in Chapter 2. I found that the adoption of crop diversification and the welfare impacts of the cultivation in the secondary cropping season are heterogeneous by farmers' economic levels. Poor farmers care about whether they fall into low income or when prices of necessities become high. In contrast, non-poor consumers does not need to care about the survival unlike poor farmers and attempt to improve the quality of livelihoods (e.g., dietary diversity, non-food consumption), rather than to preserve livelihood requirement (e.g., caloric intakes). The two livelihood strategies in this thesis reduce the possibility that farmers fall into low income and improve their food security when prices tend to be high, which leads to the heterogeneity by farmers' economic levels.

The contributions of this thesis are threefold. First, this thesis provides a convenient theoretical tool to analyze choices and welfare of poor households at subsistence levels. Previous studies often set specific functional forms and assumptions to obtain analytical solutions (e.g, using the utility function $u(x) = \log(x - R)$). In contrast, this thesis maintains the generality of the framework when deriving mathematical properties. These mathematical properties enable us to intuitively discuss a wide range of topics in development economics, without delving into detailed mathematical calculation. Second, this thesis explains how poor and non-poor farmers are differently motivated for crop diversification. Some previous studies empirically observed that the adoption of crop diversification is heterogeneous by farmers' economic levels, but did not clearly explain sources of heterogeneity (Makate et al. 2022; Makate et al. 2023). This thesis clarifies potential factors behind different motivations for crop diversification between poor and non-poor farmers. Third, this thesis shows that the cultivation in the short rainy season may be a strategic option for farmers to address economic issues in their livelihoods. Most previous studies neither focused on the cultivation in the short rainy season as a main research topic nor examined its economic roles (Holden and Shiferaw 2004; Kassie et al. 2013; Dimitrova 2021). However, the cultivation in the short rainy season is not a minor or special economic activity among farmers, as I will show in the fourth chapter. The findings of this thesis suggest that farmers who cultivate in the short rainy season can address the lack of food during the lean season and incur educational expenses for the schooling of children.

(Acknowledgement)

I first express my gratitude to my academic supervisor Dr. Aya Suzuki for her generous support throughout the doctoral course. She always welcomes a talk with anybody even though she is busy. I was able to ask for her advice anytime while feeling relaxed. I could not have finished writing this thesis or completed the doctoral course without her generous support. I admire her attitudes toward students and will attempt to practice her virtue in my academic life.

I then express my gratitude to each committee member of the PhD defense in the alphabetical order of first names. I thank Dr. Hiroyuki Nakata for his profound theoretical advice, which made me aware how to explain phenomena using microeconomics and led me to the idea of the subsistence constraint. I then thank Dr. Koshi Yoshida for his useful comments in terms of his expertise in agronomy. For my future research using hydrologic methodology, I would like to ask for his advice. I feel grateful to Dr. Takeshi Sakurai for giving useful advice and comments to my research since I attended his seminar in the master course. He also kindly undertook my academic advisor in the SPRING-GX, a fellowship program at the University of Tokyo. I feel grateful to Dr. Yuko Nakano for a lot of beneficial advice and helps in my academic activities since the master course. She gave me detailed comments on the paper of crop diversification as a discussant in the JADE conference and I owe crop-specific analyses in the paper to her.

In the acknowledgment below, I express my thanks in the alphabetical order. I thank Dr. Daisuke Oyama for his help in microeconomics and mathematics since I took his class in the master course. Because he welcomes students' visits without reservation in his weekly office hour, I was able to ask for his advice anytime when I have questions on mathematics and economics. I feel grateful to Dr. Douglas Gollin for giving me detailed inspiring comments on the subsistence constraint during his busy schedule. The talk with him during the CSAE conference is one of the memorable events in my graduate life. I thank Dr. Hisaki Kono for his useful comments in my seminar presentation. He also gave me a lot of useful advice on a theory in development economics. I appreciate the kind support by Dr. Jacob E. Ricker-Gilbert. Despite his busyness, he kindly welcomed my visit of his laboratory, arranged seminar presentations and conversation with many researchers, and gave me detailed comments on the two short-rain papers after reading them. I also thank his lab members for reading my education paper and giving me a lot of useful comments. Research talks in his lab meeting was very stimulating to me. I feel grateful to Dr. Kazushi Takahashi for his detailed comments on the education paper in a conference of Japanese Economic Association. Owing to his advice, I was able to know what points I should explain in a short-rain paper. I feel grateful to Dr. Ken Miura, who invited me to the Kyoto Environment and Development Seminar. I was very glad when he emailed me and feel honored to present in this seminar.

I feel grateful to Dr. Nobuyuki Nakamura for his help during my PhD course. He sincerely gave me information when it is necessary, which helped my graduate life a lot. I am glad that he was one of the graduates in my department. I thank Dr. Wanglin Ma for teaching me many tips for publication during his stay in my department. Without his support, I could not have published my subsidy paper.

Finally, I express my gratitude to my parents Isamu and Naomi. Their support relaxed the pressure in my academic life and enabled my research activity. I could not have written this thesis or completed the doctoral course without their continuous support. My PhD work was supported by financial aids from the Special Research Assistant program by the Graduate School of Frontier Sciences, the SPRING-GX at the University of Tokyo, and the Japan Society for the Promotion of Science (KAKENHI Grant Number JP22J11332). I bear full responsibility for all possible mistakes.

Contents

List of Figures

List of Tables

List of Abbreviations

APE	Average Partial Effects
CF	Control Function
CI	Confidence Interval
CRE	Correlated Random Effects
CV	Coefficient of Variation
DMI	Dipole Mode Index
ESR	Endogenous Switching Regression
FAO	Food and Agriculture Organization of the United Nations
FE	Fixed-Effects
FE-2SLS	Fixed-Effects Two-Stage Least Squares
GER	Gross Enrollment Rate
HDDS	Household Dietary Diversity Score
IOD	Indian Ocean Dipole
IV	Instrumental Variables
LSMS-ISA	Living Standards Measurement Study - Integrated Survey on Agriculture
MRS	Marginal Rate of Substitution
MoEST	Ministry of Education, Science and Technology
NAIVS	National Agricultural Input Voucher Scheme
NBS	National Bureau of Statistics
NPS	National Panel Survey
pp	percentage point
PSLE	Primary School Leaving Examination
SACCO	Savings And Credit Cooperatives
SE	Standard Error
SI	Simpson Index
SSA	Sub-Saharan Africa
SST	Sea Surface Temperature

Tsh	Tanzanian Shillings
WFP VAM	Vulnerability Analysis and Mapping of the World Food Programme
2SLS	Two-Stage Least Squares

Chapter 1

Introduction

1.1 Background

In Sub-Saharan Africa (SSA), farmers are known to use various strategies to secure their livelihoods under limited resources. Income diversification, which extends farmers' income sources to multiple sources such as non-farm income, enables farmers to compensate for the income loss caused by climatic and economic shocks (Loison 2015; Asfaw et al. 2018; Musumba et al. 2022). Crop diversification, which provides wide environmental adaptability by growing multiple crops in farmland, enables farmers to alleviate damages of agricultural and market shocks to their livelihoods (Asfaw et al. 2019; Makate et al., 2023). Previous studies have proposed livelihood strategies as policies to reduce poverty among SSA farmers with limited resources.

However, empirical studies suggested that SSA farmers neither uniformly adopt livelihood strategies nor evenly benefit from their adoption. For example, it has been widely observed that crop diversification improves farmers' food consumption and income among the poorest the most and contributes to poverty reduction (Michler and Josephson 2017; Asfaw et al. 2018; Asfaw et al. 2019; Tesfaye and Tirivayi 2020). In contrast, farmers must own enough agricultural resources (e.g., land and seeds) to grow multiple crops, which makes it difficult for poor farmers to adopt crop diversification (Weinberger and Lumpkin 2007; Birthal et al. 2013; Makate et al. 2022; Makate et al. 2023). Although livelihood strategies aim to reduce poverty, their adoption and welfare impacts are different by farmers' economic levels.

The heterogeneity of livelihood strategies by farmers' economic levels obscures policy implications that livelihood strategies have for poverty reduction. To utilize farmers' livelihood strategies as policy tools, we must deepen the understanding on where the heterogeneity between poor and non-poor farmers originates.

1.2 Objectives of This Thesis

Based on the research motivation in Background, this thesis aims to understand why livelihood strategies are heterogeneous by farmers' economic levels. To do so, this thesis first constructs an economic framework to analyze choices and welfare of poor households at subsistence levels. Then, I apply this framework to explain sources of heterogeneity in two livelihood strategies: crop diversification and the cultivation in a secondary cropping season. Using the Tanzanian National Panel Survey (NPS), this thesis shows that empirical findings are consistent with theoretical predictions derived using my framework. Clarifying sources of heterogeneity by economic levels is beneficial to build policies of livelihood strategies that are suitable for poor and non-poor farmers.

1.3 Outline of This Thesis

In the second chapter, this thesis develops a microeconomic framework to analyze choices and welfare of poor households at subsistence levels. I introduce a generalized functional form of the subsistence constraint, $f(x) \geq R$, into the standard consumer problem, i.e., the utility maximization under the budget constraint. For applications to wide research topics in development economics, this thesis maintains the generality of f when deriving microeconomic properties. In this thesis, I restrict the domain of utility functions to consumption bundles satisfying the subsistence constraint. I first attempt to derive a condition where the subsistence constraint is binding ($f(x) = R$), which is practically useful to obtain an analytical solution. I also attempt to show that a price ratio lies between two marginal rates of substitution for a utility function u and subsistence function f at a solution. Finally, I propose a hypothesis for subsequent chapters on farmers' livelihood strategies. That is, farmers' production decisions and welfare impacts are heterogeneous between poor and non-poor farmers, based on their different purposes: to preserve the subsistence requirement or to improve the quality of livelihoods.

In the third chapter, this thesis attempts to show that poor and non-poor farmers are differently motivated for crop diversification. Introducing the subsistence constraint into a farmer's utility maximization framework under uncertainty, this thesis illuminates three potential factors behind heterogeneous behaviors on crop diversification between poor and non-poor farmers: subsistence requirement, balanced dietary intakes, and stabilization of market income. Using the Tanzanian NPS, I then examine whether past experiences of agronomic and market shocks affect the adoption of crop diversification differently between poor and non-poor farmers. Farmers who experienced shocks in the past recognize possible risks in the future. By observing responses to shocks, I aim to reveal that poor and non-poor farmers address different risks by crop diversification. I further examine farmers' crop choices behind crop diversification.

The fourth chapter of this thesis sheds light on farmers' cultivation in the short rainy season, which is a secondary cropping period in East Africa. In Tanzania, most farmers cultivate in the long rainy season from March to May and harvest crops in July and August. The short rainy season begins in September/October and continues through December/January. Farmers who cultivate in the short rainy season produce additional food until the start of the main cropping season, which may enable them to address the lack of food available during the lean season. The conceptual framework in this chapter considers how additional agricultural income obtained in the short rainy season improve farmers' welfare in the lean season. Using six waves from 2008/09 to 2020/21 in the Tanzanian NPS, I empirically examine whether the cultivation in the short rainy season improves farmers' caloric intakes and dietary diversity. I also examine heterogeneous impacts on farmers' food consumption, depending whether farmers have off-farm jobs.

Welfare improvements by the cultivation in the short rainy season are not limited to food security. Farmers who cultivate in the short rainy season obtain additional cash income by selling harvested crops, which may enable farmers to solve wide ranges of economic issues in their livelihoods. In the fifth chapter, this thesis attempts to clarify that the double-cropping over the long and short rainy seasons improves the schooling of children among Tanzanian farmers, using the household dataset of the Tanzanian NPS. As outcome variables of children's education, I focus on the probability of children's enrollment in pre-primary, primary, and lower-secondary schools and the number of children attending these schools without dropping out.

1.4 Contributions

This thesis makes three contributions to existing studies. First, this thesis provides a convenient theoretical tool to analyze choices and welfare of poor households at subsistence levels. Previous studies often set convenient functional forms and assumptions to obtain analytical solutions (e.g, using the utility function $u(x) = \log(x - R)$). In contrast, this thesis maintains the generality of the framework when deriving mathematical properties. These mathematical properties enables us to intuitively discuss a wide range of topics in development economics, without delving into detailed mathematical calculation.

Second, this thesis explains how poor and non-poor farmers are differently motivated for crop diversification. Some previous studies empirically observed that the adoption of crop diversification is heterogeneous by farmers' economic levels, but did not clearly explain sources of heterogeneity (Makate et al. 2022; Makate et al. 2023). This thesis clarifies potential factors behind different motivations for crop diversification between poor and non-poor farmers.

Third, this thesis shows that the cultivation in the short rainy season can be a strategic option for farmers to address economic issues in their livelihoods. Most previous studies neither focused on the cultivation in the short rainy season as a main research topic nor examined its economic role (Holden and Shiferaw 2004; Kassie et al. 2013; Dimitrova 2021). However, the cultivation in the short rainy season is not a minor or special economic activity among farmers, as will be shown in Chapter 4. The findings of this thesis suggest that farmers who cultivate in the short rainy season can address the lack of food during the lean season and incur educational expenses for the schooling of children.

Chapter 2

Theoretical Framework of the Subsistence Constraint under Preferences with Restricted Domain

2.1 Introduction

In poor economies, households must secure a minimum level of food consumption for survival. This subsistence requirement forces poor households to prioritize consuming certain goods relative to others, which causes behavioral and productivity differences between poor and wealthy countries (Lagakos and Waugh 2013; Gollin et al. 2014).

Because of the importance in developing economics, many previous studies have formulated subsistence households in an economic framework. For example, some studies specified a functional form of utility as $u(c) = \log(c - R)$ $(c \geq 0, R > 0)$ (Lagakos and Waugh 2013; Gollin and Rogerson 2014; Rivera-Padilla 2020). In the theory of economic growth, an instantaneous utility function at time t is specified as $\dfrac{(c(t) - R)^{1-\theta} - 1}{1 - \theta}$ (Steger 2000; Antony and Klarl 2022). Other studies introduced the subsistence constraint $c \geq R$ (Coles and Hammond 1995; Zimmerman and Carter 2003). In all formulations, the consumption c must exceed the subsistence requirement R.

However, this subsistence constraint should be framed more generally because the meaning of subsistence is diverse, depending on research contexts. For example, consider a situation where farmers produce two crops for their self-consumption: a staple crop such as maize and rice and a non-staple crop such as vegetables. Both staple and non-staple crops contribute to farmers' dietary intakes, but with different weights. Poor farmers with limited production resources may prioritize producing and consuming the staple crop with high caloric contents for their survival. Then, the subsistence constraint that represents different weights of importance in the requirement of subsistence consumption takes the form of $\epsilon_1 c_1 + \epsilon_2 c_2 \geq R$ ($\epsilon_1 > \epsilon_2 > 0$), where c_1 and c_2 are consumption amounts of the staple and non-staple foods, respectively, and ϵ_1 and ϵ_2 are caloric amounts per unit of consumption of each food.

As another example, households may have to consume minimum levels of several dietary nutrients such as calories, proteins, and vitamins. Then, the subsistence constraint takes the form of $\min\left\{\frac{c_1}{\epsilon_1}, \frac{c_2}{\epsilon_2}\right\} \geq R$, where c_1 and c_2 are consumption amounts including each dietary nutrient. In other cases, there may be even a substitution between multiple necessities, for example, bread and water. Eating only bread may not be appropriate even in a hunger situation; starving persons also need to drink water while eating bread. It seems unlikely that aid policies to regions attacked by hunger supply either only bread or only water; they usually seem to supply both goods in a balanced proportion. Then, the subsistence constraint takes the form of $(c_1)^{\epsilon_1}(c_2)^{\epsilon_2} \geq R$, where c_1 and c_2 are consumption amounts of bread and water.

This chapter aims to develop a microeconomic framework widely applicable for analyzing households at subsistence levels. I introduce into the customary consumer problem a generalized form of the subsistence constraint, $f(x) \geq R$, where $x \in \mathbb{R}_+^N$ is a consumption bundle ($N \in \mathbb{N}$) and $R \in \mathbb{R}$ is a subsistence requirement ($\mathbb{R}_+^N$ is the non-negative domain). For wide applications to broad research topics in development economics, this chapter will then derive microeconomic properties while maintaining the generality of f.

Throughout this chapter and this thesis, I restrict the domain of utility functions to consumption bundles satisfying the subsistence constraint. In Section 2.2, I firstly show that the subsistence constraint is binding for a solution ($f(x) = R$) when the MRS there does not coincide with the price ratio. This property is practically useful to obtain an analytical solution, which is given by the intersection between the budget and subsistence lines. I also show that the price ratio lies between two marginal rates of substitution for a utility function u and subsistence function f at a solution. In Section 2.3, I propose a hypothesis for subsequent chapters on farmers' livelihood strategies. That is, farmers' production decisions and welfare impacts are heterogeneous between poor and non-poor farmers, based on their different purposes: to preserve the subsistence requirement or to improve the quality of livelihoods.

This chapter contributes to the framework of poor economics in two ways. First, the subsistence constraint can be adapted to various situations of livelihood requirement. Microeconomic properties in this chapter can be applied widely because of their generality. Second, we can easily obtain analytical solutions without delving into detailed mathematical calculation. As I will show in Section 2.2, we can easily check whether consumer solutions under the subsistence constraint are given at the intersection between the budget line and the subsistence line ($f(x) = R$). This property considerably enhances the power of graphical representation in my framework.

In this thesis, $\mathbb{R}^N_{++}$ is the interior of $\mathbb{R}^N_+$. I use the Euclidean norm $||x|| := \sqrt{x_1^2 + \cdots + x_N^2}$, where $x = (x_1, \ldots, x_N)$. I denote by $u_i(x)$ and $f_i(x)$ the partial derivatives of u and f with respect to x_i at x, respectively. Finally, let $\nabla u(x) := (u_1(x), \ldots, u_N(x))$.

2.2 Framework

2.2.1 Subsistence Constraint

This thesis first defines the consumption space $V \subset \mathbb{R}^N_+$, using the subsistence constraint below:

$$f(x) \geq R.$$

Under the subsistence constraint, the consumer first and foremost has to secure the minimum requirement in livelihoods. For example, Chapters 3 and 4 consider the requirement of caloric intakes for human bodies. If the consumer does not take enough calories for the body, he/she suffers from hunger and malnutrition, or death at worst, which forces him/her to prioritize consuming food with high caloric intakes. Chapter 5 considers two types of goods: a necessity such as food and houses and a non-necessity such as children's education. The consumer cannot afford to make his/her children take educational services until he/she secure daily food and build a house for residence. The subsistence constraint expresses the consumers' behaviors to secure livelihood requirements before enjoying free consumption choices.

Let the *subsistence set* be $V := \{x \in \mathbb{R}^N_+ \mid f(x) \geq R\}$. Then, the consumption space, on which the consumer's preference is defined, is specified as V in this thesis. The consumer cannot enjoy free consumption choices until he/she satisfies the requirement for the livelihood. Thus, this thesis assumes that the preference is formed only when the consumer satisfies the subsistence constraint. Based on the preference defined on the consumption space, I consider a utility function $u : V \to \mathbb{R}$. Throughout this chapter, $R \in \mathbb{R}$ is fixed.

The subsistence function $f : \mathbb{R}^N_+ \to \mathbb{R}$ can take various forms, depending on research contexts. Figure 2.1 displays three examples of the subsistence constraint. In Figure 2.1, indifference curves are indicated by solid lines in the domain satisfying the subsistence constraint but omitted outside V. In Figure 2.1(a), $f(x) = \epsilon_1 x_1 + \epsilon_2 x_2$, which means that the consumer faces different weights $\epsilon_1 > \epsilon_2 > 0$. Because $\epsilon_1 > \epsilon_2$, feasible consumption bundles under both the budget and subsistence constraints, or the shaded parts in Figure 2.1(a), are placed disproportionately to the side of x_1. The solution is given by x^{sub} at the intersection between the budget and subsistence lines. In x^{sub}, the consumer must prioritize consuming

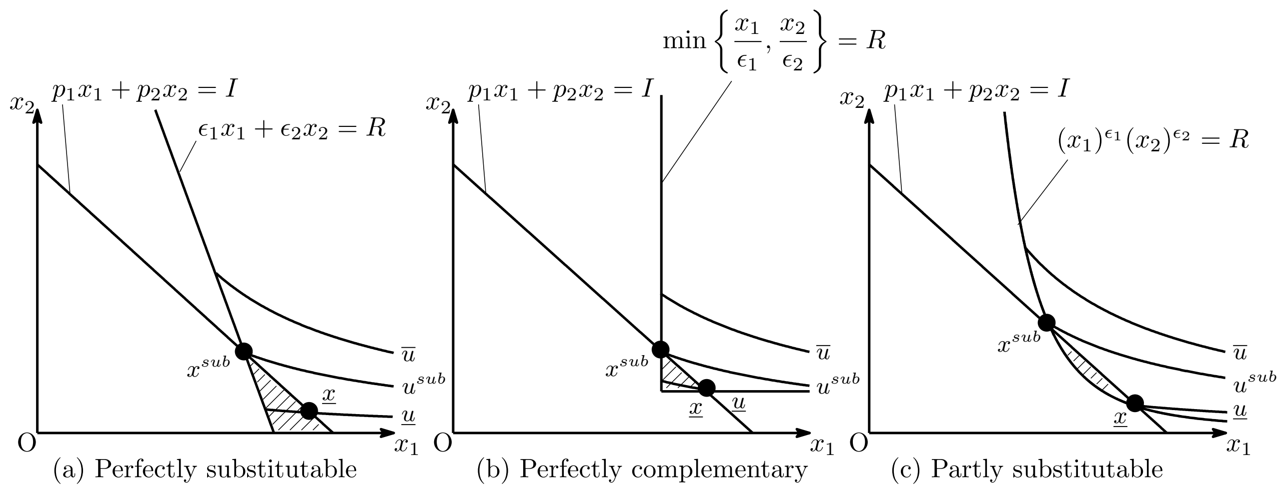


Figure 2.1: Examples of the subsistence constraint. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints.

x_1 relative to x_2 to secure the livelihood requirement.

In Figure 2.1(b), $f(x) = \left\{\frac{x_1}{\epsilon_1}, \frac{x_2}{\epsilon_2}\right\}$, which means that the consumer must consume at least a minimum level R in each good.[1] In Figure 2.1(c), $f(x) = (x_1)^{\epsilon_1}(x_2)^{\epsilon_2}$, which means that the substitution between x_1 and x_2 exists and changes depending on the proportion of consumption of these goods.

The consumer is a price taker and does not have the power to influence market price $p \in \mathbb{R}^N_{++}$. Income $I \in \mathbb{R}_{++}$ is also exogenously given. Given $(p, I) \in \mathbb{R}^{N+1}_{++}$, I define the *budget set* as $B(p, I) := \{x \in V \mid p \cdot x \leq I\}$, where $B(p, I)$ is also restricted to V. I assume that $B(p, I)$ is nonempty unless expressly stated otherwise.

Given $(p, I) \in \mathbb{R}^{N+1}_{++}$, the consumer in this chapter solves

$$\max_{x \in V} \ u(x) \text{ s.t. } x \in B(p, I),$$

where the domain of u, which would be $\mathbb{R}^N_+$ in the customary consumer problem, is restricted to V. Throughout this chapter, I denote by $X^{sub}(p, I)$ a set of solutions to the utility maximization problem above, given $(p, I) \in \mathbb{R}^{N+1}_{++}$.

2.2.2 Characterizing Solutions by Marginal Rates of Substitution, Price Ratios, and Slopes of Subsistence Functions

In Figure 2.1, the subsistence constraint is binding for the solution x^{sub}. The following theorem shows that the subsistence constraint is binding for a solution when the marginal rate of substitution (MRS) there differs from the price ratio:

Theorem 2.2.1. *Take any $(p, I) \in \mathbb{R}^{N+1}_{++}$ and any $x^{sub} \in X^{sub}(p, I)$. Suppose that u is*

[1]This subsistence constraint can be considered to represent Stone-Geary preferences with utility functions such as $u(x) = (x_1 - \bar{x}_1)^{\gamma_1}(x_2 - \bar{x}_2)^{\gamma_2} \cdots (x_N - \bar{x}_N)^{\gamma_N}$ $(\bar{x}_i \geq 0,\ 0 < \gamma_i < 1)$.

continuously differentiable in the neighborhood of x^{sub} and that f is continuous at x^{sub}. For $x^{sub} \in \mathbb{R}^N_{++}$, if $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} \neq \dfrac{p_i}{p_j}$ and $u_i(x^{sub})$, $u_j(x^{sub}) > 0$ for some i, $j \in \{1, \ldots, N\}$, then $f(x^{sub}) = R$ holds.

Proof. Without loss of generality, assume $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} > \dfrac{p_i}{p_j}$. Suppose $f(x^{sub}) > R$ for a contradiction. For $\epsilon > 0$, consider the deviation from x^{sub} by $h^\epsilon = (h^\epsilon_1, \ldots, h^\epsilon_N)$, where $h^\epsilon_i = \left(\dfrac{p_j}{p_i}\right)\epsilon$, $h^\epsilon_j = -\epsilon$, and $h^\epsilon_k = 0$ $(k \neq i, j)$. Because $x^{sub} \in \mathbb{R}^N_{++}$, f is continuous, and $f(x^{sub}) > R$, there exists $b > 0$ such that we have $x^{sub} + h^\epsilon \in \mathbb{R}^N_+$ and $f(x^{sub} + h^\epsilon) \geq R$ for all $0 < \epsilon < b$. Because u is continuously differentiable in the neighborhood of x^{sub}, it holds by the first-order Taylor expansion at x^{sub} that

$$u(x^{sub} + h^\epsilon) = u(x^{sub}) + u_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right)\epsilon + u_j(x^{sub}) \cdot (-\epsilon) + o(||h^\epsilon||),$$

where $o(\cdot)$ is a higher-order infinitesimal. By definition, $h^\epsilon \to 0$ if $\epsilon \to 0$. Thus, we can rewrite the equation above as

$$u(x^{sub} + h^\epsilon) = u(x^{sub}) + \epsilon \cdot \left\{u_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right) - u_j(x^{sub})\right\} + o(\epsilon),$$

Suppose for a contradiction that, for all $c > 0$ $(c < b)$, there exists $0 < \bar{\epsilon} \leq c$ such that $\bar{\epsilon} \cdot \left\{u_i(x^{sub}) \cdot \left(\dfrac{p_j}{p_i}\right) - u_j(x^{sub})\right\} \leq |o(\bar{\epsilon})|$. Take any sequence $\{c_n\}$ such that $c_n > 0$ for all $n \in \mathbb{N}$ and $c_n \to 0$. It follows from the supposition that, for all $n \in \mathbb{N}$, there would exist $\bar{\epsilon}_n > 0$ such that

$$\bar{\epsilon}_n \cdot \left\{u_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right) - u_j(x^{sub})\right\} \leq |o(\bar{\epsilon}_n)|$$
$$\therefore u_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right) - u_j(x^{sub}) \leq \left|\frac{o(\bar{\epsilon}_n)}{\bar{\epsilon}_n}\right|.$$

Note $\bar{\epsilon}_n \to 0$ because $c_n \to 0$. By definition, $\left|\dfrac{o(\bar{\epsilon}_n)}{\bar{\epsilon}_n}\right| \to 0$ $(\bar{\epsilon}_n \to 0)$, which implies $u_i(x^{sub}) \cdot \left(\dfrac{p_j}{p_i}\right) - u_j(x^{sub}) \leq 0$. However, $u_i(x^{sub}) \cdot \left(\dfrac{p_j}{p_i}\right) - u_j(x^{sub}) > 0$ because $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} > \dfrac{p_i}{p_j}$ and $u_i(x^{sub}),\ u_j(x^{sub}) > 0$, a contradiction.

Therefore, there exists $c > 0$ $(c < b)$ such that $u(x^{sub} + h^\epsilon) > u(x^{sub})$ for all $0 < \epsilon < c$. Note $p \cdot (x^{sub} + h^\epsilon) = p \cdot x^{sub} \leq I$ for all $\epsilon > 0$. Then, we have $u(x^{sub} + h^\epsilon) > u(x^{sub})$, $p \cdot (x^{sub} + h^\epsilon) \leq I$, and $f(x^{sub} + h^\epsilon) \geq R$ for all $0 < \epsilon < c$, which contradicts the optimality of x^{sub}. Therefore, we have $f(x^{sub}) = R$, which proves the theorem. □

The solution x^{sub} must be in the interior of $\mathbb{R}^N_+$, or $\mathbb{R}^N_{++}$, because we cannot take h^ϵ, the deviation from x^{sub}, when x^{sub} lies in the boundary of $\mathbb{R}^N_+$.

As Figure 2.1 shows, the MRS at x^{sub}, or $\dfrac{u_1(x^{sub})}{u_2(x^{sub})}$, does not coincide with the price ratio $\dfrac{p_1}{p_2}$. Then, it follows from Theorem 2.2.1 that the subsistence constraint is binding for the solution, or $f(x^{sub}) = R$.

As a corollary to Theorem 2.2.1, I obtain the following property:

Corollary 2.2.2. *Take any $(p, I) \in \mathbb{R}^{N+1}_{++}$ and any $x \in B(p, I)$ such that $\dfrac{u_i(x)}{u_j(x)} \neq \dfrac{p_i}{p_j}$ for some i, $j \in \{1, \ldots, N\}$. Suppose that u is continuously differentiable in the neighborhood of x and that f is continuous. For $x \in \mathbb{R}^N_{++}$, if $f(x) > R$, then $x \notin X^{sub}(p, I)$.*

Proof. Suppose $x \in X^{sub}(p, I)$ for a contradiction. Because $f(x) > R$, it follows from the contraposition of Theorem 2.2.1 that $\dfrac{u_i(x)}{u_j(x)} = \dfrac{p_i}{p_j}$ for all i, $j \in \{1, \ldots, N\}$, which contradicts the assumption of the corollary. □

In Figure 2.1(a), u is strictly increasing and f is nondecreasing. Thus, a solution lies within the budget line $p \cdot x = I$. However, $\underline{x}$ in Figure 2.1(a) is not a solution by Corollary 2.2.2 because $f(\underline{x}) > R$ and $\dfrac{u_1(\underline{x})}{u_2(\underline{x})} \neq \dfrac{p_1}{p_2}$.

As a property of solutions, the following theorem provides a relationship between three slopes of u, p, and f:

Theorem 2.2.3. *Take any $(p, I) \in \mathbb{R}^{N+1}_{++}$ and any $x^{sub} \in X^{sub}(p, I)$. Suppose that u and f are continuously differentiable in the neighborhood of x^{sub}. For $x^{sub} \in \mathbb{R}^N_{++}$, $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} \leq \dfrac{p_i}{p_j} \leq \dfrac{f_i(x^{sub})}{f_j(x^{sub})}$ or $\dfrac{f_i(x^{sub})}{f_j(x^{sub})} \leq \dfrac{p_i}{p_j} \leq \dfrac{u_i(x^{sub})}{u_j(x^{sub})}$ holds for all i, $j \in \{1, \ldots, N\}$ such that $u_i(x^{sub})$, $u_j(x^{sub})$, $f_i(x^{sub})$, and $f_j(x^{sub})$ are positive.*

Proof. If $N = 1$, the claim is trivial because $i = j = 1$ and $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} = \dfrac{p_i}{p_j} = \dfrac{f_i(x^{sub})}{f_j(x^{sub})} = 1$. Thus, consider $N \geq 2$. Suppose for a contradiction that there exists some $x^{sub} \in X^{sub}(p, I) \cap \mathbb{R}^N_{++}$ such that either (1) or (2) below holds for some i, $j \in \{1, \ldots, N\}$ such that $u_i(x^{sub}), u_j(x^{sub}) > 0$ and $f_i(x^{sub}), f_j(x^{sub}) > 0$:

(1) $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} > \dfrac{p_i}{p_j}$ and $\dfrac{f_i(x^{sub})}{f_j(x^{sub})} > \dfrac{p_i}{p_j}$

(2) $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} < \dfrac{p_i}{p_j}$ and $\dfrac{f_i(x^{sub})}{f_j(x^{sub})} < \dfrac{p_i}{p_j}$.

Consider Case (1) first. For $\epsilon > 0$, let $h^\epsilon = (h_1^\epsilon, \ldots, h_N^\epsilon)$, where $h_i^\epsilon = \left(\dfrac{p_j}{p_i}\right)\epsilon$, $h_j^\epsilon = -\epsilon$, and $h_k^\epsilon = 0$ $(k \neq i, j)$. Take any sufficiently small $\epsilon > 0$ so that $x^{sub} + h^\epsilon$ belongs to the neighborhood of x^{sub} for the continuous differentiability of u and f within $\mathbb{R}_+^N$. Then, because f is continuously differentiable in the neighborhood of x^{sub}, it follows from the first-order Taylor expansion at x^{sub} that

$$f(x^{sub} + h^\epsilon) = f(x^{sub}) + \epsilon \cdot \left\{ f_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right) - f_j(x^{sub}) \right\} + o(\epsilon),$$

where $o(\cdot)$ is a higher-order infinitesimal. As in the proof of Theorem 2.2.1, we have $f(x^{sub} + h^\epsilon) > f(x^{sub}) \geq R$, which implies $x^{sub} + h^\epsilon \in V$ and hence u is defined at this point. Note $p \cdot (x^{sub} + h^\epsilon) = p \cdot x^{sub} \leq I$, which implies that $x^{sub} + h^\epsilon$ is feasible under the budget and subsistence constraints. Because $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} > \dfrac{p_i}{p_j}$, we similarly have $u(x^{sub} + h^\epsilon) > u(x^{sub})$, which contradicts the optimality of x^{sub}.

Because Case (2) is equivalent to $\dfrac{u_j(x^{sub})}{u_i(x^{sub})} > \dfrac{p_j}{p_i}$ and $\dfrac{f_j(x^{sub})}{f_i(x^{sub})} > \dfrac{p_j}{p_i}$, we have proved the theorem. □

Theorem 2.2.3 shows that the price ratio lies between the MRS and the slope of f at a solution. In Figure 2.1(a), the budget line lies between the indifference curve through x^{sub} and the "lower-bound" of the consumption space V (the subsistence line of $f(x^{sub}) = R$). As a caveat, Theorem 2.2.3 does not hold in the boundary of $\mathbb{R}_+^N$ because we must take h^ϵ, the deviation from x^{sub}, in the proof. Appendix 2-I shows its counterexample.

Theorem 2.2.3 cannot be applied to a case where $f_i(x^{sub}) = 0$, as in Figure 2.1(b). The following theorem complements Theorem 2.2.3:

Theorem 2.2.4. *Take any $(p, I) \in \mathbb{R}_{++}^{N+1}$ and any $x^{sub} \in X^{sub}(p, I)$. Suppose that u and f are continuously differentiable in the neighborhood of x^{sub}. For $x^{sub} \in \mathbb{R}_{++}^N$, $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} \leq \dfrac{p_i}{p_j}$ holds for all i, $j \in \{1, \ldots, N\}$ such that $u_i(x^{sub}), u_j(x^{sub}) > 0$, $f_i(x^{sub}) > 0$, and $f_j(x^{sub}) = 0$.*

Proof. Suppose for a contradiction that $\dfrac{u_i(x^{sub})}{u_j(x^{sub})} > \dfrac{p_i}{p_j}$ for some i, $j \in \{1, \ldots, N\}$ such that $f_i(x^{sub}) > 0$ and $f_j(x^{sub}) = 0$. For $\epsilon > 0$, let $h^\epsilon = (h_1^\epsilon, \ldots, h_N^\epsilon)$, where $h_i^\epsilon = \left(\dfrac{p_j}{p_i}\right)\epsilon$, $h_j^\epsilon = -\epsilon$, and $h_k^\epsilon = 0$ $(k \neq i, j)$. Take any sufficiently small $\epsilon > 0$ so that $x^{sub} + h^\epsilon$ can belong to the neighborhood of x^{sub} for the continuous differentiability of u and f within $\mathbb{R}_+^N$. Then, because f is continuously differentiable in the neighborhood of x^{sub} and $f_j(x^{sub}) = 0$, it follows from the first-order Taylor expansion at x^{sub} that

$$f(x^{sub}+h^{\epsilon}) = f(x^{sub}) + \epsilon \cdot f_i(x^{sub}) \cdot \left(\frac{p_j}{p_i}\right) + o(\epsilon),$$

which implies $f(x^{sub}+h^{\epsilon}) > f(x^{sub}) \geq R$. Because $\frac{u_i(x^{sub})}{u_j(x^{sub})} > \frac{p_i}{p_j}$, we also have $u(x^{sub}+h^{\epsilon}) > u(x^{sub})$, which contradicts the optimality of x^{sub}. □

Theorem 2.2.4 allows for "$f_i/f_j = 0$" or "$f_i/f_j = \infty$" in Theorem 2.2.3 (When "$f_i/f_j = 0/0$," any ordering between u_i/u_j and p_i/p_j is allowed. See Appendix 2-I). In Figures 2.1(b) and 2.1(c), the subsistence constraint is binding for $\underline{x}$, similarly to x^{sub}. However, $\frac{u_1(\underline{x})}{u_2(\underline{x})} < \frac{p_1}{p_2}$ while $f_1(\underline{x}) = 0$ and $f_2(\underline{x}) > 0$ in Figure 2.1(b). In Figure 2.1(c), $\frac{u_1(\underline{x})}{u_2(\underline{x})} < \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{p_1}{p_2}$. By Theorem 2.2.3 and Theorem 2.2.4, $\underline{x}$ is not a solution.

The following theorem shows that the ordering between the three slopes of u, p, and f also works as a sufficient condition for a solution in a two dimensional case:

Theorem 2.2.5. *Let $N = 2$. Take any $(p, I) \in \mathbb{R}^3_{++}$ and any $\bar{x} \in B(p, I)$ such that $f(\bar{x}) = R$ and $p \cdot \bar{x} = I$. Suppose that u is strictly quasi-concave, strictly increasing, and continuously differentiable in the neighborhood of $\bar{x}$. Suppose also that f is quasi-concave, nondecreasing, and continuously differentiable in the neighborhood of $\bar{x}$, Finally, suppose $u_i(\bar{x}) > 0$ and $f_i(\bar{x}) > 0$ $(i = 1, 2)$. For $\bar{x} \in \mathbb{R}^2_{++}$, if $\frac{u_1(\bar{x})}{u_2(\bar{x})} < \frac{p_1}{p_2} < \frac{f_1(\bar{x})}{f_2(\bar{x})}$ or $\frac{f_1(\bar{x})}{f_2(\bar{x})} < \frac{p_1}{p_2} < \frac{u_1(\bar{x})}{u_2(\bar{x})}$ holds, then $\bar{x}$ is the unique solution, or $X^{sub}(p, I) = \{\bar{x}\}$.*

In case $f_i(x) = 0$ ($i = 1$ or 2), the following theorem complements Theorem 2.2.5:

Theorem 2.2.6. *Let $N = 2$. Take any $(p, I) \in \mathbb{R}^3_{++}$ and any $\bar{x} \in B(p, I)$ such that $f(\bar{x}) = R$ and $p \cdot \bar{x} = I$. Suppose that u is strictly quasi-concave, strictly increasing, and continuously differentiable in the neighborhood of $\bar{x}$. Suppose also that f is quasi-concave, nondecreasing, and continuously differentiable in the neighborhood of $\bar{x}$, Finally, suppose $u_k(\bar{x}) > 0\,(k = 1, 2)$, $f_i(\bar{x}) > 0$, and $f_j(\bar{x}) = 0$. For $\bar{x} \in \mathbb{R}^2_{++}$, if $\frac{u_i(\bar{x})}{u_j(\bar{x})} < \frac{p_i}{p_j}$ holds, then $\bar{x}$ is the unique solution, or $X^{sub}(p, I) = \{\bar{x}\}$.*

The proofs are shown in Appendix 2-I. The two theorems above show that, in the two dimensional cases of Figures 2.1(a)-(c), a solution is uniquely determined at x^{sub}, which is located at the intersection between the budget line ($p \cdot x = I$) and the subsistence line ($f(x) = I$).

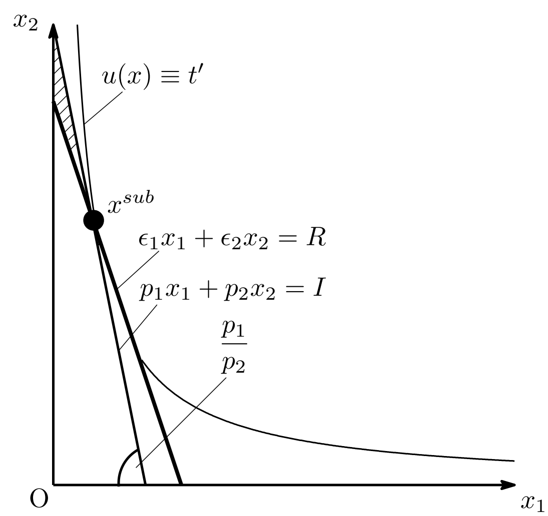


Figure 2.2: The second channel in which the subsistence constraint becomes binding. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints.

2.3 Implications

The theorems in the previous subsection suggest two channels in which the subsistence constraint binds: (1) when income is extremely low and (2) when prices of necessity goods are very high. To illustrate this, consider two goods, x_1 and x_2, and the subsistence constraint $f(x) = \epsilon_1 x_2 + \epsilon_2 x_2$, where $\epsilon_1 > \epsilon_2 > 0$. Because $\epsilon_1 > \epsilon_2$, consuming x_1 is more crucial for satisfying the subsistence constraint than consuming x_2. Thus, x_1 is a necessity good, whereas x_2 is a non-necessity good. Suppose that the marginal utility is positive and diminishing.

To explain the first channel (1), note that a poor consumer with low income has to prioritize consuming x_1 relative to x_2 for survival, as Figure 2.1(a) shows. Consequently, u_1 is considerably low relative to u_2. Then, the MRS is below the price ratio, or $\dfrac{u_1}{u_2} < \dfrac{p_1}{p_2}$. Although a poor consumer wishes to consume x_1 and x_2 to the point where the MRS coincides with the price ratio, they cannot do so because the consumer's income is too low. By Theorem 2.2.1 and Corollary 2.2.2, $f(x^{sub}) = R$ holds for the solution x^{sub}, which is the corner solution of the consumption space.

To explain the second channel (2), suppose that the price of the necessity good, p_1, rises dramatically relative to p_2 due to some external shocks. Then, the price ratio p_1/p_2 rises, leading to $p_1/p_2 > \epsilon_1/\epsilon_2$, which is shown in Figure 2.2. Although the consumer wants to consume x_1 up to the point where $u_1/u_2 = p_1/p_2$, he/she cannot do so because of the lack of income to consume enough amounts of the high-priced necessity. Thus, $\epsilon_1/\epsilon_2 < p_1/p_2 < u_1/u_2$ holds at x^{sub} in Figure 2.2. By Theorem 2.2.5 and Theorem 2.2.6, the subsistence constraint is binding for the solution.

It is notable that when such cases occur, the optimal choice is the corner solution, rather than on the interior of the feasible consumption space. As farmers wish to avoid facing such situations, they can take measures against this possibility through their production decisions that take place one period before. Based on the theories developed thus far, I hypothesize that poor farmers, who are more likely to fall into such situations where the subsistence constraint binds, and non-poor farmers, who are less likely to suffer from such situations, take different strategies in their production decisions. Further, as their strategies are heterogeneous, the welfare impacts are also expected to be heterogeneous depending upon the farmers' economic levels.

In the following chapters, I focus on two production decisions, which are crop diversification and the cultivation in a secondary cropping season. These production strategies have two roles in preventing farmers from falling into situations where the subsistence constraint binds. First, the two livelihood strategies may increase overall income so that farmers' income will not be extremely low as discussed as channel (1). Crop diversification increases farmers' production volume through agronomic synergies between crops. The cultivation in a secondary cropping season provides additional food and income to farmers within a year. Both of the strategies have a role to prevent farmers from falling into strikingly low income by expanding farmers' "production frontiers." Second, the two livelihood strategies weaken effects of the volatility in crop prices even in situations where the market price of one good surges as discussed as channel (2). By planting a wide range of crops for self-consumption, farmers can improve the food security when a food price dramatically rises. By cultivating in a secondary cropping season, farmers can improve the food security during the lean season when prices of staple crops tend to be high.

Because crop diversification and the cultivation in a secondary cropping season have these two roles, poor and non-poor farmers are predicted to adopt these strategies, based on different purposes: to preserve the subsistence requirement or to improve the quality of livelihoods. I will test in subsequent chapters whether the motivation to adopt these livelihood strategies and their welfare impacts are different by farmers' economic levels.

2.4 Conclusion

This chapter has introduced the subsistence constraint, $f(x) \geq R$, into the customary consumer theory. Because of a generalized functional form of the subsistence constraint, my framework can apply to wide research topics in developing economics. I first show that the subsistence constraint is binding for a solution ($f(x) = R$) when the MRS does not coincide with the price ratio. I also show that the price ratio lies between two marginal rates of substitution for u and f at a solution.

Based on these findings, I propose the hypothesis for subsequent chapters on farmers' livelihood strategies. The subsistence constraint becomes binding when consumers fall into low income or prices of necessities become high. Poor consumers, whose subsistence constraint is binding, cannot lower the consumption level below R and are expected to take measures in the previous production stage to prevent them from falling into low income or being severely affected by the price hikes. In contrast, non-poor consumers do not need to care about the subsistence constraint and are thus expected to attempt to improve the quality of livelihood rather than to preserve the subsistence requirement. Therefore, I hypothesize that farmers' production decisions and their welfare impacts are heterogeneous between poor and non-poor farmers. I will test this hypothesis in subsequent chapters which focus on crop diversification and the cultivation in a secondary cropping season as farmers' livelihood strategies.

Chapter 3

Different Strategies of Crop Diversification Between Poor and Non-Poor Farmers: Concepts and Evidence from Tanzania

3.1 Introduction

Crop diversification, or growing multiple crops in farmland, has received attention as a strategy that smallholders can employ for poverty reduction. Cultivating a wide range of crops with different environmental adaptabilities offers the endurance of drought and flood and prevents outbreaks of crop diseases and pests (Asfaw et al. 2019). Diverse compositions of harvested crops for sale stabilize farmers' income, which is susceptible to price fluctuations in markets (Asfaw et al. 2018).

Numerous empirical studies showed that crop diversification improves farmers' welfare through increased crop productivity, agricultural income, food consumption, and dietary diversity (Di Falco and Chavas 2008; Di Falco et al. 2010; Jones et al. 2014; Birthal et al. 2015; Dillon et al. 2015; Hirvonen and Hoddinott 2017; Arslan et al. 2018; Asfaw et al. 2018; Ecker 2018; Asfaw et al. 2019; Bellon et al. 2020; Mulwa and Visser 2020; Tesfaye and Tirivayi 2020). Furthermore, crop diversification stabilizes crop income and reduces its downside risks (Smale et al. 1998; Di Falco and Perrings 2005; Di Falco and Chavas 2009; Bozzola and Smale 2020; Maggio and Sitko 2021). Importantly, crop diversification benefits the poorest the most and raises farmers out of poverty (Michler and Josephson 2017; Asfaw et al. 2018; Asfaw et al. 2019; Tesfaye and Tirivayi 2020).

Although crop diversification is conducive to reducing poverty, poverty may constrain farmers' capabilities of handling multiple crops in farmland. In the allocation of limited agricultural resources, poor farmers may first and foremost prioritize cultivating staple crops with high-caloric intakes for their survival (Rivera-Padilla 2020). Most studies examining determinants of crop diversification found that agricultural resources such as land sizes, asset levels, and livestock holdings are positively associated with farmers' adoption of crop diversification (Di Falco et al. 2010; McCord et al. 2015; Michler and Josephson 2017; Arslan et al. 2018; Asfaw et al. 2018; Asfaw et al. 2019; Bellon et al. 2020; Bozzola and Smale 2020; Tesfaye and Tirivayi 2020; Maggio and Sitko 2021; Makate et al. 2022; Makate et al. 2023). The adoption of crop diversification is more difficult for poor farmers with low wealth levels, which conflicts with the role of crop diversification for poverty reduction.

Because of this conflicting nature, farmers may take two opposite paths of crop diversification when they experience shocks. If farmers who experienced shocks recognize the importance of growing multiple crops for risk reduction, they may adopt crop diversification to address future shocks. In contrast, poor farmers with limited assets may be vulnerable to shock exposure (Dercon 2002). If farmers incur long-standing damages by past shocks, they may lose capabilities of growing multiple crops. Empirical studies observed the two opposite paths of crop diversification. Some studies showed that crop diversity in farmland increased in response to past drought, crop disease, and food and input price fluctuation (Asfaw et al. 2018; Malawi and Niger in Asfaw et al. 2019; Bozzola and Smale 2020; Mulwa and Visser 2020; Tesfaye and Tirivayi 2020; Coromaldi et al. 2015). Conversely, other studies showed that crop diversity decreased in response to past drought and flood (Maggio and Sitko 2021; Zambia in Asfaw et al. 2019; Coromaldi et al. 2015).

Given these contradictory findings, it is unclear whether farmers are really motivated to adopt crop diversification as a strategy to reduce poverty. Makate et al. (2022) and Makate et al. (2023) found that past drought reduces crop diversity among poor Ethiopian farmers with low wealth levels and increases it among wealthy Tanzanian and Malawian farmers, which suggests that poverty constrains poor farmers' adoption of crop diversification. However, we are still unsure in their findings how poor farmers are motivated to adopt crop diversification under poverty constraints. Furthermore, if the adoption of crop diversification is heterogeneous by wealth levels, how are non-poor farmers motivated to adopt crop diversification, differently from poor farmers?

To fill in this research gap, this chapter first formalizes a microeconomic framework to analyze farmers' heterogeneous behaviors on crop diversification. By introducing a subsistence constraint into the customary consumer problem, I highlight three potential factors behind heterogeneous behaviors between poor and non-poor farmers: namely, subsistence requirement, balanced dietary intakes, and stabilization of market income.

Using the Tanzanian National Panel Survey, this chapter then examines whether farmers' experiences of agronomic and market shocks affect their adoption of crop diversification. I distinguish between poor and non-poor farmers based on food consumption levels, because subsistence levels seem to work as a source of their heterogeneous behaviors. By observing farmers' responses to these shocks separately by consumption levels, I can reveal that poor and non-poor farmers adopt crop diversification to address different risks in their livelihoods.

To clarify the heterogeneity, this chapter introduces a threshold model that assumes an unobserved breakpoint in farmers' food consumption levels, or a threshold. All farmers are divided into two regimes: poor farmers below the threshold and non-poor farmers above the threshold. Using the fixed-effects threshold model of Hansen (1999), this chapter examines whether farmers who experienced shocks adopt crop diversification differently in the two regimes.

Finally, this chapter examines what type of crops poor and non-poor farmers decide to cultivate in response to shocks. The crop-specific analysis uncovers farmers' crop choices behind crop diversification and reveals important agronomic and market traits in their strategies of crop diversification (Maggio and Sitko 2021).

The contributions are twofold. First, this study deepen the argument on where the heterogeneity between poor and non-poor farmers originates in the context of crop diversification. Makate et al. (2022) and Makate et al. (2023) suggested that poor farmers are constrained from adopting crop diversification. Then, a new question arises on how we should place crop diversification in relation to its role of poverty reduction. This study attempts to answer this question by illuminating the three potential factors as sources of heterogeneity. Second, this study contributes to a better understanding of crop diversification in terms of *ex-ante* and *ex-post* strategies. Ex-ante strategies are risk management to prepare for future shocks in advance (Dercon 2002). Ex-post strategies are feedback actions coping with damages after unanticipated shocks occur (Asfaw et al. 2018). In the context of crop diversification, farmers may diversify crop portfolios *ex ante* to stabilize crop income (Smale et al. 1998; Di Falco and Perrings 2005; Di Falco and Chavas 2009; Bozzola and Smale 2020; Maggio and Sitko 2021). In contrast, farmers whose productive resources were damaged by shocks may specialize in growing staple crops *ex post* to preserve subsistence levels (Makate et al. 2022). Based on the three potential factors illustrated in the conceptual framework, this study attempts to explain why farmers' strategies of crop diversification can be ex ante in some cases and ex post in others.

This chapter proceeds as follows: Section 3.2 provides the conceptual framework; Section 3.3 explains the dataset and the estimation strategy; Section 3.4 presents the estimation results; Section 3.5 explores crop-specific analysis; and Section 3.6 concludes this chapter.

3.2 Conceptual Framework

I construct a conceptual foundation based on a state-contingent framework to characterize the risk-reducing nature of crop diversification.

3.2.1 Theoretical Framework

Structure of Farmer's Decisions under Uncertainty

Crop diversification is a risk-reducing strategy and improves farmers' welfare by alleviating uncertainty of future outcomes. To describe decisions under uncertainty, I introduce a set of states of nature, Ω, and consider a farmer who maximizes the expected utility over Ω. In each state, the farmer encounters different climatic and market conditions. In one state, he/she cultivates crops in a climatically favorable condition; in another state, he/she obtains low on-farm profits because of worse market conditions.

To highlight a risk-reducing role of crop diversification, I suppose that the process of farmer's decisions comprises two stages. In the first stage, the farmer determines the degree of crop diversification before a state is realized. In the second stage after a state is realized, consumption bundles are chosen.

First-Stage Decision on Crop Diversification

In the first stage, the farmer is uncertain which state is realized, and therefore maximizes an expected utility, based on his probability assessments for each state. Denote by $\pi(\omega; h)$ $(0 < \pi(\omega; h) < 1)$ the subjective probability for state $\omega \in \Omega$ conditional on farmer's personal history h, and by $E = (e_1, \ldots, e_J)$ the land allocation, where e_j is the amount of land allocated to crop j. Then, the farmer in the first stage chooses E to maximize the expected utility $U(E)$ below:

$$\max_{E} \ U(E) = \sum_{\omega \in \Omega} \pi(\omega; h) V(\omega, E), \tag{3.2.1}$$

where $V(\omega, E)$ is a maximized utility in the second stage, given ω and E, as will be defined later. The subjective probability $\pi(\omega; h)$ reflects the idea that probability assessments may be affected by farmer's personal experience h. Farmers who experienced a shock may assess the probability of the same shock occurring again more greatly than those who did not.

Second-Stage Decision on Consumption in Each State

Given the first-stage decision on crop diversification, suppose state ω is realized. In the second stage, the farmer maximizes a utility function $u(c, q)$, where $c = (c_1, \ldots, c_J)$ is a food consumption bundle and $q = (q_1, \ldots, q_K)$ is a non-food consumption bundle. I assume u

is strictly concave, strictly increasing, and differentiable (its implication will be discussed later).

The farmer chooses (c, q) under two constraints: a budget constraint and a subsistence constraint. The budget constraint is the same as that in the customary consumer theory, but varies by ω and E. Define the price vectors of the goods c and q by $p_c(\omega)$ and $p_q(\omega)$, which is state-dependent. Then, the budget constraint in ω is written as

$$\sum_{j=1}^{J} p_{c_j}(\omega)c_j + \sum_{k=1}^{K} p_{q_k}(\omega)q_k \leq I_f(\omega, E) + I_n, \tag{3.2.2}$$

where $I_f(\omega, E)$ and I_n in the right side are on-farm and non-farm income, respectively. Although I_n is exogenously given, $I_f(\omega, E)$ changes by the choice of E in the first stage, as well as ω.

I further introduce the subsistence constraint to the customary consumer theory. The farmer is required to consume enough food for survival so that a weighted average of food consumption outweighs a subsistence level R:

$$\sum_{j=1}^{J} \epsilon_j c_j \geq R, \tag{3.2.3}$$

where a nonnegative weight ϵ_j is allocated to each food consumption c_j. For example, staple crops (e.g., maize) have high caloric content (Rivera-Padilla 2020) and hence high weights; cash crops (e.g., coffee) have low energy content and low or zero weights. The subsistence constraint is assumed to be intrinsic to all individuals and is provided independently of states.

Under the budget and subsistence constraints, the farmer obtains the optimal consumption $(c(\omega, E), q(\omega, E))$ to maximize u:

$$u\big(c(\omega, E), q(\omega, E)\big) = \max_{(c,q)} \ u(c, q) \ \text{ subject to Equations (3.2.2) and (3.2.3)} \tag{3.2.4}$$

The maximized utility in the second stage, $u\big(c(\omega, E), q(\omega, E)\big)$, corresponds to $V(\omega, E)$ in Equation (3.2.1).

The two optimizations in the first and second stages follow a backward solution: we first obtain $(c(\omega, E), q(\omega, E))$ and $V(\omega, E)$ in the second stage, given ω and E; using $V(\omega, E)$, we then solve Equation (3.2.1) to obtain E in the first stage.

Convex and Risk-Averse Preference of Utility Function

The utility function in the second stage, u, is assumed to be strictly concave, which implies two key properties on the farmer's consumption preference (see Appendix 3-I for mathematical details). First, the preference is convex: an indifference curve is convex-shaped to the origin (see Figure 3.1). This property is attributable to the fact that u is *strictly quasi-concave* if it is strictly concave. The convex preference expresses the farmer's desire to achieve balanced consumption. Second, the preference is risk-averse: the farmer prefers to achieve consumption with certainty rather than with uncertainty. This property motivates the farmer to adopt crop diversification as a risk-reducing strategy because he/she maximizes the expected utility in Equation (3.2.1).

Constraint for Subsistence Requirement

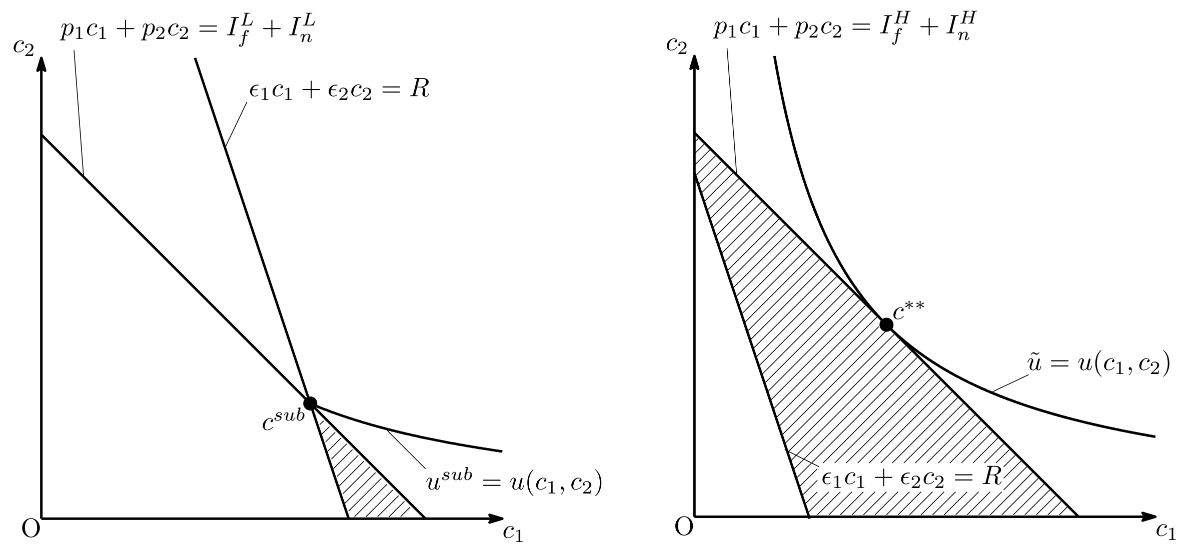
To explain how the subsistence constraint affects the farmer's consumption choice in the second stage, Figure 3.1 depicts choices of two food consumption bundles, c_1 and c_2, by poor and non-poor farmers. In Figure 3.1(a), the low income of poor farmers, $I_f^L + I_n^L$, shrinks the shaded part of feasible consumption bundles that satisfy both the budget and subsistence constraints. Then, the solution is given by the intersection between the budget and subsistence constraints, c^{sub}, and $f(c^{sub}) = R$ holds. In c^{sub}, poor farmers prioritize consuming the high-weight food c_1 relative to the low-weight one c_2. The marginal rate of substitution (MRS) in c^{sub} does not coincide with the slope of the budget constraint. Consequently, poor farmers attain the low utility level $u^{sub} = u(c^{sub})$.

In contrast, non-poor farmers have sufficiently high income $I_f^H + I_n^H$. In Figure 3.1(b), the optimal consumption under the budget constraint, c^{**}, satisfies the subsistence constraint. Moreover, $\epsilon_1 c_1^{**} + \epsilon_2 c_2^{**} > R$ and the MRS coincides with the price ratio. No consumption distortion arises because non-poor farmers have relatively high income and a set of feasible consumption bundles is sufficiently large.

The Lagrange function for Equation (3.2.4) also reveals how the subsistence constraint affects farmers' consumption choices. Differentiating the Lagrange function for Equation (3.2.4) with respect to c_j and q_k, I obtain:

$$\frac{\partial u}{\partial c_j} = \lambda_1 p_{c_j}(\omega) - \lambda_2 \epsilon_j, \quad \frac{\partial u}{\partial q_k} = \lambda_1 p_{q_k}(\omega), \tag{3.2.5}$$

where λ_1 is the Lagrange multiplier for the budget constraint and λ_2 is for the subsistence constraint. A Lagrange multiplier is nonnegative as its property: it is positive if its constraint is binding and zero if it is slack. Therefore, if the subsistence constraint is slack, as in Figure 3.1(b), we have $\lambda_2 = 0$. The farmer's consumption choice is unaffected by the subsistence constraint. However, if the subsistence constraint is binding, as in Figure 3.1(a),



(a) Poor farmers: subsistence constraint is binding (b) Non-poor farmers: subsistence constraint is slack

Figure 3.1: Graphical representation of the subsistence constraint between poor and non-poor farmers. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints

we have $\lambda_2 > 0$ and $\dfrac{\partial u}{\partial c_j}$ declines by $\lambda_2\epsilon_j > 0$. If the marginal utility is positive and diminishes, the farmer prioritizes consuming more c_j than he/she would choose without the subsistence constraint. This effect becomes larger as ϵ_j gets larger. Non-food consumption, q, has no weight in the subsistence constraint and therefore weakly demanded relative to c if the constraint is binding.

Functional Form of On-Farm Income

In the following argument, I specify on-farm income $I_f(\omega, E)$ as a linear combination of the allocated land to crop j:

$$I_f(\omega, E) = \sum_{j=1}^{J} p_{c_j}(\omega)\alpha_j(\omega)e_j + \sum_{k\neq l} d_{kl}(\omega)e_k e_l, \tag{3.2.6}$$

where $p_j(\omega)$ and $\alpha_j(\omega)$ are price and productivity shocks for crop j ($j = 1, \ldots, J$) in ω, respectively. I further introduce agronomic synergies between crops by adding the second term $\sum_{k\neq l} d_{kl}(\omega)e_k e_l$. The coefficient $d_{kl}(\omega) \geq 0$ represents a complementary effect between crops k and l, which depends on state ω. The total amount of land is fixed as $\overline{E}$, which means $\sum_{j=1}^{J} e_j \leq \overline{E}$.

The on-farm income $I(\omega, E)$ is volatile because of the price shock $p_j(\omega)$ and the productivity shock $\alpha_j(\omega)$. Crop diversification affects the distribution of the on-farm income in two channels. First, it stabilizes the on-farm income. Even if a shock decreases the revenue obtained through the production of crop j, farmers can alleviate the damage by planting another crop l. By allocating land to the production of a wide range of crops (e.g., $e_j = \bar{E}/J$

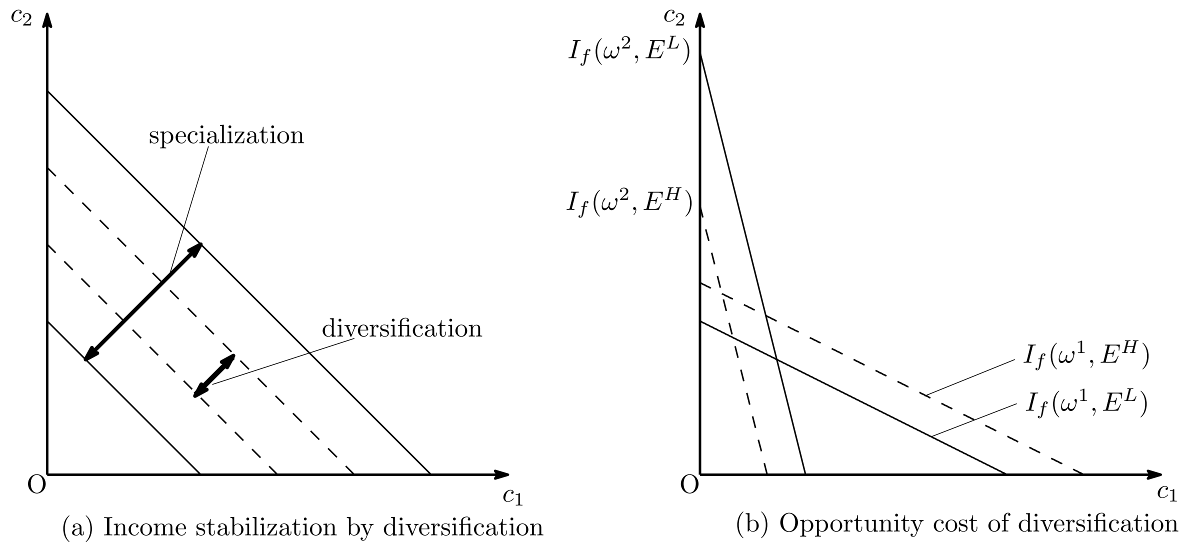


Figure 3.2: Graphical representation of trade-off between diversification and specialization. The budget lines under specialization are depicted by solid lines. The budget lines under diversification are depicted by dotted lines.

for all j), farmers can reduce the volatility of $I(\omega, E)$. Second, crop diversification increases the on-farm income. In Equation (3.2.6), planting different crops produces complementary effects ($d_{kl}(\omega)$). For example, intercropping legumes provides soil nutrients for staple crops, which increases the total production.

Due to the price and productivity shocks, $p_j(\omega)$ and $\alpha_j(\omega)$, farmers face a trade-off between diversification and specialization, which is illustrated in Figure 3.2. The budget lines under specialization are depicted by solid lines, whereas the budget lines under diversification are depicted by dotted lines.

As a merit of crop diversification, on-farm income is stable over states. If farmers plant only high-priced crops and face no shocks, they earn high on-farm income. However, the on-farm income declines considerably if farmers plant only one crop (e.g., $e_j = \bar{E}$ for some crop j) and a shock hits its price or production. Figure 3.2(a) shows that diversifying crop portfolios weakens the fluctuation of the on-farm income over states relative to specializing in certain crops.

However, farmers who diversify crop portfolios may incur an opportunity cost for not planting profitable crops. This opportunity cost becomes large if a crop price becomes high. In Figure 3.2(b), crop 1 has a low price relative to crop 2 in ω^1. In ω^1, farmers who diversify obtain higher on-farm income ($I(\omega^1, E^H)$) than those who specialize in crop 1 do ($I(\omega^1, E^L)$), because the diversifying farmers allocate less land to crop 1. However, suppose that the price of crop 1 becomes higher in ω^2 due to some external shocks. Then, farmers who specialize in crop 1 obtain higher income ($I(\omega^2, E^L)$) than those who diversify do ($I(\omega^2, E^H)$). Moreover, the diversifying farmers must pay a higher price when they purchase crop 1. As I will show, this opportunity cost is serious for poor farmers whose subsistence constraint tends to bind.

The on-farm income $I(\omega, E)$ is restricted by the total land endowment $\bar{E}$. If farmers own low amounts of land when they determine the land allocation in the first stage, the potential amount of $I(\omega, E)$, which is realized in the second stage, is limited. Then, the subsistence constraint tends to be binding in the second stage. Thus, poor farmers with low amounts of land may utilize the roles of crop diversification to secure at least the subsistence level in any state. In contrast, non-poor farmers, who own enough amounts of the total land endowments for survival in the first stage, may utilize crop diversification to stabilize the on-farm income over states and increase the expected income in the second stage.

3.2.2 Potential Factors Behind Farmers' Heterogeneous Behaviors

Using the defined setup, I illustrate that three potential factors (subsistence requirement, balanced dietary intakes, and stabilization of market income) affect decisions on crop diversification differently between poor and non-poor farmers.

In what follows, I assume perfect neoclassical markets to simplify the argument. As a caveat, three potential factors behind crop diversification are characterized by the subsistence constraint, and not by the existence of markets or market frictions. Thus, the three factors still work even in imperfect markets. Appendix 3-I explains that the essence of my argument on the three factors is the same, regardless of whether markets are perfect or imperfect.

Moreover, this assumption seems relatively reasonable because, as I will show, my dataset indicates that poor and non-poor farmers have similar participation rates in both input and output markets, which also motivates us to explain without assuming market frictions (e.g., interregional trade costs).

Subsistence Requirement: Poverty Traps

The highest priority of poor farmers is to satisfy the subsistence constraint in any state.[1] As shown in Figure 3.1(a), their feasible consumption bundles are restricted by the subsistence constraint, and hence susceptible to the fluctuation of on-farm income over states. To preserve the subsistence-level consumption, poor farmers may prioritize cultivating staple crops with high calories, which constrains their capabilities of growing multiple crops (poverty traps).

Suppose there are two crops, the primary crop c_p with a high-caloric amount and the secondary crop c_s with a low-caloric amount. Let $E = (e_p, e_s)$ be land allocated to each crop, with the total land $\overline{E} \equiv e_p + e_s$ fixed. Based on Equation (3.2.6), $I_f(\omega, E)$ becomes

$$I_f(\omega, E) = p_p(\omega)\alpha_p(\omega)e_p + p_s(\omega)\alpha_s(\omega)e_s. \tag{3.2.7}$$

To highlight poverty traps, I omit complementary effects between e_p and e_s for simplicity.

[1] I implicitly assume that, for some E, there exists a feasible consumption bundle for all $\omega \in \Omega$.

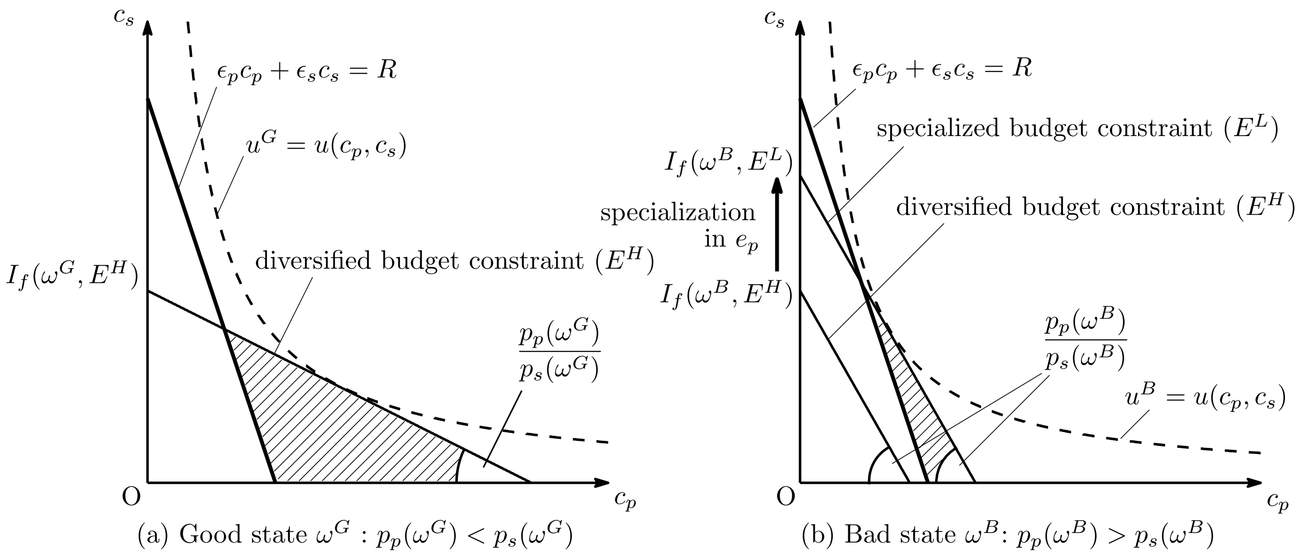


Figure 3.3: Graphical representation of poverty traps. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints

Land allocation E is determined in the first stage, and is diversified between e_p and e_s under high crop diversity E^H or specialized under low diversity E^L.

Consider two states, the good state ω^G and the bad state ω^B. In ω^G, farmers face no productivity shock ($\alpha_j(\omega) = 1$). Because the primary crop is popularly consumed, it is assumed to have a lower price than the secondary crop ($p_p(\omega^G) < p_s(\omega^G)$). Under high crop diversity E^H, Figure 3.3(a) illustrates feasible consumption bundles of poor farmers in ω^G (the shaded part), with the indifference curve and the budget and subsistence constraints depicted together (I_n is ignored for simplicity). Poor farmers satisfy the subsistence constraint under the diversified budget constraint $I_f(\omega^G, E^H)$.

However, in ω^B, an economy-wide productivity shock hits the primary crop and $\alpha_p(\omega)$ declines.[2] Because the primary crop is a necessity for households, its price elasticity is considerably low in the economy. Therefore, the economy-wide decline in the supply of the primary crop raises $p_p(\omega)$ dramatically. Because the budget constraint becomes steeper in ω^B than in ω^G, the diversified income $I_f(\omega^B, E^H)$ provides no feasible consumption bundle in Figure 3.3(b). Because endowments are allocated to the secondary crop under crop diversification, the primary crop is insufficiently self-produced for satisfying the subsistence constraint, but has too high a price for poor farmers to purchase. Therefore, allocating endowments to the secondary crop in the first stage involves the risk of violating the subsistence constraint.

Conversely, if poor farmers allocate most endowments to the primary crop (E^L), the high

[2]Damages specific to certain crops are observed in crop diseases and pests: many plant pathogens parasitize specific crop species and cannot be transmitted to others (host specificity); food preferences of some pests are narrow (oligophagy).

price of the primary crop may raise their income level from $I_f(\omega^B, E^H)$ to $I_f(\omega^B, E^L)$ even if the productivity declines, which enables them to have feasible consumption bundles in ω^B. Even if $\pi(\omega^B; h)$ is low, poor farmers prioritize producing the primary crop to satisfy the subsistence constraint in any state, which exemplifies poverty traps in the context of crop diversification.

Non-poor farmers, who have enough total endowments and non-farm income I_n, can satisfy the subsistence constraint in ω^B, regardless of endowment allocations. Because non-poor farmers maximize not the sole utility in ω^B but the expected utility in Equation (3.2.1), they do not prioritize producing the primary crop in endowment allocations if $\pi(\omega^B; h)$ is low.

Subsistence Requirement: Risk Dispersion

Subsistence requirement, in which poverty traps work previously, may conversely induce poor farmers to adopt crop diversification. Growing multiple crops enables poor farmers to disperse risks and ensures subsistence-level consumption in a bad state where, without crop diversification, they would obtain low on-farm income below the subsistence level.

In Equation (3.2.6), suppose that a productivity shock hits some of the crops randomly in the second stage. Figure 3.4(a) considers again two crops c_1 and c_2 and the good state ω^G and the bad state ω^B. For viewability, I consider a simplified form of $I_f(\omega, E) = p_1(\omega)\alpha_1(\omega)e_1 + p_2(\omega)\alpha_2(\omega)e_2$ (The essence is the same even if I introduce complementary effects). In Figure 3.4(a), if poor farmers specialize in planting the crop for c_1 (e.g., $E^L = (\bar{E}, 0)$) and they face no productivity shock in ω^G (i.e., $\alpha_1(\omega^G) = 1$), they obtain high on-farm income $I_f(\omega^G, E^L)$. Consequently, they gain the high utility $\overline{u}$ by choosing $\overline{c}$. However, if a productivity shock hits the crop for c_1 in ω^B, poor farmers obtain low income $I_f(\omega^B, E^L)$, which results in no feasible consumption bundle.

In contrast, if poor farmers diversify crop portfolios (e.g., $E^H = (\bar{E}/2, \bar{E}/2)$), they disperse risks over multiple crops and obtain stable on-farm income. Figure 3.4(a) considers a case where the on-farm income is the same in both states ($I_f(\omega^G, E^H) = I_f(\omega^B, E^H)$). Consequently, feasible consumption bundles exist in both states. To satisfy the subsistence constraint in any state, poor farmers adopt crop diversification.

In summary, if farmers specialize in specific crops and the shock does not hit these crops, they obtain high income; they obtain low income if the shock hits these crops. In contrast, farmers can disperse risks over crops by allocating endowments widely. Whatever crop the shock hits, farmers can maintain relatively stable income.

Farmers' behaviors of dispersing risks by crop diversification seem to conflict with specializing behaviors illustrated in poverty traps. However, both behaviors reflect the same motivation to preserve subsistence-level consumption in any state. Farmers may or may not

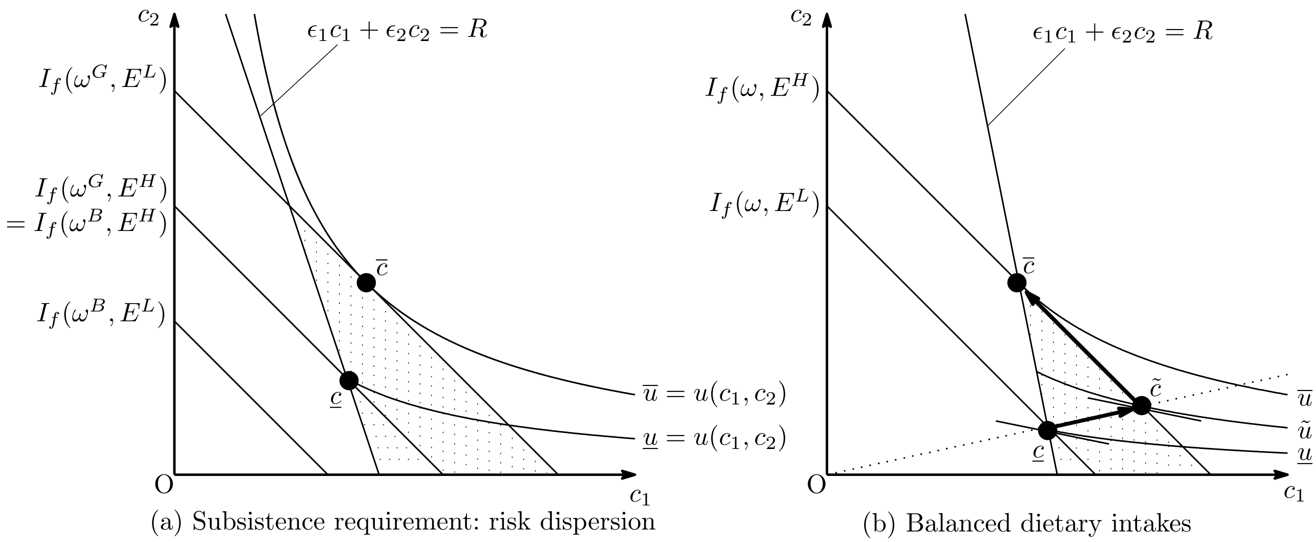


Figure 3.4: Graphical representation of risk dispersion and balanced dietary intakes. The shaded parts represent feasible consumption bundles when choosing $\underline{c}$ (horizontal dotted lines) and $\overline{c}$ (horizontal and vertical dotted lines)

diversify crop portfolios, depending on what state they recognize is possible and hence on their personal experiences h.

Balanced Dietary Intake

Crop diversification contributes to welfare improvements of poor farmers through agronomic synergies among crops. In Figure 3.4(b), the subsistence constraint is binding for $\underline{c}$ under the land allocation with low crop diversity (E^L). However, crop diversification (E^H) raises on-farm income from $I_f(\omega, E^L)$ to $I_f(\omega, E^H)$ in ω, because the productivity increases by complementary effects between crops, $d_{kl}(\omega)e_k e_l$, as shown in Equation (3.2.6). In fact, crop diversification has been empirically found to increase income, in particular among the poorest (Asfaw et al. 2018; Asfaw et al. 2019). Consequently, the welfare of poor farmers improves from the utility $\underline{u}$ realized in the consumption $\underline{c}$ to $\overline{u}$ in $\overline{c}$.

There are two channels of welfare improvements from $\underline{u}$ to $\overline{u}$. For an illustrative purpose, we suppose that farmers' preferences are homothetic and can be represented by a utility function homogeneous of degree one. As a property of homothetic preferences, the MRS, or the slope of an indifference curve, is constant along straight lines from the origin. Thus, the MRS is the same in $\underline{c}$ and $\tilde{c}$, where $\tilde{c}$ is placed in the intersection between the budget line of $I_f(\omega, E^H)$ and the straight dotted line through $\underline{c}$ from the origin.

In the first channel, welfare improvements arise from $\underline{u}$ to $\tilde{u}$. Because on-farm income rises from $I_f(\omega, E^L)$ to $I_f(\omega, E^H)$, farmers' consumption expands from $\underline{c}$ to $\tilde{c}$, while maintaining the proportion of c_1 to c_2. This improvement is a direct income effect caused by crop diversification. In the second channel, farmers can further improve welfare by moving from $\tilde{c}$ to $\bar{c}$. Increased income by crop diversification enlarges a set of feasible consumption bundles by the triangle $\triangle\bar{c}\underline{c}\tilde{c}$, which enables farmers to take an MRS closer to the slope of the budget line than in $\tilde{c}$. In $\tilde{c}$, farmers continue to prioritize consuming c_1 relative to c_2, as in $\underline{c}$. Once the subsistence constraint is weakened by the income increase, farmers with the convex preference prefer more balanced dietary consumption $\bar{c}$ to $\tilde{c}$.

Importantly, this indirect income effect in the second channel arises only if the subsistence constraint is binding for the consumption before the income increase ($\underline{c}$). As shown in Equation (3.2.5), λ_2 equals zero if the subsistence constraint is slack, which generates no distortion of MRS. Therefore, the desire for balanced dietary intakes works as a motivation for crop diversification among poor farmers, rather than non-poor farmers.

Stabilization of Market Income

Unlike poor farmers, the subsistence constraint is unlikely to be binding or violated among non-poor farmers. If the subsistence constraint is slack, λ_2 in Equation (3.2.5) equals zero. Hence, non-food consumption q (and possibly luxury food with zero subsistence weights) are more important among non-poor farmers.

Suppose that non-poor farmers consume q by using market income obtained from the difference between produced crops and self-consumed ones (they do not produce q by themselves). Then, crop diversification is appealing for non-poor farmers with risk-averse preferences if it stabilizes market income necessary for consuming q.

Stabilizing market income may also be important for poor farmers, if they sell harvests to purchase food in markets. However, non-poor farmers depend in their utility more on consumption unnecessary for preserving the subsistence level. Therefore, non-poor farmers seem to have stronger incentives to stabilize market income by crop diversification.

Summary

Table 3.1 summarizes how the three potential factors affect decisions on crop diversification differently between poor and non-poor farmers ("+" and "−" indicate positive and negative impacts, respectively). Table 3.1 shows that the effects of the three potential factors overlap or conflict with each other. Therefore, it is inconclusive in theoretical perspectives how farmers' experiences of each shock affect decisions on crop diversification. For example, poor farmers may perceive drought as a devastating shock for their livelihoods from their experience, which may induce them to plant drought-tolerant crops for risk dispersion. Conversely, poor farmers who obtained little income owing to past drought may devote their limited productive resources to producing staple crops for survival. Empirical results depend on farmers' recognition of possible states and future paths of productive resources after shocks.

Table 3.1: Three Potential Factors Behind Crop Diversification between Poor and Non-Poor Farmers

	Subsistence Requirement		Balanced Dietary Intake	Stabilization of Market Income
	Poverty Traps	Risk Dispersion		
Poor Farmers	$-$	$+$	$+$	
Non-Poor Farmers				$+$

Note: "+" means a potential factor positively affects farmers' adoption of crop diversification; "−" means its negative impact.

3.3 Dataset and Estimation Strategy

3.3.1 Dataset Description

This chapter exploits the Tanzanian National Panel Survey (NPS) of the World Bank's Living Standards Measurement Study - Integrated Survey on Agriculture (LSMS-ISA). The NPS data comprise agricultural and socioeconomic information from nationally representative households nationwide.

Indicators on Crop Diversification and its Regional and Temporal Standardization

My analysis relies on two indicators of crop diversification: the simple counting of crop numbers and the Simpson index (SI). For simple counting, I sum the number of planted crops in farmland for each farmer. For the SI, I calculate the following indicator for each farmer:

$$SI = 1 - \sum_{c} s_c^2 \tag{3.3.1}$$

where s_c is the area share of crop c in total farmland. The SI is large if a farmer plants many crops evenly on the farmland. Conversely, we have $SI = 0$, or complete specialization, if only one crop is planted on the whole land. The SI values the evenness of land shares of planted crops as diversity (Duelli and Obrist 2003). Whereas the indicator of crop numbers simply counts all crops equally, the SI becomes low if a certain crop dominates farmland. Appendix 3-II illustrates how different meanings of "diversity" in the two indicators illuminate farmers' various motivations for crop diversification.

The definition of the SI has a close relationship with crop diversity in the Conceptual Framework. In the Conceptual Framework, I represent crop diversity by land allocations to each crop, as shown in Equation (3.2.6). In Figure 3.4(a), the effects of the risk dispersion in the subsistence requirement depend on how much farmers can stabilize on-farm income by allocating land evenly to each crop, as well as by just planting multiple crops. In Figure 3.4(a), the effects of the balanced dietary intakes depend on the degree in which farmers plant different crops over the same land. Thus, using the SI as well as crop numbers is helpful in terms of the link between the theoretical and empirical parts.

To control for regional and temporal heterogeneity of crop diversity, I further standardize the two indicators by region and year:

$$\frac{CD_{ijkt} - \overline{CD}_{jkt}}{sd_{jkt}} \tag{3.3.2}$$

where CD_{ijkt} is an indicator k of crop diversification for farmer i in region j in year t, and $\overline{CD}_{jkt}$ and sd_{jkt} are its average and standard deviation for each region and year, respectively.

I take this regional and temporal standardization because agro-ecological conditions such as soil fertility, rainfall, and temperature vary widely among regions every year. Because regional dummies cannot be included in my threshold model owing to computational problems in estimation, I take the standardization in Equation (3.3.2) to control for regional and temporal heterogeneity of crop diversity.[3]

Food Consumption as a Wealth Indicator

This chapter distinguishes poor and non-poor farmers by food consumption levels. As illustrated in Section 3.2, food consumption levels directly determine whether the subsistence constraint is binding or slack.[4]

Food consumption levels are based on household food consumption aggregates in the NPS, which are reported as the amount per 28 days in Tanzanian shillings.[5] They include all the monetary values of food consumed by a household: food that was produced by the household, purchased in markets, eaten outside home, and received as a gift. Using this data, I calculate food consumption values per adult equivalent for each household.[6]

Shock Variables

Shock variables comprise six agronomic and market shocks over the past five years before cultivation: drought/flood, crop disease/pests, large falls in crop prices for sale, large increases in food prices for purchase, large increases in agricultural input prices, and livestock loss.

Shock variables are dummy variables constructed from the questionnaire that asks, “Over the past five years, was your household severely affected negatively by any of the following events?” Thus, shock variables may be subjective in that farmers may list a shock in the

[3] Michler and Josephson (2017) had the same concern. They used a total crop number grown in each village and year to eliminate heterogeneity.

[4] In the robustness check later, asset levels will be used as another wealth indicator.

[5] Spatial and temporal price differences within each survey wave and across waves are adjusted using Fisher price indices.

[6] My food consumption data are based on monetary values to reflect farmers’ affordability to consume food. Luxury food (e.g., tea) may have small kilograms and calories but large monetary values in markets. I discuss relations to caloric consumption in Appendix 3-VII.

questionnaire when they felt a threat to their livelihoods. For example, poor farmers with vulnerable food securities may easily feel threatened by a rise in food price and tend to list it as a past shock in the questionnaire. Nevertheless, self-reported shock variables may be favorable because farmers' adoption of crop diversification depends on subjective expectations for future states in my conceptual framework. Moreover, I find no significant difference for all shock variables between poor and non-poor farmers in the summary statistics later.

The questionnaire also provides information on the timing of shocks. To avoid the reverse causality that crop diversity prevents agricultural shocks, this chapter sets the reference period of shock variables as exactly five years before the *cultivation*, not before the *interview*.[7]

Dataset for Analysis

I restrict my dataset to farmers in Tanzania's mainland and focus on farmers' activities in the long rainy season (*masika*), the main cultivation season from March to May in Tanzania. I also exclude households that split off from the original households in past waves owing to marriage or migration, because it is uncertain whether they inherited the unobserved characteristics of their original households.

There are five longitudinal waves from Wave 1 (2008/09) to Wave 5 (2018/19) available in the NPS. Of the five waves, this chapter constructs a balanced panel of Wave 2 (2010/11) and Wave 3 (2012/13). As will be discussed, I use food consumption data in the preceding, not contemporary, wave. Thus, the first wave is used to obtain preceding data on food consumption. Waves 4 (2014/15) and 5 (2018/19) are difficult to use owing to the discontinuous structure of the NPS. The NPS refreshed the sampled households in Wave 4. Then, there are only 314 farmers who had the corresponding data in Waves 2 and 3 under the aforementioned data limitation. Because Hansen's threshold model used in this chapter requires a balanced panel, including Waves 4 and 5 would cause large attrition over waves. Moreover, the information on the timing of shocks is not provided after Wave 4, which makes it difficult to avoid the reverse causality that crop diversity prevents agricultural shocks.

In Wave 2, there are 1,632 farmers who cultivate crops in the long rainy season, who reside in the mainland, and who have preceding data on food consumption in Wave 1. Of these farmers, 1,533 farmers were re-interviewed in Wave 3 (split-off households between Waves 2 and 3 are excluded), and 1,336 farmers continued to cultivate crops in Wave 3. By eliminating farmers with missing data, I consequently obtain a balanced panel of 1,332 households in each wave.

[7] Because the NPS interval between waves is two years, some may recommend the two-years reference period. Theoretically, time-demeaning treatment in fixed-effects methods just eliminates households' time-constant heterogeneity; definitions of variables seem unessential to obtain consistent estimates in fixed-effects methods. However, even though I restrict the reference period in Wave 3 to two years for robustness, I still obtain similar findings.

In the robustness check later, I will tackle the limited sample size of my dataset by allowing an unbalanced panel of all waves and including split-off households. I will also incorporate the short rainy season, the secondary cropping season in Tanzania, into my analysis as the robustness check.

3.3.2 Threshold Regression Model in Non-dynamic Panel

To estimate farmers' heterogeneous behaviors on crop diversification in panel data, I rely on Hansen's (1999) fixed-effects threshold model:

$$SCD_{ikt} = \begin{cases} \alpha_{ik} + x'_{it}\beta_{Lk} + s'_{it}\gamma_{Lk} + \epsilon_{ikt} \text{ if } C_{i,t-1} < \lambda_k \\ \alpha_{ik} + x'_{it}\beta_{Uk} + s'_{it}\gamma_{Uk} + \epsilon_{ikt} \text{ if } C_{i,t-1} \geq \lambda_k \end{cases} \tag{3.3.3}$$

for farmer i in year t. The dependent variable SCD_{ikt} is an indicator k of crop diversification, standardized in each region and year, as in Equation (3.3.2). In the set of covariates, x_{it} is a column vector of household characteristics; s_{it} is a column vector of dummy variables for experiencing each of the six agricultural shocks over the past five years before the cultivation season. The error term ϵ_{ikt} is independently and identically distributed with mean zero and finite variance. The individual-specific effects α_{ik} are introduced. Because I estimate Equation (3.3.3) separately by indicator type k, coefficients $(\beta_{Lk}, \gamma_{Lk}, \beta_{Uk}, \gamma_{Uk})$, the error term ϵ_{ikt}, and the individual-specific effects α_{ik} differ by k.

Equation (3.3.3) splits all farmers into two groups by comparing the magnitudes between the threshold variable $C_{i,t-1}$ and the threshold parameter λ_k (λ_k differs between the two indicators). The variable $C_{i,t-1}$ is the level of farmers' food consumption in the preceding wave, and λ_k is the unobserved breakpoint behind the regime shift. If $C_{i,t-1} < \lambda_k$, farmer i belongs to the lower regime with coefficients β_{Lk} and γ_{Lk}. If $C_{i,t-1} \geq \lambda_k$, farmer i belongs to the upper regime with β_{Uk} and γ_{Uk}. This chapter aims to estimate the threshold λ_k and investigate the difference in γ_{Lk} and γ_{Uk} for each indicator k.

The regime structure of threshold models may fit my conceptual framework. In food consumption levels below the threshold, the subsistence constraint affects farmers' behaviors. Once farmers consume food above the threshold, the constraint has no effect, however high the food consumption levels become. In threshold models, coefficients remain constant unless the regime switches. This regime structure is not captured in interaction terms between shock variables and the *continuous* food consumption variable. If interaction terms are used, food consumption levels must continue to change farmers' behaviors uniformly over all consumption ranges (even above the threshold).[8]

[8] If I use interaction terms, most targeted coefficients are insignificant, though their signs coincide with those in the threshold model.

Furthermore, the regime structure of threshold models is in line with previous studies. Makate et al. (2022) analyzed farmers' heterogeneous behaviors by adding interaction terms between shock variables and two *dummy* (not *continuous*) variables representing poverty. The two dummy variables are equal to one if farmers' asset levels and farm size fall in the bottom 40% of the sample distribution, respectively. In contrast, Makate et al. (2023) split samples into three quantile subgroups (low, medium, high) using the same wealth data and conducted estimation for each group. These studies also adopted the regime structure for estimation, although "thresholds" were specified in advance by these studies.

Fixed-effects Estimation of Coefficients

Equation (3.3.3) can be rewritten as a single equation:

$$\begin{aligned} SCD_{ikt} &= \alpha_{ik} + I(C_{i,t-1} < \lambda_k)(x'_{it}\beta_{Lk} + s'_{it}\gamma_{Lk}) + I(C_{i,t-1} \geq \lambda_k)(x'_{it}\beta_{Uk} + s'_{it}\gamma_{Uk}) + \epsilon_{ikt} \\ &= \alpha_{ik} + X_{it}\beta_k + \epsilon_{ikt}, \end{aligned} \tag{3.3.4}$$

where $\beta_k = [\beta'_{Lk}, \gamma'_{Lk}, \beta'_{Uk}, \gamma'_{Uk}]'$ and $X_{it} = [I(C_{i,t-1} < \lambda_k)(x'_{it}, s'_{it}), I(C_{i,t-1} \geq \lambda_k)(x'_{it}, s'_{it})]$ with the indicator function $I(\cdot)$. If the threshold λ_k is known, the estimator for β_k is obtained as the fixed-effects estimator by subtracting each SCD_{ikt} and X_{it} by the time-averaged values and conducting the ordinary least squares in the time-demeaned data. I choose the estimated threshold for λ_k from the actual data of $C_{i,t-1}$ so that it minimizes the sum of squared residuals (SSR) in the time-demeaned form of Equation (3.3.4).[9]

Reverse Causality Between Food Consumption and Crop Diversification

In Equation (3.3.4), X_{it} contains food consumption data inside the indicator function $I(\cdot)$. Because the food consumption data is constructed from the information over the week prior to the interview, the timing of C_{it} chronologically follows the cultivation season in each wave. If I used C_{it} as the threshold variable, the estimated coefficients would reflect positive causal impacts of crop diversification on current food consumption.

To avoid the reverse causality, I use the past food consumption $C_{i,t-1}$ as the threshold variable. Although recent literature on threshold models has addressed the endogeneity of threshold variables (e.g., cross-sectional models by Kourtellos et al. (2016) and Yu and Phillips (2018); dynamic panel models by Seo and Shin (2016) and Gørgens and Würtz (2019)), there is limited accumulation in static panel models. Because using a lagged threshold variable is common in empirical corporate finance (Seo and Shin 2016), I follow this practice.

[9] I avoid choosing extreme values of threshold parameters by eliminating the smallest and largest few percents of food consumption data as candidates. To avoid arbitrariness, the excluded percents are set as two percents in crop numbers and one percent in the SI throughout this chapter.

In the conceptual framework, poor farmers who own low amounts of the total endowment $\bar{E}$ in the first stage must determine the land allocation carefully because the subsistence constraint tends to bind when the production is realized in the second stage. Although the food consumption level in the first stage is not incorporated in the model, the total amount of endowments, which farmers can allocate to the production in the second stage, is closely linked to farmers' economic levels in the first stage, which justifies the use of past food consumption levels as a threshold variable.

Linearity Test of Regime Shift

The existence of a regime shift is tested for each indicator k. The null hypothesis is H_0 : $(\beta_{Lk}, \gamma_{Lk}) = (\beta_{Uk}, \gamma_{Uk})$, which supports a linear model without the regime shift, whereas the alternative hypothesis is $H_1 : (\beta_{Lk}, \gamma_{Lk}) \neq (\beta_{Uk}, \gamma_{Uk})$, which supports a threshold model with a regime shift. In this linearity test, the F-statistic follows a nonstandard asymptotic distribution. Thus, its test relies on bootstrapping procedures (Hansen 1996).

Covariates for Household Characteristics

Household characteristics x_{it} in Equation (3.3.3) include age, gender, and years of education of the household head. Adult equivalents are also included as a household-size indicator. I also add the total acres of owned land as farmers' agricultural production scales.

Easy access to markets enables farmers to purchase seeds and seedlings. Thus, I add to x_{it} the distance from each household to the nearest agricultural market. I also add the distance to the district headquarter because opportunities for non-farm activities may affect farmers' on-farm strategies. Cooperatives may provide access to agricultural inputs. Moreover, saving and credit cooperatives (SACCO) may alleviate credit constraints on agricultural investments. Thus, I introduce two dummy variables that equal one if a farmer belongs to a cooperative and a SACCO, respectively. Finally, I add the average rainfall in the long rainy season over the past five years in each district because farmers choose crops that are environmentally adaptable to the expected level of rainfall.

3.3.3 Descriptive Statistics

Estimated Thresholds for Each Indicator of Crop Diversification

Table 3.2 shows estimated thresholds for each indicator of crop diversification and the numbers and proportions of households with per capita food consumption below and above the threshold in each wave. Both estimated thresholds are around 200,000 in Tanzanian shillings. The proportions of farmers below each threshold (lower regime) are roughly 15%.

Table 3.2: Estimated Threshold Parameter and Number and Proportion of Households in Lower and Upper Regimes in Each Wave (food consumption levels per adult equivalent per 28 days in Tanzanian shillings)

		Wave 2 (2010/11)		Wave 3 (2012/13)	
Type of Indicator	Estimated Threshold	Lower Regime	Upper Regime	Lower Regime	Upper Regime
Simple Count of Crop Number	206,276	211 (0.16)	1,121 (0.84)	186 (0.14)	1,146 (0.86)
Simpson Index	204,938	209 (0.16)	1,123 (0.84)	185 (0.14)	1,147 (0.84)

Note: There are 1,332 households in each wave. The proportion of households to the total is in parenthesis.

Figure 3.5 shows the sample distribution of food consumption per adult equivalent below the 99th percentile. The two thresholds are shown in the red solid line for crop numbers and the blue dotted line for the SI. As will be discussed later, farmers below the estimated thresholds suffer from poor energy intakes. Around 80% of them consume calories below 2,300 kcal, a poverty line set by Dercon et al. (2009). The estimated thresholds seem to reflect subsistence requirement.

Crop Diversity in Farmland in the Two Regimes

To compare crop diversity between the two regimes, I classify all crops into five broad types: staples, sorghum/tubers/roots, legumes/oil, fruits/vegetables, and cash crops. Appendix 3-III enumerates all crop items included in each type. Figure 3.6 illustrates the proportion of farmers planting each crop type on the left-hand side, and the average land shares of each crop type on the right-hand side. Farmers are separated into two regimes by the threshold for crop numbers on the left and for the SI on the right.[10]

Surprisingly, farmers' crop diversity seems quite similar in both regimes and does not change by the food consumption level itself. In Figure 3.6, most farmers plant crops and allocate large land shares to staples, which comprise main foods such as maize and rice. This is followed by legumes/oil (a variety of beans) and sorghum/tubers/roots (e.g., millet and potatoes). High-value crops of fruits/vegetables and cash crops (e.g., cotton and tobacco) are limited in planting proportions and land shares, although lower-regime farmers have slightly higher proportions and land shares of cash crops than upper-regime farmers.

Summary Statistics

Table 3.3 displays mean values of variables in each regime; farmers are divided by the threshold for crop numbers in the left-hand side and by the threshold for the SI in the right-hand

[10] As a way of crop diversification, it is uncertain whether farmers plant different crops within one agricultural plot or plant different crops over multiple plots, with only one crop planted in each plot. I examine the number of planted crops in each plot by constructing the plot-level dataset corresponding to the household-level dataset in this chapter. Then, only one crop is planted in 15.3% of the plot-level dataset; farmers plant more than one crop in the rest of the plots. Thus, it seems prevalent that farmers plant different crops within one agricultural plot.

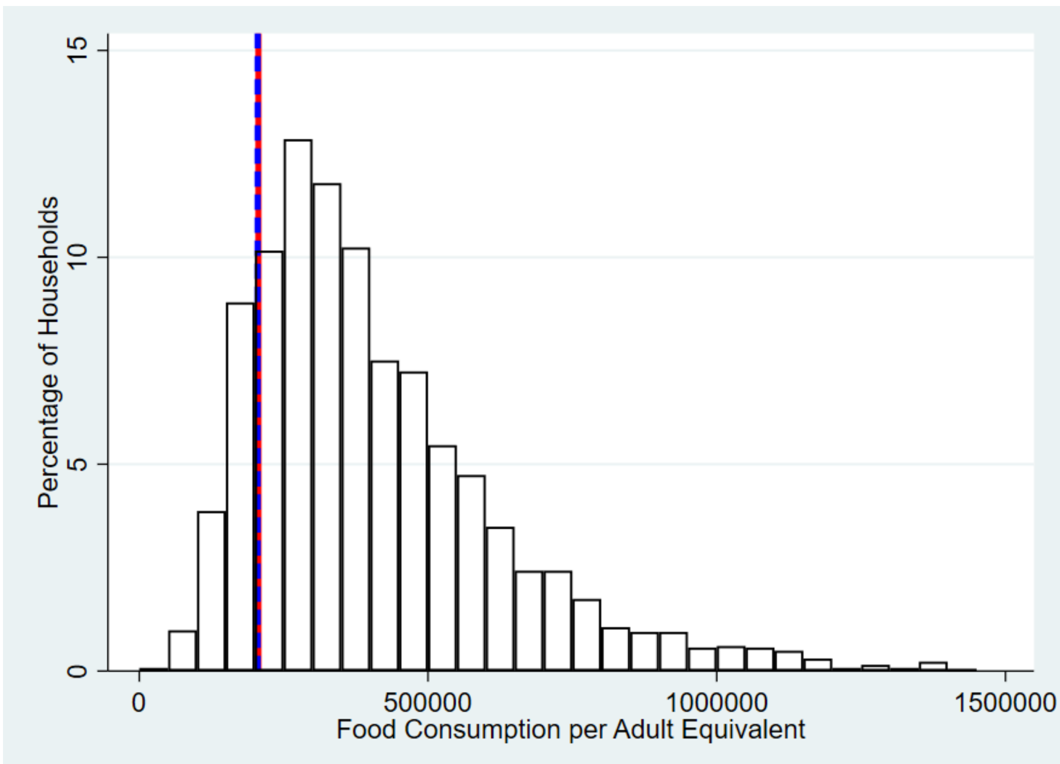


Figure 3.5: Sample Distribution of Food Consumption Per Adult Equivalent Per 28 Days (Tsh) below the 99 Percentile with Each Threshold for Crop Numbers (Red Solid Line) and the Simpson Index (Blue Dotted-Line)

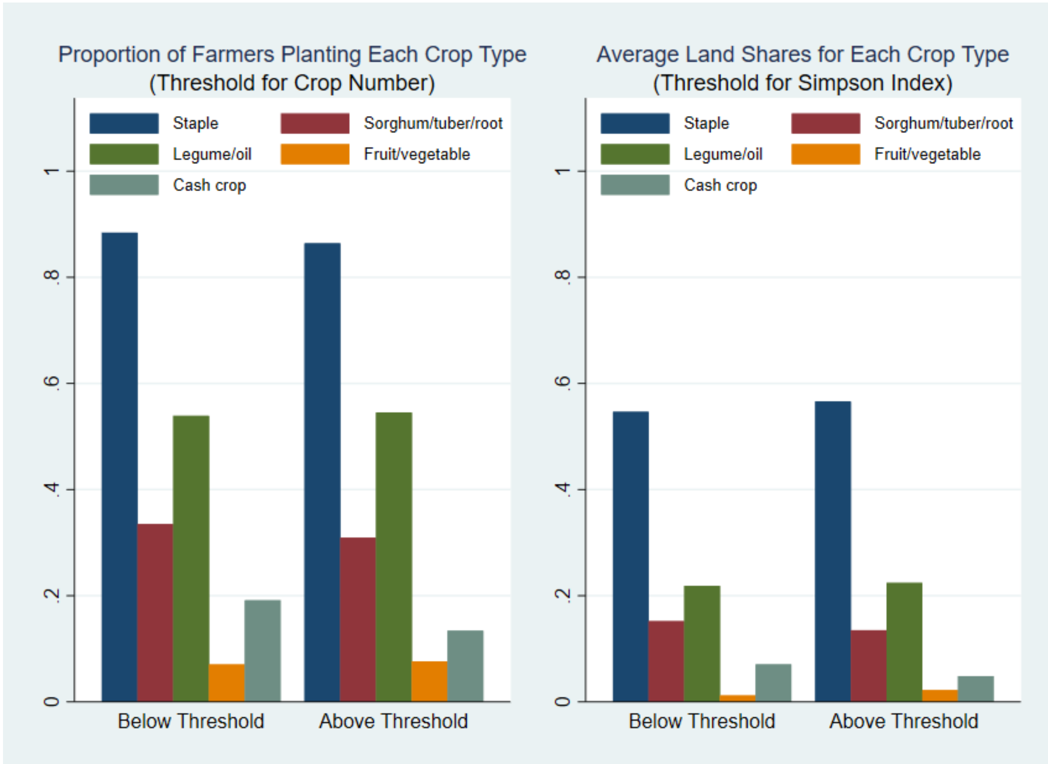


Figure 3.6: Crop Diversity of Farmers in the Lower and Upper Regimes

side. Mean differences are tested between the two regimes via t-tests in each side.

In Table 3.3, the two indicators of crop diversification, which are not standardized in region and year, are close in value between the two regimes. Similar crop diversity in the two regimes is consistent with that in Figure 3.6. Obviously, lower-regime farmers consume less in the preceding wave than upper-regime farmers. Household heads are more educated and have smaller adult equivalents in the upper regime than in the lower regime. I find no significant difference in acres of owned land, the distance to agricultural markets and district headquarters, and the belonging to cooperatives between the two regimes. Upper-regime farmers are more likely to belong to a SACCO.

Upper-regime farmers reside in environmentally advantaged areas with high rainfall, compared with lower-regime farmers. I find no significant difference in all shock variables between the two regimes and no evidence of subjective biases on shock variables. To check the bias associated with farmers' self-reporting, I also conduct the fixed-effects regression of each shock variable on a dummy variable that are equal to one if farmers belong to the lower regime. Although the results are not reported here, I still find no significant association between all shock experiences and the regime shift.

Table 3.3: Summary Statistics for Farmers in the Lower and Upper Regimes Divided by the Thresholds for Each Indicator of Crop Diversification

Mean Values of Variables	Crop Number: Lower Regime	Crop Number: Upper Regime	Simpson Index: Lower Regime	Simpson Index: Upper Regime
Indicators of crop diversification:				
Simple counting of crop numbers	2.40	2.34	2.40	2.34
Simpson Index	0.39	0.39	0.40	0.39
Threshold variable:				
Food consumption per HH member in preceding wave	159,717***	467,488	159,370***	467,142
Household characteristics:				
Age of household head	51.47	50.18	51.49	50.18
Male-headed household (dummy)	0.77	0.78	0.77	0.78
Years of education of household head	4.01***	4.80	4.02***	4.80
Adult equivalents	5.43***	4.50	5.43***	4.50
Acres of owned land	6.38	7.68	6.40	7.68
Distance to nearest agricultural market (km)	81.55	78.67	81.65	78.66
Distance to district headquarter (km)	101.21	102.77	101.43	102.73
Belongs to cooperative (dummy)	0.13	0.13	0.13	0.13
Belongs to SACCO (dummy)	0.04**	0.07	0.04**	0.07
Regional characteristics:				
Avg. *masika*-rainfall over past five years (mm)	356.84***	387.84	357.12***	387.75
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.23	0.22	0.23	0.22
Experienced crop disease/pests (dummy)	0.12	0.14	0.12	0.14
Experienced large fall in sale prices for crops (dummy)	0.12	0.14	0.12	0.14
Experienced large rise in price of food (dummy)	0.26	0.24	0.26	0.24
Experienced large rise in agricultural input prices (dummy)	0.14	0.14	0.14	0.14
Experienced livestock death or theft (dummy)	0.10	0.08	0.09	0.08
Observations	397	2,267	394	2,270

Note: Mean differences are tested between two regimes by year at significance levels of * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

3.4 Results

Table 3.4 displays the results of the Hansen's fixed-effects threshold model and the traditional fixed-effects (FE) model without a threshold, where standardized indicators of crop diversification, crop number and the SI, are regressed on household characteristics and past agricultural shocks. Standard errors are clustered at the household level. Because the two indicators are standardized, a one unit increase of a covariate changes an indicator of crop diversification by $\hat{\beta}_k$ times its standard deviation (SD) for each region and year, sd_{jkt} in Equation (3.3.2). As indicated as "log" in parenthesis in Table 3.4, some covariates with large values are transformed into logarithmic form. To conserve space, Appendix 3-IV sees the estimated effects of household and regional characteristics on crop diversification in Table 3.4.

The results of the linearity test using bootstrapping with 500 repetitions (the F-statistics and their p-values) are shown in Columns (1) and (4) at the end of the table. The linearity tests significantly support the threshold model for both crop numbers and SI, which implies that the regime shift in the threshold of food consumption fits the reality where the data were observed.

Crop Number

Column (1) of Table 3.4 indicates that lower-regime farmers who experienced drought/flood significantly increase crop numbers. For simplicity, I consider the SD of crop numbers among lower-regime farmers in the entire dataset (not in each region and year). Because it is 1.18, experiencing drought/flood induces lower-regime farmers to increase crop number by 0.33 ($= 0.28 \times 1.18$) on "average." As Table 3.1 suggests, lower-regime farmers who experienced drought/flood may perceive it as a devastating shock for their livelihoods and introduce climatically resistant crops to disperse risks so that they can preserve subsistence levels in case the shock occurs again.

This finding is contrary to Makate et al. (2022) and Makate et al. (2023), who found that the rainfall shortage one year before the growing season discouraged crop diversification among poor farmers in Ethiopia and encouraged it among wealthy farmers in Tanzania and Malawi. Because my shock variables consider the five-years interval before the growing season, these opposite results may be explained by the different reference periods of shock variables. In fact, Makate's studies showed that the statistical significance disappeared in two-years lagged rainfall shortage (Makate et al. 2022), and that crop diversity was positively associated with the long term rainfall variability (1980-2018) even among poor farmers (Makate et al. 2023). Drought may reduce crop diversity among poor farmers in the short run, but may increase it in the long run.

Table 3.4: Estimates of Indicator of Crop Diversification on Household Characteristics and Past Experiences of Agricultural Shocks

	Crop Number			Simpson Index		
	Threshold Model		FE	Threshold Model		FE
Explanatory Variable	(1) Lower Regime	(2) Upper Regime	(3) All Sample	(4) Lower Regime	(5) Upper Regime	(6) All Sample
Household and regional characteristics:						
Age of household head	-0.04*	-0.04	-0.04	-0.04	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Years of education of household head	0.09***	0.02	0.03	0.08***	0.03	0.04*
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Adult equivalents	-0.07**	-0.05*	-0.05*	-0.07**	-0.04*	-0.04*
	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Acres of owned land	0.00	-0.00	-0.00	0.02*	-0.00	-0.00
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Distance to nearest agricultural market (km) (log)	0.14	0.30*	0.31*	-0.13	0.05	0.07
	(0.19)	(0.17)	(0.16)	(0.21)	(0.19)	(0.19)
Distance to district headquarter (km) (log)	0.03	0.01	0.01	0.02	0.01	0.01
	(0.07)	(0.03)	(0.02)	(0.06)	(0.03)	(0.03)
Belongs to cooperative (dummy)	0.07	0.15**	0.13*	0.06	0.11	0.10
	(0.18)	(0.08)	(0.07)	(0.18)	(0.08)	(0.07)
Belongs to SACCO (dummy)	0.46	0.18*	0.21*	0.12	0.13	0.11
	(0.42)	(0.11)	(0.12)	(0.34)	(0.11)	(0.11)
Avg. *masika*-rainfall over past five years (mm) (log)	1.20***	1.09**	1.11**	1.03*	0.96*	0.92*
	(0.46)	(0.46)	(0.46)	(0.56)	(0.56)	(0.55)
In Wave 3 (2012/13) (dummy)	0.44***	0.03	0.09	0.46***	0.02	0.08
	(0.16)	(0.08)	(0.07)	(0.16)	(0.08)	(0.08)
Agricultural shocks over past five years:						
Experienced drought/flood (dummy)	0.28*	0.02	0.06	0.38***	-0.00	0.06
	(0.15)	(0.07)	(0.07)	(0.15)	(0.07)	(0.06)
Experienced crop disease/pests (dummy)	-0.21	0.10	0.07	-0.20	0.07	0.05
	(0.17)	(0.08)	(0.07)	(0.18)	(0.08)	(0.07)
Experienced large fall in sale prices for crops (dummy)	-0.06	0.13*	0.10	0.05	0.04	0.04
	(0.14)	(0.08)	(0.07)	(0.17)	(0.08)	(0.07)
Experienced large rise in price of food (dummy)	0.25*	0.06	0.10*	0.37**	0.03	0.08
	(0.14)	(0.06)	(0.05)	(0.16)	(0.06)	(0.06)
Experienced large rise in agricultural input prices (dummy)	-0.04	0.20**	0.19***	-0.17	0.09	0.08
	(0.15)	(0.08)	(0.07)	(0.16)	(0.07)	(0.07)
Experienced livestock death or theft (dummy)	-0.17	0.07	0.04	-0.35*	0.04	-0.01
	(0.21)	(0.10)	(0.09)	(0.20)	(0.09)	(0.09)
Linearity test						
F statistic (p-value in parenthesis)	76.71*			76.29*		
	(0.07)			(0.08)		

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level.

Column (1) also indicates that experiencing a large rise in food price significantly induces lower-regime farmers to increase crop numbers by 0.30 ($= 0.25 \times 1.18$) on "average." The rise in food prices may easily threaten the vulnerable food security of lower-regime farmers. As Table 3.1 suggests, lower-regime farmers may grow multiple crops for their self-consumption to disperse market risks and to secure balanced dietary intakes.

In Column (2), upper-regime farmers who experienced a large price fall in output markets and a large price rise in input markets significantly increase crop numbers. Because the SD is 1.16 in the whole upper-regime farmers, the increased crop numbers are 0.15 ($= 0.13 \times 1.16$) and 0.23 ($= 0.20 \times 1.16$) on "average," respectively. Because these two market shocks are closely related to market income, upper-regime farmers may have strong incentives to stabilize market income by crop diversification, as Table 3.1 suggests. Coromaldi et al. (2015) and Asfaw et al. (2018) also found that shocks of output and input prices promote crop diversification, though households are not separated by wealth levels.

In the traditional FE estimation of Column (3), I find significantly positive coefficients for rises in food and input prices, as in Columns (1) and (2). However, I find no statistical significance for drought/flood and a fall in crop price. Thus, the FE estimation may provide intermediate results between the two regimes and conceal different behaviors on crop diversification between poor and non-poor farmers.

Simpson Index

Column (4) shows that lower-regime farmers who experienced drought/flood and a large rise in food price significantly increase the SI, as in Column (1). Because the SD of the SI among all lower-regime farmers is 0.26, experiencing either shock induces lower-regime farmers to increase the SI by 0.10 ($= 0.38/0.37 \times 0.26$) on "average." In Column (5), the coefficients of experiencing a fall in crop price and an increase in input price are insignificant in upper-regime farmers, although they are positive.

The key finding in the SI's analysis is past livestock loss significantly decreases the SI among lower-regime farmers by 0.09 ($= 0.35 \times 0.26$) "on average." Negative impacts of livestock loss were also found in Makate et al. (2022) (though not examined separately by wealth levels). The loss of livestock as productive assets may damage the capability for lower-regime farmers to plow land or plant new seeds. However, they may not afford to repurchase it, which may force lower-regime farmers to allocate limited agricultural resources to growing staple crops, as Table 3.1 suggests.

In the FE estimation in Column (6), I find no statistical significance for all shock variables. Because the FE estimation is based on the entire dataset without sample division, it may ignore the minority of lower-regime farmers (see Table 3.2).

3.5 Discussion

3.5.1 Crop-Specific Analysis

This section examines what crop farmers decide to cultivate in response to each agricultural shock. In the previous results, index-based measurements of crop diversification, crop numbers and the SI, consider all crops uniformly, regardless of the differences in agronomic traits of each crop (Maggio and Sitko 2021).

This section classifies all crops into five crop types in terms of crop-specific characteristics, as in Figure 3.6: staples as main foods and dietary sources of carbohydrates, sorghum/roots/tubers as low-valued but drought-resistant crops, legumes as dietary sources of protein and nitrogenous suppliers to the soil, vegetables/fruits as dietary sources of minerals and vitamins, and cash crops as income sources. In Appendix 3-V, the two indicators of crop diversification are reconstructed using the five crop types. The results in Appendix 3-V are similar to the main results, suggesting that crop diversification mainly arises between different crop types.

Based on the five crop types, I consider five categories of farmers, according to their cropping in farmland. The first category is farmers who cultivate only staple crops in their farmland. As Figure 3.6 shows, planting staples is highly common among farmers. In this category, farmers specialize in the production of staple crops. The other four categories are farmers who cultivate not only staples but also each one of the other four crop types (sorghum/roots/tubers, legumes/oil, vegetables/fruits, and cash crops). Farmers in these four categories may diversify crop portfolios to utilize crop-specific characteristics of each crop type. This categorization of farmers is analogous to Maggio and Sitko (2021), who set maize mono-cropping as the reference cropping system and investigated how Malawian and Zambian farmers deviate from the maize mono-cropping to intercropping with legumes, non-maize staples, and cash crops.

Using the Hansen's fixed-effects threshold model, I estimate in Table 3.5 the probability that farmers fall into each category. Household and regional characteristics are included in the analysis but omitted from the table. All estimated threshold parameters are relatively close to those in my main analysis (around 200,000 in Table 3.2). Thus, Table 3.5 seems helpful to interpret my main results, though F-statistics are insignificant in three crop types.

Table 3.5 provides three findings related to my main results in Table 3.4. First, farmers' specialization in the production of staple crops in Columns (1)-(2) contrasts with results in Table 3.4. I find significantly negative coefficients for a rise in food price in Column (1) and for a rise in input price in Column (2). Crop diversification may proceed as deviations from the mono-cropping of staple crops, if lower-regime farmers recognize risks of food insecurities and if upper-regime farmers recognize risks of market shocks. Conversely, I find a significantly

Table 3.5: Estimates of Probability of Cultivating Each Crop Type on Shock Variables

	Only Staple		Staple & Sorghum/Root/Tuber		Staple & Legume		Staple & Vegetable/Fruit		Staple & Cash crop	
Explanatory Variable	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime	(5) Lower Regime	(6) Upper Regime	(7) Lower Regime	(8) Upper Regime	(9) Lower Regime	(10) Upper Regime
Shock Variables:										
Experienced drought/flood (dummy)	-0.08	0.02	0.28***	0.00	0.17*	-0.02	-0.00	0.03	-0.06	0.01
	(0.10)	(0.02)	(0.08)	(0.03)	(0.09)	(0.03)	(0.04)	(0.02)	(0.06)	(0.02)
Experienced crop disease/pests (dummy)	-0.03	-0.03	0.19**	0.01	-0.10	0.06	0.04	0.00	-0.06	0.02
	(0.08)	(0.03)	(0.09)	(0.03)	(0.11)	(0.04)	(0.04)	(0.02)	(0.07)	(0.02)
Experienced large fall in sale prices for crops (dummy)	-0.31**	-0.00	0.16	0.01	-0.02	0.06*	0.01	0.03	-0.01	-0.00
	(0.14)	(0.03)	(0.10)	(0.03)	(0.11)	(0.03)	(0.05)	(0.02)	(0.08)	(0.02)
Experienced large rise in price of food (dummy)	-0.22*	-0.04*	0.05	0.04	0.16*	-0.02	0.03	0.03*	0.16***	-0.01
	(0.11)	(0.02)	(0.09)	(0.02)	(0.09)	(0.03)	(0.03)	(0.01)	(0.06)	(0.02)
Experienced large rise in agricultural input prices (dummy)	0.06	-0.06**	-0.18*	0.05	-0.08	0.04	0.03	0.03	-0.11**	-0.02
	(0.13)	(0.03)	(0.10)	(0.03)	(0.11)	(0.03)	(0.04)	(0.02)	(0.06)	(0.02)
Experienced livestock death or theft (dummy)	0.30**	-0.02	-0.16	-0.03	-0.11	0.02	-0.08*	0.01	0.10	-0.00
	(0.13)	(0.03)	(0.10)	(0.04)	(0.13)	(0.04)	(0.05)	(0.03)	(0.08)	(0.03)
Estimated threshold	143,550		171,201		182,362		220,418		186,306	
Linearity test										
F statistic (p-value in parenthesis)	84.48**		61.74		94.76***		50.87		50.67	
	(0.05)		(0.41)		(0.00)		(0.76)		(0.78)	

Note: $N = 2,664$, Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Other covariates are omitted from the table.

positive coefficient for livestock loss in Column (1). The loss of productive abilities induces lower-regime farmers to specialize in the production of staple crops, which may discourage crop diversification. This contrasts with negative coefficients for livestock loss among lower-regime farmers in Columns (3), (5), and (7).

Second, lower-regime farmers cultivate sorghum/roots/tubers and legumes/oil to reduce damages by agronomic shocks. As Columns (3) and (5) indicate, lower-regime farmers increase the probability of planting staples plus sorghum/roots/tubers or legumes/oil if they experienced drought/flood. Moreover, Column (3) shows that lower-regime farmers are also more likely to plant sorghum/roots/tubers in response to crop diseases/pest outbreaks. Lower-regime farmers may plant sorghum/roots/tubers and legumes/oil to build farming systems robust to agronomic shocks, which coincides with Maggio and Sitko (2021), who observed that exposure to drought encourages Malawian farmers to adopt a diversified cropping system containing sorghum/roots/tubers and legumes.

Third, farmers may rely on cropping legumes/oil to address market shocks. The probability of planting staples plus legumes/oil increases significantly if lower-regime farmers experienced rises in food prices in Column (5) and if upper-regime farmers experienced falls in crop prices in Column (6). The coefficient for rises in input prices is also positive in Column (6), though it is insignificant. Legumes/oil may be used to secure dietary intakes of protein, to alleviate income volatility caused by movements of sales price for staples, and to complement soil nutrients necessary for the growth of staple crops. Figure 3.6 indicates that more than half of the farmers cultivate legumes/oil in both regimes and that legumes/oil is the second most common crop type following staples. Because legume/oil is also used to address drought/flood, these observations suggest that legumes/oil may be a main driver for crop diversification. This claim seems consistent with the fact that staple-legume intercropping has been common historically (Morgan et al. 2019).

For other results in Table 3.5, see Appendix 3-VI.

Three Potential Factors Behind the Regime Shift

I discuss how the three potential factors behind the regime shift (subsistence requirement, balanced dietary intake, and stabilization of market income) work for crop diversification. Using the estimated threshold parameter for crop numbers in Table 3.2, I split the dataset into lower-regime and upper-regime farmers.[11] For each farmer group, I provide descriptive statistics associated with the three potential factors. Though these descriptive statistics provide no causality, they may benefit interpretations of the previous results in terms of the three potential factors.

[11]Because the estimated threshold parameters are quite similar between crop numbers and the SI, as shown in Table 3.2, a choice of estimated threshold parameters is unproblematic.

Subsistence Requirement

To discuss sources of subsistence requirement, I calculate daily calories from kilograms of each consumed food in a household by relying on Lukmanji et al. (2008). Dercon et al. (2009) and Michler and Josephson (2017) set a food poverty line of 2,300 kcal per adult per day.[12] Thus, examining the proportion of farmers below this poverty line in each regime is helpful for discussing whether the estimated thresholds in Table 3.4 divide the dataset by subsistence requirement.

The left-hand side of Figure 3.7 presents the proportion of farmers whose daily food consumption per adult equivalent is less than 2,300 kcal.[13] Around 80% of lower-regime farmers consume less than 2,300 kcal, whereas around 20% of upper-regime farmers consume less. Because there are some farmers below the poverty line in the upper regime, the estimated thresholds in Table 3.4 may not be perfectly identical with the poverty line. Nevertheless, the poor energy intakes of lower-regime farmers seem to fit the interpretation of subsistence requirement.

As was shown previously in Tables 3.4 and 3.5, livestock loss is associated with less crop diversification and more specialization in the production of staple crops. The right-hand side of Figure 3.7 depicts the proportions of lower-/upper-regime farmers owning, borrowing, or renting oxen as farm implements, separately by the experience of livestock loss. Out of farmers who experienced no livestock loss, around 20% utilized oxen in their production in both regimes. In contrast, out of farmers who experienced livestock loss, more than 40% utilized oxen in the lower regime, higher than in the upper regime by as many as 17%. Lower-regime farmers who experienced livestock loss depend more on livestock in their production than upper-regime farmers do. Owing to the higher dependence on livestock, lower-regime farmers may be more damaged by its loss than upper-regime farmers. Because staple crops are planted as crucial sources of calories (Rivera-Padilla 2020), lower-regime farmers with limited calorie intakes may prioritize the production of staple crops, if their productive abilities are restricted owing to livestock loss (As will be shown later, the productive ability of upper-regime farmers may rely on hiring labor, rather than using livestock).

In Table 3.6, I estimate effects of each shock experience on the probability for farmers to

[12]The World Bank also uses a similar value of the minimum nutritional requirements (2,200 kcal per adult per day) (`https://www.nbs.go.tz/nbs/takwimu/hbs/Tanzania_Mainland_Poverty_Assessment_Report.pdf`)

[13]In Figure 3.7, food consumption data in calories are divided by adult equivalents, not by the number of household members. Thus, the estimated calories of consumed food in Figure 3.7 are larger than those in many previous studies. Ecker and Qaim (2011) calculated the average calories of Malawian farmers as 2,171 kcal. My estimated average calories of all farmers, including both regimes, are 3,830. However, even if I divide daily food consumption by the number of household members, the estimated average calories are 3,100, which far exceeds the level of Ecker and Qaim (2011). My estimation of calories may experience large measurement errors and only show a rough comparison of consumed calories in the two regimes.

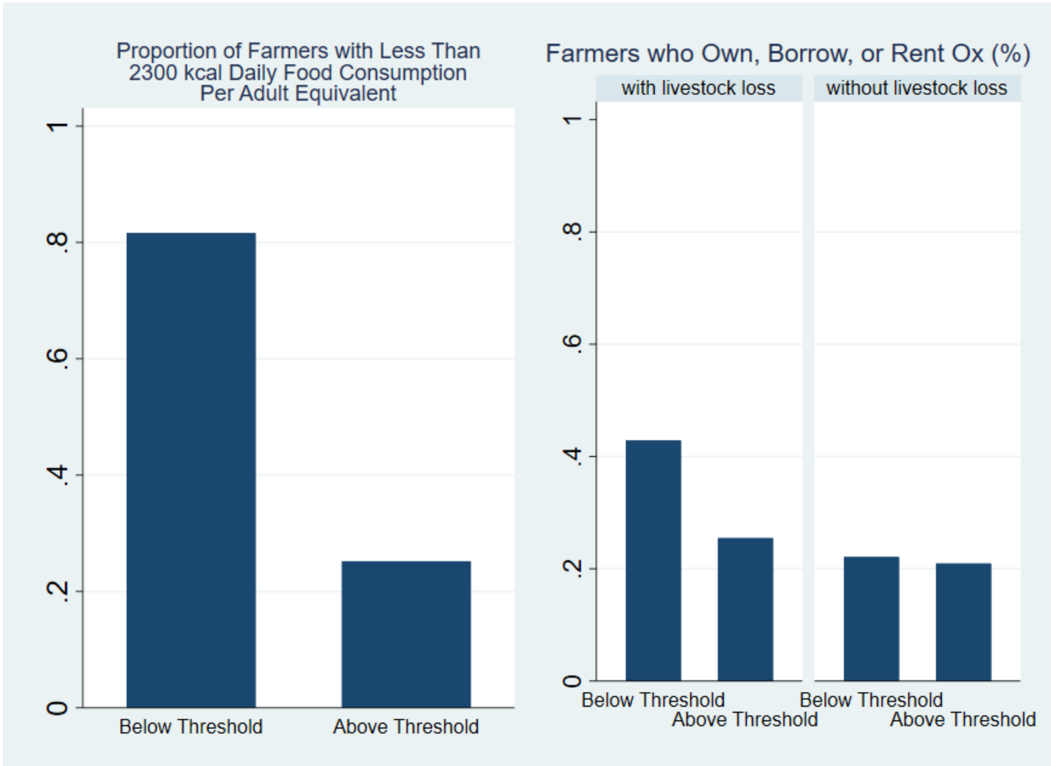


Figure 3.7: The Levels of Farmers' Food Consumption and Usage of Oxen as Farm Implements

rent-in land and acres of planted land during the long rainy season, because farmers who are damaged by past shock events may have difficulty in renting in land and cultivating land. In Table 3.6, I separate the sample into the two regimes by using the threshold in crop numbers and conduct the FE regression separately in each regime, using the same covariates as the main analysis.

In Column (1), all coefficients for shock variables are negative in the lower regime, except for drought/flood. Thus, past shock events might make it more difficult for poor farmers to rent in land. However, because they are insignificant, I find no clear evidence that poor farmers are less likely to rent in land after shocks. In Column (2), upper-regime farmers are less likely to rent in land when they experienced a rise in food price. Non-poor farmers who faced a rise in food price may have incurred more expenses for food and quitted renting in land to save cash.

In Column (3), lower-regime farmers cultivate less land when they experienced crop diseases/pest outbreaks. Crop diseases/pest outbreaks may directly decrease the amount of harvested seeds for the next production, which may lower the acre of cultivated land after the shock occurred. In contrast, lower-regime farmers cultivate more land when they experienced a fall in crop prices. Poor farmers who experienced a fall in crop prices for sales may decide to expand the cultivated land for self-consumption, rather than rely on sales activities. Therefore, negative impacts of shocks on poor farmers' capability to cultivate land may

Table 3.6: FE Estimates of the Effects of Shocks on the Probability of Renting-in Land and Acres of Planted Land during the Long Rainy Season

	Probability to rent-in		Acres of planted land	
Explanatory Variable	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime
Shock variables:				
Experienced drought/flood (dummy)	0.04	0.01	0.23	-0.54
	(0.04)	(0.02)	(0.50)	(0.44)
Experienced crop disease/pests (dummy)	-0.03	0.00	-1.41**	-0.57
	(0.03)	(0.02)	(0.66)	(0.36)
Experienced large fall in sale prices for crops (dummy)	-0.02	-0.03	1.79*	-0.08
	(0.02)	(0.02)	(0.92)	(0.34)
Experienced large rise in price of food (dummy)	-0.00	-0.03*	0.30	-0.02
	(0.01)	(0.02)	(0.53)	(0.31)
Experienced large rise in agricultural input prices (dummy)	-0.03	0.01	0.43	0.11
	(0.03)	(0.02)	(1.15)	(0.27)
Experienced livestock death or theft (dummy)	-0.10	0.00	0.47	0.13
	(0.09)	(0.02)	(0.76)	(0.31)

Note: Sample size: $N = 397$ in the lower regime and $N = 2,267$ in the upper regime. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to shock variables are omitted from the table.

depend on the type of shocks. If shocks damage agricultural resources poor farmers own for production, they may have long-standing negative impacts on the acre of cultivated land thereafter. In Column (4), I find no statistical significance.

To consider a risk-reducing role of crop diversification, I examine whether crop diversification enhances farmers' crop income and reduces their volatility. In my conceptual framework, poor farmers are motivated to adopt crop diversification to prevent heavy damages in a bad state (risk dispersion in subsistence requirements). For estimation, I follow a two-step procedure in Di Falco and Chavas (2009). In the first step, I estimate the impacts of crop diversification on gross values of farmers' harvested crops. Gross values of harvested crops were reported by farmers in Tanzanian shillings, regardless of whether these crops were sold or not. These values are transformed in logarithm in the regression. In the second step, I take the residuals in the previous regression and regress its square on crop diversity to estimate the impacts on the volatility of crop income. Both the two estimations are conducted by fixed-effects (FE) regressions.

Table 3.7 shows results of the two-steps FE estimations.[14] The indicators of crop diversification are based on crop numbers and the SI in my main analysis and standardized in each region and year. The two indicators are not included together in covariates. Rather, I estimate the respective impact of increasing crop numbers and the SI, by separately adding the two indicators to covariates. The set of covariates are the same as my main analysis. In

[14]I exclude a few missing values of harvested crops in the regressions.

Table 3.7: Estimates of the Effects of Crop Diversification on the Natural Log of Gross Values of Harvested Crops and its Volatility

	Simple Counting		Simpson Index	
Explanatory Variable	(1) Crop Income	(2) Income Volatility	(3) Crop Income	(4) Income Volatility
Indicators of crop diversification:				
Crop numbers standardized in each region and year	0.42*** (0.03)	-0.36*** (0.12)		
Simpson index standardized in each region and year			0.42*** (0.03)	-0.33*** (0.11)

Note: Sample size: $N = 2,597$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to indicators of crop diversification are omitted from the table.

Columns (1) and (3), both indicators of crop diversification are positively associated with gross values of harvested crops, whereas both indicators are negatively associated with the volatility in gross values of harvested crops in Columns (2) and (4), which is consistent with risk-reducing roles of crop diversification.

Using the thresholds in Table 3.2, I further divide farmers into lower and upper regimes and conduct the two-steps regressions separately in each regime in Table 3.8. In Columns (1), (3), (5), and (7), both indicators of crop diversification are positively associated with crop income. Farmers in the lower regime earns more crop income by crop diversification than farmers in the upper regime. In Columns (2), (4), (6), and (8), both indicators of crop diversification are negatively associated with the volatility of crop income. Moreover, the reduction in the volatility of crop income in the lower regime considerably outweighs that in the upper regime. These results require caution because I do not control for endogeneity of crop diversification. Nevertheless, these findings suggest that crop diversification contributes to the risk reduction among whole farmers and prevents devastating damages among poor farmers (risk dispersion in subsistence requirements).

Balanced Dietary Intake

To assess balanced dietary intake as a potential factor for the regime shift, I calculate the Household Dietary Diversity Score (HDDS) (Swindale and Bilinsky 2006). The HDDS categorizes food items into 12 food groups: cereals, root/tubers, vegetables, fruits, meat/poultry/offal, eggs, fish/seafood, pulses/legumes/nuts, milk and milk products, oil/fats, sugar/honey, and miscellaneous. The HDDS then counts the number of food groups from which any food item is consumed by a household.

The left-hand side of Figure 3.8 displays the average HDDS of the preceding wave for each regime.[15] In the range from 0 to 12 in the HDDS, upper-regime farmers consume two more

[15] Figure 3.8 displays the HDDS in the preceding, not contemporary, wave to ascertain whether the desire

Table 3.8: Estimates of the Effects of Crop Diversification on the Natural Log of Gross Values of Harvested Crops and its Volatility in Lower and Upper Regimes

	Simple Counting				Simpson Index			
	Lower Regime		Upper Regime		Lower Regime		Upper Regime	
Explanatory Variable	(1) Crop Income	(2) Income Volatility	(3) Crop Income	(4) Income Volatility	(5) Crop Income	(6) Income Volatility	(7) Crop Income	(8) Income Volatility
Indicators of crop diversification:								
Crop numbers standardized in each region and year	0.55^{***} (0.13)	-2.03^{***} (0.68)	0.43^{***} (0.04)	-0.29^{*} (0.15)				
Simpson index standardized in each region and year					0.47^{***} (0.14)	-2.29^{***} (0.75)	0.43^{***} (0.04)	-0.29^{*} (0.15)

Note: Sample size: $N = 391$ in the lower regime and $N = 2,206$ in the upper regime in the simple counting; $N = 388$ in the lower regime and $N = 2,209$ in the upper regime in the Simpson index. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to indicators of crop diversification are omitted from the table.

food groups than lower-regime farmers (the difference is statistically significant at the one percent level). Although the HDDS is supposed to be calculated over the previous 24 hours in Swindale and Bilinsky (2006), my reference period is the previous seven days before the questionnaire owing to data limitations. Because my reference period is long, the difference of the HDDS between the two regimes tends to shrink and be understated. Nevertheless, lower-regime farmers have less diverse intake of diets than upper-regime farmers.

Using the Hansen's fixed-effects threshold model, I estimate the impacts of crop diversification on the HDDS of the contemporary wave in Table 3.9. As in Table 3.7, I add to covariates the standardized crop numbers and SI separately. The set of covariates are the same as my main analysis. For brevity, Table 3.9 displays estimated coefficients only for each indicator of crop diversification. F-statistics are statistically significant for the two cases. In Table 3.9, estimated thresholds are considerably larger than those in my main analysis (around 200,000 in Table 3.2). I may have different threshold parameters because farmers' dietary diversity is also related to other factors than farmers' decisions on crop diversification.

Columns (1) and (2) show that increasing crop numbers raises the HDDS in both regimes by the same amount. Columns (3) and (4) also show that increasing the SI raises the HDDS in both regimes by the same amount. Although farmers' decisions on crop diversity in their farmland may still be endogenous in the fixed-effects model, these results show that crop diversification is positively associated with farmers' dietary diversity, regardless of farmers' food consumption levels. Because the HDDS is lower among lower-regime farmers in Figure 3.8, they may be more motivated to promote crop diversification, if agricultural shocks are perceived as threats to their balanced dietary intake.

To explain balanced dietary intakes as a potential factor, my conceptual framework as-

for balanced dietary intakes works as farmers' motivation for crop diversification. In contrast, Table 3.9 later relies on the HDDS in the contemporary wave to examine how crop diversification affects the HDDS.

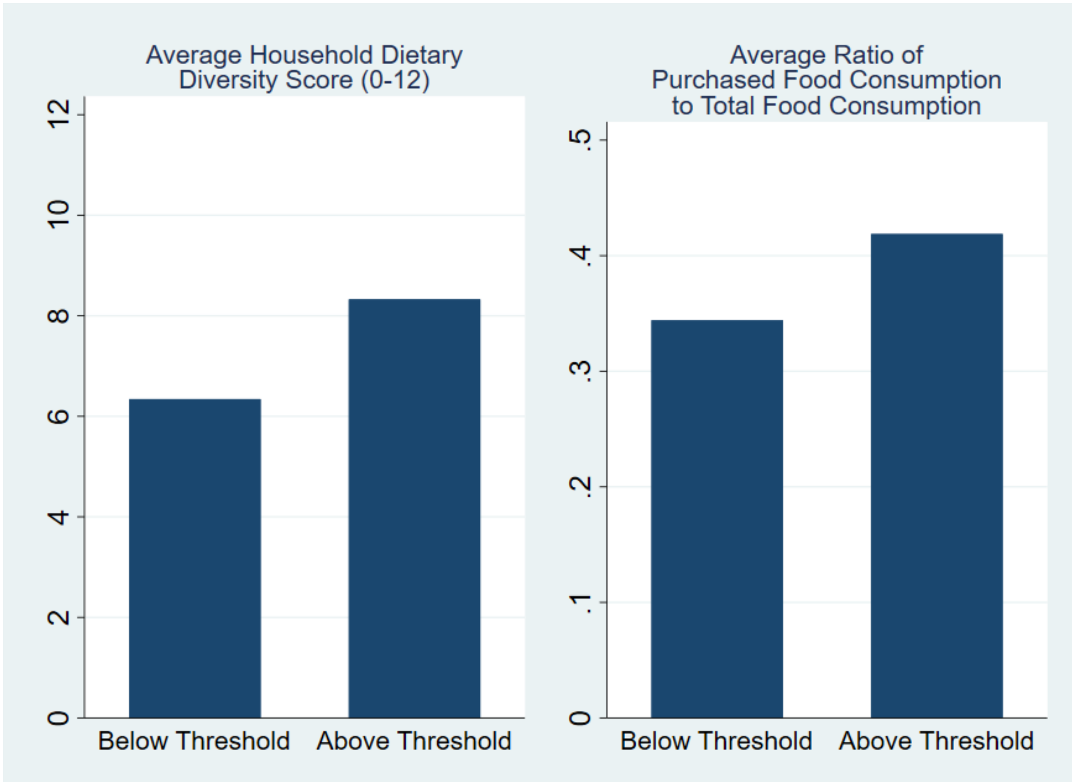


Figure 3.8: The Levels of Dietary Intake and Dependence on Purchased Food

Table 3.9: Estimates of the Effects of Crop Diversification on the HDDS

Explanatory Variable	Simple Counting		Simpson Index	
	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime
Indicators of crop diversification:				
Crop numbers standardized in each region and year	0.20*** (0.06)	0.20*** (0.07)		
Simpson index standardized in each region and year			0.19*** (0.06)	0.19** (0.08)
Estimated threshold	375,145		375,145	
Linearity test				
F statistic (p-value in parenthesis)	133.93*** (0.00)		133.60*** (0.00)	

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to indicators of crop diversification are omitted from the table.

sumes that the income of poor farmers is increased by crop diversification. As Table 3.8 previously showed, crop income is increased by crop diversification in the lower regime, which seems consistent with my assumption.

As was shown previously in Table 3.4, a rise in food price is positively associated with crop diversification. Lower-regime farmers may realize balanced dietary intakes that are unaffected by fluctuation in food price. However, it is uncertain whether lower-regime farmers consume food from not only their own production but also purchases in output markets.

The right-hand side of Figure 3.8 displays the average ratio of purchased food consumption to the total food consumption in the two regimes. To calculate this ratio, I follow Dolislager et al. (2022). The NPS provides data on expenditure on purchased food as well as its purchased quantities over the previous seven days before the interview. By dividing the expenditure by its purchased quantities, I calculate market prices of each food.[16] Then, these market prices are multiplied by total quantities of consumed food coming from own production, purchases, and gifts. Finally, I divide the expenditure on purchased food by the predicted market values of the total food consumption.

The right-hand side of Figure 3.8 shows that more than 30% of the total food consumption comes from purchases among lower-regime farmers. As Dolislager et al. (2022) shows, purchases are an important source of food consumption even for poor farmers in Tanzania. Owing to limited income of lower-regime farmers, their food securities may be threatened easily by rises in the price of purchased food, relative to upper-regime farmers.

Stabilization of Market Income

I assess farmers' participation in output and input markets in the two regimes. In Figure 3.9, the two bar graphs in the top depict the proportion of farmers' selling each type of crops and the respective average sales revenue when they sell. Although the selling rates are similar in the two regimes, upper-regime farmers achieve higher crop sales in all crop types, once they decide to sell in output markets. As shown in Figure 3.7, food securities are even less severe for upper-regime farmers than for lower-regime ones. Thus, upper-regime farmers may obtain more surplus of harvested crops for sales. To reduce the income loss caused by falls in crop prices, upper-regime farmers may have more incentives to grow multiple crops for sales.

The two graphs at the bottom of Figure 3.9 depict the proportions of farmers' purchasing inorganic fertilizer and hybrid/improved seed, and their average expenditures when they purchase. Although the purchasing proportions are statistically higher for upper-regime farmers in the two inputs, relatively sizable proportions of lower-regime farmers also purchase

[16] Using the Fisher price indices provided in the Basic Information Document by the NPS, I adjust market prices of food in each household, depending on spatial price differences.

Figure 3.9: Farmers' Participation in Output and Input Markets

them. Moreover, conditional on farmers' input purchases, I find similar average expenditures on inorganic fertilizer in both regimes, though upper-regime farmers expend more on hybrid/improved seed. Specific amounts of fertilizer may need to be applied, once farmers determine its application.

These observations suggest that lower-regime farmers may also depend on purchases in input markets and respond to a rise in input price, as upper-regime farmers do. Appendix 3-V, which redefines the two indicators of crop diversification by five crop types, finds that a rise in input price discourages the diversity at the crop-type level among lower-regime farmers, which is contrasted with positive responses by upper-regime farmers in Table 3.4. The crop-specific analysis in Table 3.5 also shows that lower-regime farmers who experienced a rise in input price are less likely to plant cash crops (see Column (9)). Lower-regime farmers may not afford to purchase higher-priced inputs because of limited income. They may fail in obtaining necessary seeds/seedlings or fertilizer and abandon planting cash crops, which may be the evidence of poverty traps.

Finally, I investigate farmers' demands for hiring agricultural labor. Hired labor also constitutes an agricultural input in markets. Figure 3.10 shows the percentage of farmers hiring labor for agricultural activities in the left-hand side and total days of hired labor (conditional on farmers' hiring) in the right-hand side, separately in the two regimes. In each side, Figure 3.10 further depicts purposes farmers hire labor: land preparation and

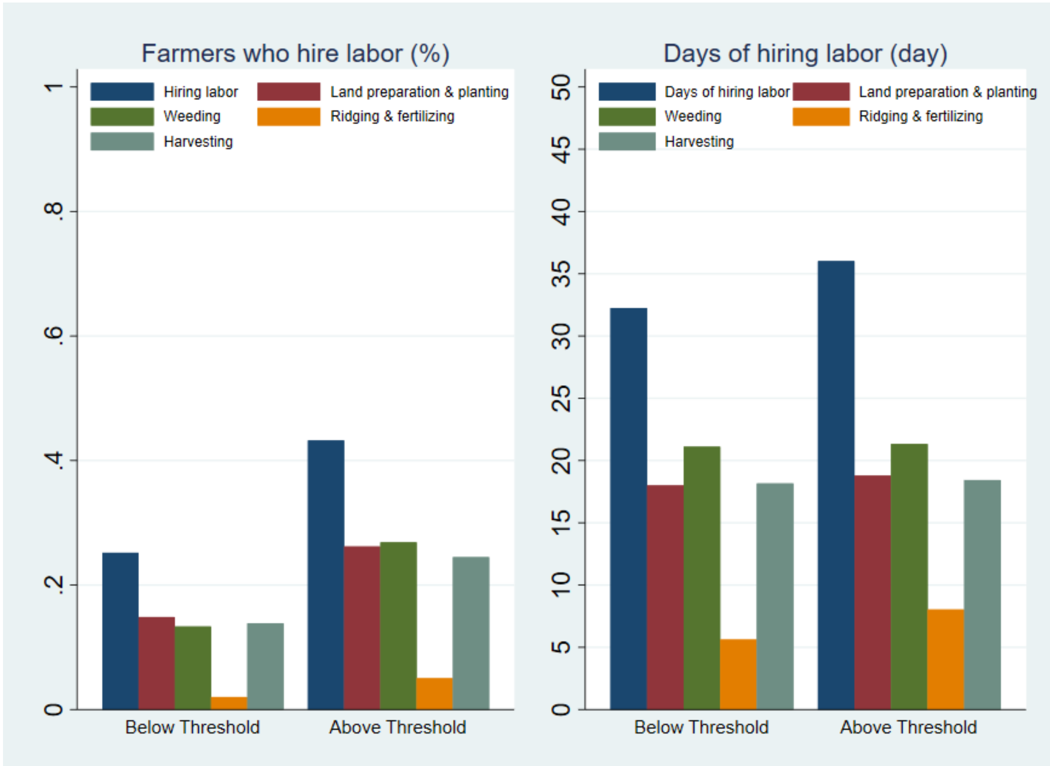


Figure 3.10: Percentage of Farmers Hiring Labor for Agricultural Activities and Total Days of Hired Labor Conditional on Farmers' Hiring

planting, weeding, ridging and fertilizing, and harvesting.

The left-hand side of Figure 3.10 shows that more upper-regime farmers hire labor for their agricultural activities than lower-regime farmers (18% higher in upper-regime farmers). Furthermore, this tendency applies to all purposes of hiring labor. In contrast, in the right-hand side of Figure 3.10, upper-regime farmers demand only 4 more labor days in total than lower-regime farmers, when they hire labor. Moreover, days of hiring labor are similar in the two regimes in all purposes. In Table 4.3, there is no significant difference in land sizes between the two regimes. Thus, once farmers decide to hire labor, labor days necessary for achieving agricultural purposes may be the same.

In conclusion, upper-regime farmers hire labor for their agricultural activities more actively than lower-regime farmers do, though demanded labor days are similar between the two regimes. Hired labor may play a main role in the productive ability of upper-regime farmers, unlike lower-regime farmers who rely on livestock. Moreover, the usage of tractors is more prevalent in the upper regime; out of the lower-regime farmers, 5.5% owned, borrowed, or rented tractors, whereas 10.8% of the upper-regime farmers did. Thus, agricultural resources for production tend to come from market transactions among upper-regime farmers.

3.5.2 Robustness check

In Appendix 3-VII, I construct crop numbers and the SI by excluding fruit trees and permanent cash crops because the cultivation of permanent crops may be unsusceptible to shocks in the short run once farmers plant them. I then check the robustness of my main results and find similar results in the SI.

In Appendix 3-VII, I also separately estimate the probability that farmers cultivate each of legumes (including pigeon peas), vegetables, fruits, annual cash crops, and permanent cash crops in addition to staples. Because farmers may plant annual and permanent crops in a different time span, I separate vegetables/fruits and annual and permanent cash crops in the crop-specific analyses. Moreover, the classification of pigeon pea is moved from cash crops to legumes. Although pigeon pea is classified as permanent cash crops in the NPS (see Appendix 3-III), it also has crop characteristics as legumes. I observe similar results in legumes, vegetables, and permanent cash crops.

In Appendix 3-VIII, I incorporate the short rainy season, the secondary cropping season in Tanzania, by aggregating indicators of crop diversification over different rainy seasons. I still find similar results.

Appendix 3-IX tackles the limited sample size of my dataset and the possible endogeneity of food consumption. My dataset is expanded as an unbalanced panel comprising all waves from 2008 to 2018. Furthermore, I adopt a different threshold specified by Makate et al. (2022), that is, the bottom 40% of the sample distribution of asset holdings. I find that lower-regime farmers adopt crop diversification in response to drought/flood and disadopt it in response to livestock loss. In contrast, upper-regime farmers may promote crop diversification in response to a fall in crop price, though the evidence is weak.

Finally, I must admit two limitations of this chapter. First, this study relies on the use of self-reported shock variables. Although I find no significant association between all shock experiences and the regime shift in the FE estimation, subjective measures may contain heterogeneity by wealth levels. In contrast, it is subjective perception of future shocks that influences farmers' choices of crop diversity. Although my analysis may not fully remove heterogeneity of subjective measures, they still provide useful insights to farmers' decisions on crop diversification.

Second, this study does not separate effects of the three potential factors in each shock. As Table 3.1 shows, the effects of the three potential factors overlap or conflict with each other. However, this study does not extract effects of each factor in the main results of farmers' responses to each shock event. Future research should design an empirical framework to directly examine specific effects of each factor.

3.6 Conclusion

Introducing the subsistence constraint into a farmer's utility maximization framework under uncertainty, this chapter first highlighted three potential factors behind heterogeneous behaviors on crop diversification between poor and non-poor farmers. First, subsistence requirement has opposite effects, depending on poor farmers' perception of future states: if poor farmers predict insufficient energy intakes under diversified crop portfolios, they are forced to specialize in growing staple crops with high-caloric intakes for survival. Conversely, if specialization in a certain crop involves the risk of heavy damages, subsistence requirement induces poor farmers to disperse risks by crop diversification. Second, the desire for balanced dietary intakes encourages crop diversification among poor farmers, who prefer to consume various crops as well as necessary ones for survival. Third, stabilization of market income is a major motivation for crop diversification among non-poor farmers, who need market income to purchase luxury food and non-food items.

Using two waves of the Tanzanian NPS, I then examined whether experiences of agronomic and market shocks affect farmers' adoption of crop diversification differently by the level of food consumption. Farmers' heterogeneous behaviors on crop diversification are formulated as a regime shift by Hansen's (1999) fixed-effects threshold model.

I find that poor farmers below the threshold adopt crop diversification in response to drought/flood and a rise in food price. Poor farmers may disperse risks by cultivating climatically resistant crops and realize balanced dietary intakes that are unaffected by food price fluctuation. Conversely, the loss of livestock (productive assets for tillage) traps poor farmers in low crop diversity. In contrast, non-poor farmers above the threshold adopt crop diversification in response to a fall in crop price for sale and a rise in input price for purchase. They may stabilize market income by crop diversification.

Crop-specific analyses show that legumes/oil are a main driver for crop diversification. Poor farmers who experienced drought/flood and rises in food prices tend to cultivate legumes/oil, whereas non-poor farmers who experienced falls in crop prices tend to cultivate them. Poor farmers also plant sorghum/roots/tubers, low-valued but drought-resistant crops, to address agronomic shocks. Conversely, poor farmers who lost productive resources (livestock) specialize in the production of staple crops.

These results imply that farmers' motivations for crop diversification reflect their priorities in their livelihoods: food security in poor farmers and stabilization of market income in non-poor farmers. Farmers decide to diversify crop composition in farmland if a shock is likely to threaten farmers' priorities in their livelihoods. Moreover, poor farmers may specialize in the cultivation of staple crops, if they predict insufficient energy intakes under crop diversification.

To promote crop diversification as policy tools, we must consider different livelihood priorities between poor and non-poor farmers. When we promote crop diversification among poor farmers, we should encourage them to plant drought-resistant crops (e.g., sorghum, legumes) to prevent devastating climatic damages that threaten their food security. In contrast, we should encourage non-poor farmers to plant crops which are expected to produce market income stably. One possible policy tool is a subsidy program that provides different seeds by farmers' economic levels. Subsidizing the purchase of agricultural resources (e.g., oxen) may also be useful when shocks attack poor farmers, because they may suffer from the lack of agricultural resources to adopt crop diversification.

Chapter 4

Welfare Impacts of Farmers' Cultivation in a Secondary Cropping Season: Concepts and Evidence from Tanzania

4.1 Introduction

Food security during the lean season is an important issue in Sub-Saharan Africa (SSA) (Dercon and Krishnan 2000). In the lean season, farmers must manage the consumption of food within the amount of harvested crops until the next harvest season. Moreover, grain prices tend to rise considerably in the lean season because of the lack of food supplies in markets, which imposes a heavy burden on farmers' livelihoods (Gilbert et al. 2017; Burke et al. 2019). In Malawi, households' caloric intakes and dietary diversity decline by around 20% in the lean season relative to the post harvest season (Gelli et al. 2017).

As a policy tool to reduce hunger in the lean season, previous studies paid little attention to a secondary cropping season, in which farmers produce food after a main cropping period. One strand of policies gaining attention is the proper postharvest management, which reduces a harvest loss to increase available food during the lean season (Ndegwa et al. 2016; Omotilewa et al. 2018; Brander et al. 2021; Channa et al. 2022; Chegere et al. 2022; Ricker-Gilbert et al. 2022; Nindi et al. 2023). Another strand of policies is to combine credit with price seasonality. Farmers who receive loans in the harvest season can afford to store harvested crops for sales in higher prices during the lean season (Aggarwal et al. 2018; Burke et al. 2019; Fink et al. 2020; Channa et al. 2022; Delavallade and Godlonton 2023; Nindi et al. 2023). Safety net programs providing employment, cash, and in-kind support during the lean season also attracted attention (Berhane et al. 2014; Beegle et al. 2017; Hoddinott et al, 2018). Although these studies did not focus on a secondary cropping, farmers may be able to alleviate hunger in the lean season by cultivating again after a main cropping period.

To explore this possibility, this chapter examines whether Tanzanian farmers can improve their food security by cultivating in the short rainy season, which is a secondary cropping period in East Africa. In Tanzania, the main cropping period is the long rainy season (*masika*) from March to May, and most farmers cultivate in this season. If farmers harvest crops in July/August and do not cultivate thereafter until the next long rain, they must endure the lean season by consuming the remaining harvests. In contrast, some areas of Tanzania have the short rainy season from September/October through December/January. By cultivating in this season, Tanzanian farmers can produce additional food before the start of the next long rainy season, which may improve farmers' food security in the lean season (Oestigaard 2016).

This chapter has two main objectives. First, I empirically examine whether farmers obtain more caloric intakes and realize more dietary diversity by cultivating in the short rainy season. I further examine how the impacts of the cultivation in the short rainy season are changed, depending on whether farmers have non-agricultural occupations. This is because the hunger during the lean season may be alleviated by the existence of off-farm income.

Second, I attempt to provide the entire picture of farmers' activities in the short rainy season, which is unfamiliar owing to the low research interest by previous studies. Previous studies frequently focused on the long rainy season, while excluding the short rainy season from the analysis (Bevis et al 2019; Bhargava et al. 2019). Exceptionally, a few studies investigated welfare impacts of seasonal droughts in the short rainy season and farmers' perceptions on its long-run rainfall variability in Ethiopia (Holden and Shiferaw 2004; Kassie et al. 2013; Dimitrova 2021). However, even these studies did not examine the cultivation in the short rainy season as a main research topic. To consider the importance of the short rainy season as a research topic, I compare the prevalence, crop compositions, and productivity in the short rainy season with those in the long rainy season.

My dataset comes from six waves of data from 2008/09 to 2020/21 in the Tanzanian National Panel Survey (NPS). The descriptive statistics show that the cultivation in the short rainy season is not a minor or special economic activity in terms of the prevalence among farmers, crop compositions, and productivity. I use the fixed-effects two-stage least squares (FE-2SLS) regression. My results show that farmers obtain more caloric intakes and dietary diversity by cultivating in the short rainy season. Moreover, smallholders with small farm size benefit more per acre of planted land in the short rainy season. I also use an endogenous switching regression (ESR) model proposed by Murtazashvili and Wooldridge (2016). My results show that farmers who have no occupation except agriculture obtain more caloric intakes by cultivating in the short rainy season, whereas farmers who have non-agricultural occupations realize more dietary diversity by doing so.

This chapter makes two contributions to previous literature. First, this chapter explores a new policy tool to improve farmers' food security in the lean season. To my knowledge, this chapter is the first study to link farmers' secondary cropping to their food security in the lean season. My findings may also be applicable to other East African countries which have the short rainy season. Second, I clarify that the cultivation in the short rainy season has a potential role to reduce farmers' economic inequality. My results suggest that land-poor farmers gain more food consumption per acre of planted land in the short rainy season than land-rich farmers do.

This chapter proceeds as follows. Section 4.2 provides background information. Section 4.3 presents the conceptual framework. Section 4.4 explains the estimation strategy. Section 4.5 presents analytical results. Section 4.6 provides discussion. Finally, I conclude this chapter.

4.2 Background Information

4.2.1 Dataset Description

Because the short rainy season was not a main research topic in previous studies, I first attempt to provide the entire picture of farmers' activities in the short rainy season, using the Tanzanian National Panel Survey (NPS). In the NPS, the Tanzanian National Bureau of Statistics (NBS) collected agricultural and socioeconomic information from nationally representative households through six waves from 2008/09 to 2020/21.

Table 4.1 displays the number of households and agricultural plots in my dataset. My dataset focuses on farmers who have agricultural plots in the Tanzanian mainland. My panel dataset are unbalaned and comprise households and plots with at least two waves of data. After the first three waves (2008/09-2012/13), the composition of households was refreshed in 2014/15. The majority households are 1582 households who were newly added in this wave. They were interviewed again in 2020/21. The remaining households in 2014/15 are 504 households who were interviewed in earlier waves and were also interviewed in 2018/19. NBS (2022) draws a clear picture on the history of these data collection over waves, which is provided in Appendix 4-I. To prevent outlier effects in my analysis later, I exclude households with the largest 1% of calorie intakes and farm size from the household-level dataset.

Table 4.2 roughly sketches the timeline of agricultural data in the NPS, taking a case of Wave 2012/13. In each wave, the period of the long rainy season was the same over all interviewed farmers (March-May in 2012 in Table 4.2). In contrast, farmers were asked of the most recent short rainy season from the date of interview. Thus, farmers who were interviewed early provided information on the short rainy season *before* the long rainy season (October

Table 4.1: Number of Households and Plots in Each Wave

	2008/09	2010/11	2012/13	2014/15	2018/19	2020/21	Total Households
Number of Households	1,799	2,185	2,204	2,086	499	1,582	10,355
Number of Plots	3,651	4,874	4,670	4,650	1,151	3,448	22,444

Table 4.2: Timeline of NPS in Wave 2012/13

	2011			2012												2013								
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Short rainy season (*before*)	Cultivate			Harvest																				
Long rainy season						Cultivate			Harvest															
Short rainy season (*after*)													Cultivate			Harvest								
Month of interview in NPS													Interview											

to December in 2011 in Table 4.2). Conversely, farmers who were interviewed late provided information on the short rainy season *after* the long rainy season (October to December in 2012 in Table 4.2). All waves in the NPS have the same chronological structure. This chronological difference of the short rainy season will be used in descriptive statistics and the construction of instrumental variables and rainfall variables later.

Table 4.2 roughly sketches each agricultural period for clarity. Appendix 4-I displays the distributions of harvest months, which largely coincide with time intervals in Table 4.2.

4.2.2 Farmers' Activities in the Short Rainy Season

Regional Distribution of Farmers Cultivating in the Short Rainy Season

Using the household-level data previously explained, Figure 4.1 illustrates the number of farmers who (1) cultivated only in the long rainy season, (2) cultivated only in the short rainy season, and (3) cultivated in both rainy seasons. Figure 4.1 divides the Tanzanian mainland into five geographical areas: Northeast, Coastal, Central, Lakes, and Southern Highlands.[1]

Figure 4.1 shows that the number of farmers cultivating in the short rainy season is not small. More than half of the farmers in Northeast (e.g., Kilimanjaro region) and Lakes (i.e., regions around Lake Victoria) cultivate in both long and short rainy seasons, which coincides with the typical bimodal rainfall pattern in these areas. The rest of the country tends to have a unimodal rainfall pattern of the long rainy season. However, the bimodal rainfall

[1]Each geographical area in Figure 4.1 contains the following regions under old administrative boundaries before 2012 ("region" is the largest administrative unit in Tanzania): Northeast for Arusha, Kilimanjaro, Manyara, and Tanga regions; Coastal for Dar es Salaam, Lindi, Morogoro, Mtwara, and Pwani regions; Central for Dodoma, Kigoma, Singida, and Tabora regions; Lakes for Kagera, Mara, Mwanza, and Shinyanga regions; Southern Highlands for Iringa, Mbeya, Rukwa, and Ruvuma regions.

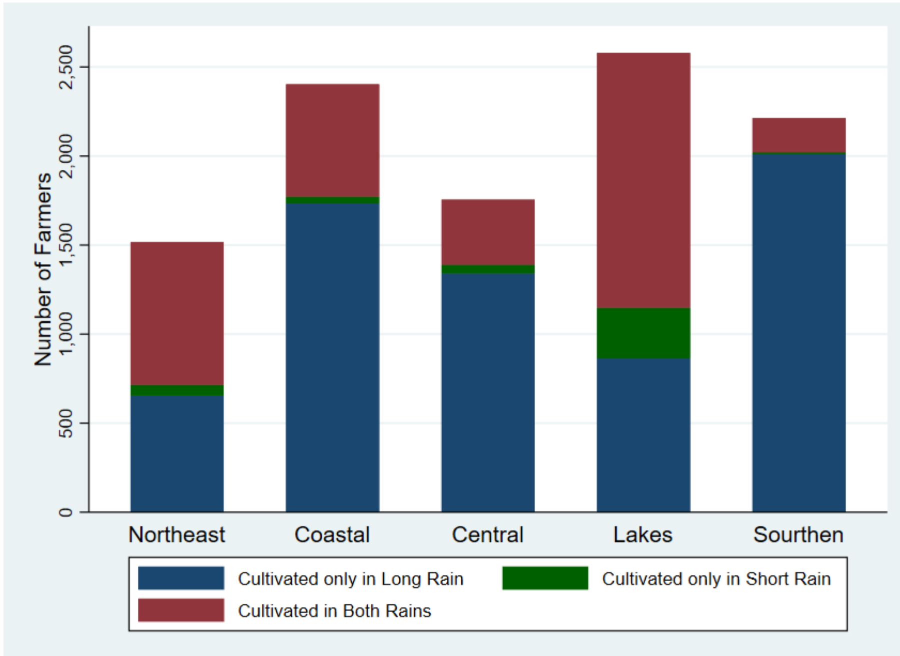


Figure 4.1: Number of Farmers Cultivating in Each Rainfall Season

pattern is also observed partly in Coastal, Central, and Southern Highlands areas (FAO).[2] In Figure 4.1, some farmers in these areas cultivate in both two rainy seasons.

Composition of Planted Crops in the Short Rainy Season

Using the the plot-level dataset, Figure 4.2 displays the top 10 crops farmers reported they cultivate as a main crop in their plots in each long and short rainy season. Because compositions of planted crops in farmland strongly depend on agro-ecological conditions, I focus on plots in Northeast and Lakes areas, where the cultivation in the short rainy season is prevalent (see Figure 4.1).

Figure 4.2 shows that compositions of planted crops are considerably similar between the two rainy seasons. Nine out of the ten crops are the same in the two rainy seasons. Maize is planted as a main crop in nearly 50% of plots in both two rainy seasons, which suggests that annual crops are widely planted in the short rainy season. Perennial fruit trees and crops such as banana and cassava are cultivated in some plots in the short rainy season, but this pattern also applies to the long rainy season. Farmers do not to plant minor, non-staple, or special crops in the short rainy season, compared with the long rainy season.

Economic Scales and Productivity in the Short Rainy Season

In Figure 4.3, I then examine farmers' production in the short rainy season in Northeast and Lakes areas. In Figure 4.3, 95% confidence intervals (CI) are overlaid for mean values.

[2]See https://www.fao.org/3/w7958e/w7958e00.htm.

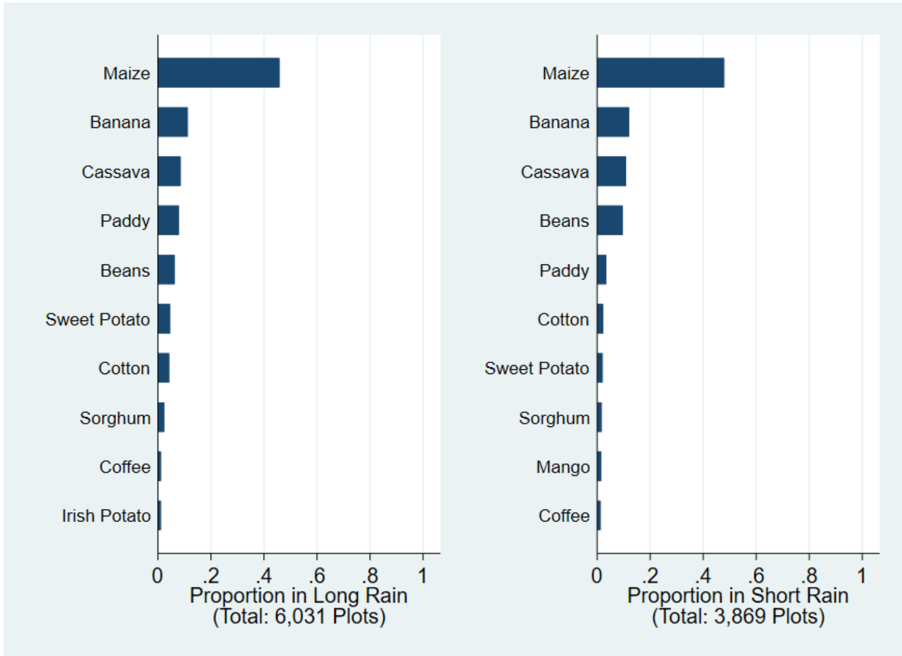


Figure 4.2: Top 10 Crops Cultivated as a Main Crop in the Long and Short Rainy Seasons in Northeast/Lakes

Using the household-level dataset, Figure 4.3(a) displays how many acres of land are used for planting annual crops on average in each long and short rainy season. The average size of planted land in the short rainy season amounts to around two-thirds of that in the long rainy season.

Using the plot-level dataset, Figure 4.3(b) displays the average kilograms of harvested maize per acre in the long and short rainy seasons. The average maize productivity is almost the same in the two rainy seasons. This is surprising because rainfall tends to be lower and more uncertain in the short rainy season (Appendix 4-I depicts the rainfall pattern in Kilimanjaro region).

If farmers cultivate in the short rainy season, it may deplete soil nutrients and lead to low productivity in the next long rainy season. In Figure 4.3(c), I examine whether maize yields per acre in the long rainy season decrease if a plot is cultivated in the last short rainy season. Interestingly, the maize productivity in the long rainy season is similar, regardless of whether farmers cultivate in the previous short rainy season.

Appendix 4-I also shows that maize prices in the short rainy season tend to be higher than those in the harvest season after long rains, which indicates the potential for farmers to earn more income by cultivating in the short rainy season. Moreover, maize prices in Northeast/Lakes tend to be lower in the harvest season of short rains than those in other areas, which suggests that other households than farmers also enjoy welfare improvements through lower food prices.

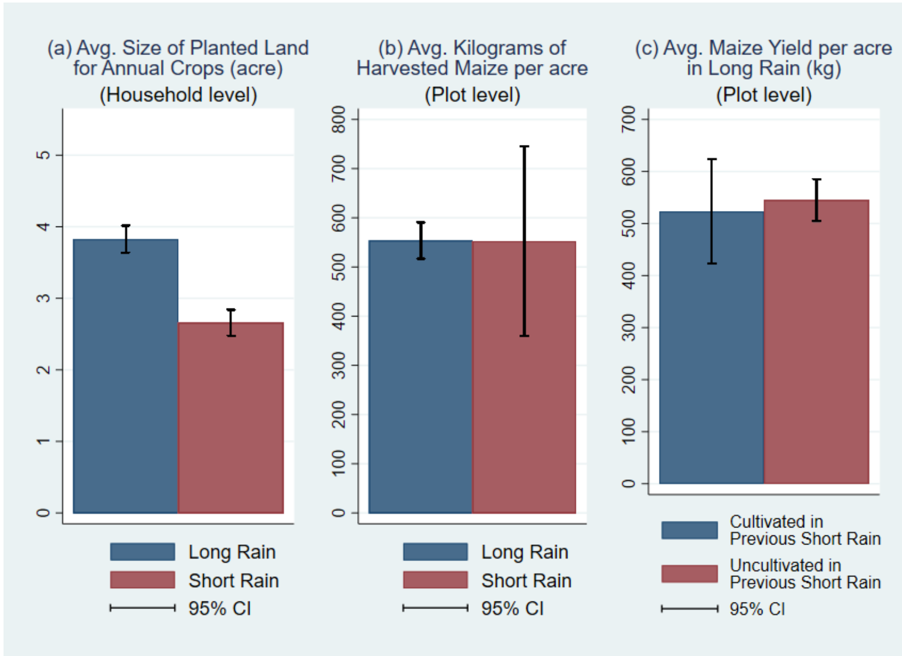


Figure 4.3: Economic Scale of Production and Maize Productivity in Northeast/Lakes. 95% Confidence intervals (CI) for mean values are overlaid.

Relationship between Non-Farm Activities and Cultivation in the Short Rainy Season

If farmers have other income sources than agriculture during the lean season, farmers may not have to rely on the cultivation in the short rainy season. Thus, we must attend to the existence of farmers' off-farm income during the short rainy season. Unfortunately, the NPS does not provide information on when farmers obtain off-farm income. Instead, this chapter uses data on whether any member within a household has a non-agricultural occupation such as the engagement in non-agricultural sectors (fishing, mining, and tourism), the employment by organizations (the central/local governments, private sector, and NGO/religious groups), and the operation of own business.[3];

Using the household-level data nationwide, Figure 4.4 depicts proportions of households whose member has a non-agricultural occupation. As a proxy for farmers' economic levels, Figure 4.4 divides farmers by the 10th, 25th, 50th, 75th, and 90th percentiles of the total expenditure including food and non-food consumption per 28 days in Tanzanian shillings. Regardless of whether they cultivate in the short rainy season in addition to the long one, the proportions of farmers having non-agricultural occupations are similar, except for the poorest below 10th percentile and the richest above the 90th percentile. This suggests that farmers may not determine the cultivation in the short rainy season, depending on the existence of

[3]Non-farm occupations do not include agricultural wage because farm employment during the short rainy season highly depends on whether other farmers in the community cultivate in this season and may lack the steady nature of off-farm income.

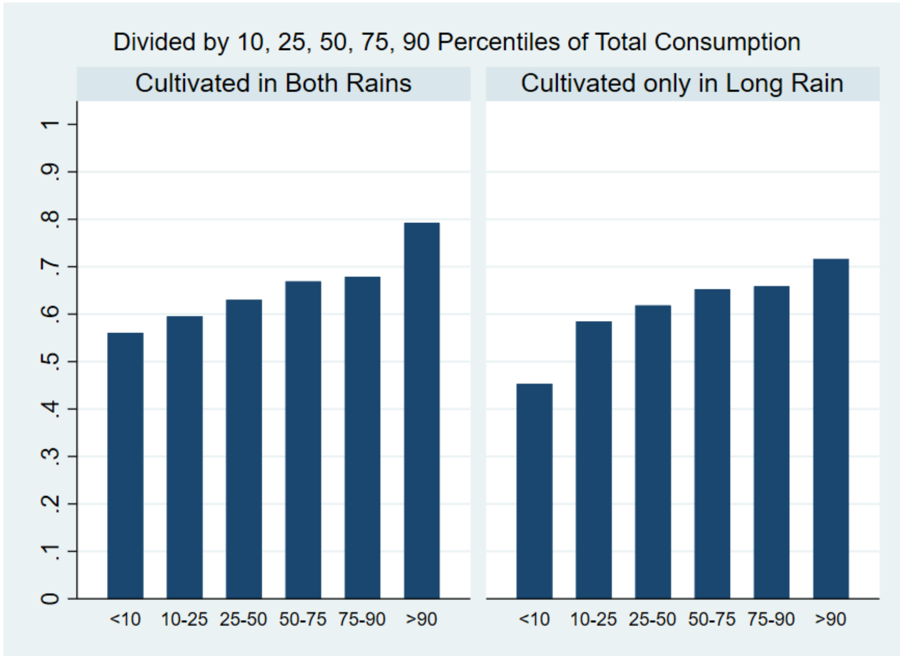


Figure 4.4: Proportions of Households whose Members Have Non-Farm Occupations

off-farm income sources within a household.

Moreover, Figure 4.4 shows that poor farmers are less likely to have non-agricultural occupations, which suggests that the existence of off-farm income is closely related to farmers' economic levels.

Appendix 4-I shows that, in all types of non-agricultural occupations, the average monthly non-farm income per individual is lower than the average monthly food consumption per adult equivalent. Thus, non-farm income may still be insufficient to fully satisfy farmers' food demands during the lean season. Harvested crops during the short rainy season and their cash income from sales may improve the quantity or quality of farmers' food consumption, even if they have off-farm income.

4.3 Conceptual Framework

4.3.1 Economic Framework on the Lean Season

This section analyzes how additional agricultural income in the short rainy season affects farmers' food consumption levels. In the interviews with Malawian farmers, Walls et al. (2023) reported that many families regularly eat only one meal per day in the lean season; some families do not eat at all. In this extreme poverty, farmers have to prioritize consuming staple foods with high energy (e.g., maize) for survival. They cannot afford to enjoy a variety of foods until they ease hunger. Previous studies on cash transfer support the possibility that farmers' priorities on food consumption change depending on the level of their caloric intakes.

Very poor households in rural Niger tended to make bulk purchases of cheap grain, rather than purchases of a variety of foods, when they received the cash transfer (Hoddinott et al. 2018). In contrast, the cash transfer improved the dietary diversity of relatively poor households in Ecuador and Mexico (Cunha 2014; Hidrobo et al. 2014).

Therefore, farmers' priorities of food consumption seem to change between basic caloric intakes and dietary diversity. In the customary consumer problem, a consumer maximizes a utility function $u(x)$ under the budget constraint $\sum_{i=1}^{N} p_i x_i \leq I$, where $x = (x_1, \ldots, x_N)$ is a consumption bundle, $p = (p_1, \ldots, p_N)$ is a price vector, and I is income. I further introduce the subsistence constraint below:

$$\sum_{i=1}^{N} \epsilon_i x_i \geq R, \tag{4.3.1}$$

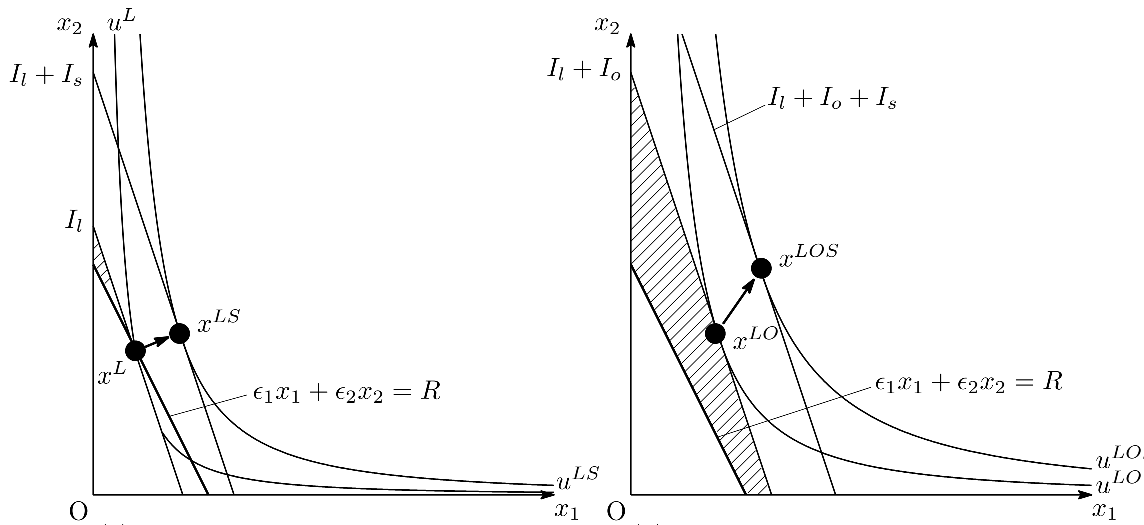
where $R > 0$ represents the requirement for basic caloric needs and $\epsilon_i \geq 0$ is a caloric amount per unit of consumption for x_i. For example, staple crops (e.g., maize) have high values of ϵ_i because of their high caloric content, whereas cash crops with low caloric content (e.g., coffee) have low values of ϵ_i.

4.3.2 Farmers' Welfare Improvements through Cultivation in the Short Rainy Season

I illustrate how additional income during the lean season could result in farmers' welfare improvements. I suppose that all farmers cultivated in the long rainy season before the lean season and have the rest of agricultural income unconsumed in the long rainy season (I_l).

In reality, farmers who determine whether to cultivate in the short rainy season face opportunity costs for enjoying leisure and earning off-farm income. However, food prices, in particular those of staples, tend to be high during the lean season because foods are less supplied in markets in this season than in the harvest season. Thus, if farmers have no income sources during the lean season, they have to care about the subsistence constraint. Even if the subsistence constraint does not bind, the marginal utility of consuming high-priced foods is large relative to that of consuming leisure during the lean season. As long as farmers can maintain nutrients in the soil, access to water for planting, and have the sufficient labor force in a household, they may have strong incentives to cultivate in the short rainy season. As was shown previously in Figure 4.4, farmers decide to cultivate in the short rainy season, regardless of the existence of off-farm income.

In Figure 4.5(a), I consider poor farmers who do not have income except I_l in the lean season. Although staple crops have high caloric content and are necessities for households, these crops are insufficiently supplied in markets during the lean season. Thus, $\frac{p_1}{p_2} > \frac{\epsilon_1}{\epsilon_2}$ in Figure 4.5(a). In order to survive in the lean season, the poor farmers have to rely on only I_l.



(a) Poor farmers without off-farm income (b) Non-poor farmers with off-farm income

Figure 4.5: Impacts of additional agricultural income in the short rainy season. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints in x^L and x^{LO}.

Then, the subsistence constraint is binding for the solution x^L. In x^L, the poor farmers wish to consume x_1 sufficiently for the marginal rate of substitution (MRS) to coincide with the price ratio, but they cannot do so because of the lack of income to purchase the high-priced staple crop x_1.

Suppose that the poor farmers obtain additional agricultural income (I_s) by cultivating in the short rainy season. Then, the solution moves from x^L to x^{LS}, which mainly increases x_1 and thus raises caloric intakes. The poor farmers who survive at subsistence levels ($f(x^L) = R$) spend additional income on increasing their caloric intakes. Therefore, this chapter first presents the following hypothesis:

Hypothesis 1. *Farmers who have no off-farm income improve their caloric intakes by cultivating in the short rainy season.*

In contrast, some farmers may have off-farm income I_o as well as I_l in the lean season. In Figure 4.5(b), I consider non-poor farmers who have $I_l + I_o$. Because of the additional income in the lean season, the subsistence constraint is not binding for the solution x^{LO}. Although the non-poor farmers already consume x_1 sufficiently for the MRS to coincide with the price ratio, they may still be eager to obtain more income so that they can enjoy a variety of food in the lean season.

Suppose again that the non-poor farmers obtain additional agricultural income by cultivating in the short rainy season (I_s). Then, the solution moves from x^{LO} to x^{LOS}, which mainly improves the dietary diversity between x_1 and x_2. Because the non-poor farmers already consume enough calories at x^{LO} under $I_l + I_o$, they spend additional income I_s on enjoying a variety of food. Therefore, this chapter presents the following second hypothesis:

Hypothesis 2. *Farmers who have off-farm income improve their dietary diversity by cultivating in the short rainy season.*

Using the household-level dataset in Table 4.1, Appendix 4-I shows that households who meet the basic caloric needs are more likely to consume food such as meat and legumes, rather than staple foods, which suggests the conceptual process here.

There are two caveats on the conceptual framework. First, my framework does not exclude the possibilities that farmers without off-farm income improve dietary diversity and that farmers with off-farm income improve caloric intakes. If farmers obtain sufficiently large income by cultivating in the short rainy season, these possibilities arise. My framework just stresses that main channels of welfare improvements may be different, depending on the extent of poverty. Second, the level of each income (I_l, I_o, and I_s) was assumed to be exogenously given to all farmers so that I can highlight welfare impacts of additional income in the lean season. In reality, farmers may have to decide whether to cultivate in the short rainy season and whether to do non-farm activities. To consider exogenous effects of I_o and I_s, my empirical analysis needs to control for farmers' decisions to cultivate in the short rainy season and their self-selection into non-farm activities.

4.4 Methodology

4.4.1 Dataset

Using the household-level dataset in Table 4.1, I will test the two hypotheses in Conceptual Framework. Because my panel dataset is unbalanced, I test for the sample selection bias by examining the statistical significance for a lagged selection indicator in the fixed-effects model (see pp 832-833 in Wooldridge (2010) for details). Because I find no evidence of the sample selection bias, I proceed without correcting for it.

4.4.2 Welfare Indicators of Farmers' Food Consumption

To estimate welfare impacts of farmers' cultivation in the short rainy season, I first calculate household's caloric intakes (kcal) per adult equivalent on a daily basis. To measure households' dietary diversity, I then calculate the Household Dietary Diversity Score (HDDS) (Swindale and Bilinsky 2006). In the HDDS, all food items are classified into 12 food groups: cereals, root/tubers, vegetables, fruits, meat/poultry/offal, eggs, fish/seafood, pulses/legumes/nuts, milk and milk products, oil/fats, sugar/honey, and miscellaneous. I calculate the HDDS by counting the number of food groups that a household consumed in the previous week before the interview.

The NPS provides food consumption data over the past week before the interview. Thus, the period of my consumption data is not limited to the lean season before the long rainy season. After the main analysis, I will discuss whether the cultivation in the short rainy season tends to improve farmers' food consumption in the lean season.

4.4.3 Estimation Strategy

Statistical Models

I first estimate welfare impacts of the cultivation in the short rainy season, without taking off-farm income into consideration:

$$y_{it} = \alpha_0 + \alpha_1 S_{it} + \alpha_2 S_{it}^2 + \boldsymbol{\alpha}_3^{'} \boldsymbol{x}_{it} + \mu_i + \epsilon_{it}, \tag{4.4.1}$$

where y_{it} is daily caloric intakes per adult equivalent (kcal) or weekly HDDS of farmer i in year t. The variables of my interest are S_{it} and S_{it}^2, acres of land for planting annual crops during the short rainy season and their squared term, respectively. I focus on the planting of annual crops because perennial crops such as fruit trees and permanent cash crops may be regularly harvested every year after their initial planting. I focus on new additional income caused by the planting of annual crops, rather than regular income by perennial crops.[4] In Equation (4.4.1), the squared term, S_{it}^2, reflects the inverse relationship between farm productivity and farm size, which is widely observed in developing countries (Bevis and Barrett 2020). The additional increase of harvests per land may decrease as the farm size increases (Appendix 4-III provides its supporting evidence on maize harvests). Thus, welfare returns to land may also diminish over planting scales. Therefore, I hypothesize $\alpha_1 > 0$ and $\alpha_2 < 0$ in Equation (4.4.1).[5] Household characteristics are included as covariates, $\boldsymbol{x}_{it}$. Farmers' time-constant and time-variant heterogeneity are also included as μ_i and ϵ_{it}, respectively.

I then estimate welfare impacts of the cultivation in the short rainy season, separately by the existence of off-farm income sources:

$$\begin{aligned} y_{it} = \beta_0 + &\beta_1 S_{it} + \beta_2 S_{it}^2 + \boldsymbol{\beta}_3^{'} \boldsymbol{x}_{it} \\ &+ \textit{off-farm income}_{it} \times (\gamma_0 + \gamma_1 S_{it} + \gamma_2 S_{it}^2 + \boldsymbol{\gamma}_3^{'} \boldsymbol{x}_{it}) + \mu_i + \epsilon_{it}, \end{aligned} \tag{4.4.2}$$

where *off-farm income*$_{it}$ is a dummy variable that is equal to one if any household member

[4]Moreover, the NPS provides detailed information on planting areas of annual crops.

[5]I do not include acres of land cultivated in the long rainy season because most farmers cultivate in this season, as shown in Figure 4.1. Moreover, it is difficult to find appropriate IVs that are correlated with planted acres in the long rainy season as well as those in the short one, but are not correlated with the error term. Instead, I include total acres of owned land as covariates later.

of farmer i has a non-agricultural occupation. As explained in Background Information, a household seems to have reliable off-farm income during the short rainy season if its member has a non-agricultural occupation.[6] To examine heterogeneous impacts by the existence of off-farm income, I introduce the interaction terms with *off-farm income*$_{it}$ in Equation (4.4.2).[7] Therefore, Hypothesis 1, in which farmers without off-farm income benefit from improved caloric intakes, claims $\beta_1 > 0$ and $\beta_2 < 0$ if y_{it} is caloric intakes. In contrast, Hypothesis 2, in which farmers with off-farm income benefit from improved dietary diversity, claims $\beta_1 + \gamma_1 > 0$ and $\beta_2 + \gamma_2 < 0$ if y_{it} is the HDDS.

Empirical Issues

There are two empirical challenges in estimating the two equations above. First, S_{it} and S_{it}^2 are endogenous. Farmers tend to cultivate in the short rainy season in environmentally favorable conditions with bimodal rainfall patterns. Thus, we must control for areas of farmers' residence and their time-constant heterogeneity. Moreover, estimation results could reflect the reverse causality that high wealth levels lead to large production scales. Wealthy farmers with high levels of food consumption may own and cultivate large areas of land relative to poor farmers with low levels of food consumption. Consequently, the coefficients for S_{it} and S_{it}^2 could be estimated in favor of wealthy farmers with high levels of food consumption, so that the estimated summation of welfare benefits, $\hat{\alpha}_1 S_{it} + \hat{\alpha}_2 S_{it}^2$ in Equation (4.4.1), would tend to be large for wealthy farmers cultivating more land during the short rainy season. This problem is serious because poor farmers facing a lack of income in the lean season may obtain more benefits by cultivating in the short rainy season than wealthy farmers.

The second empirical challenge is that farmers may hold non-agricultural occupations as a result of their choices. Because my conceptual framework considers exogenous impacts of off-farm income, I need to control for farmers' choices to hold non-agricultural occupations.

Endogeneity for Cultivation in the Short Rainy Season

To control for endogeneity of S_{it} and S_{it}^2, I conduct the fixed-effects two-stage least squares (FE-2SLS) regression in Equation (4.4.1). I rely on meteorological findings to construct appropriate instrumental variables (IV). It is known that the rainfall variability in the short rainy season is closely linked to both El Niño and Indian Ocean Dipole (IOD) events (Mbigi and Xiao 2021). If the sea surface temperature (SST) in the tropical eastern Pacific Ocean is warmed more than usual (El Niño events), Tanzania tends to have more rainfall during

[6] As a robustness check, I also tighten the definition *off-farm income*$_{it}$ as a dummy variable that are equal to one if any household member is engaged in non-farm activities as a "main" occupation. I still obtain similar findings.

[7] Other covariates than S_{it} and S_{it}^2 are also interacted with *off-farm income*$_{it}$ because of the statistical structure of an ESR model, which I will use in estimation.

the short rainy season (Mutai and Ward 2000). In contrast, if the SST in the tropical Indian Ocean are cooled in the western part and warmed in the eastern part more than usual (negative IOD), droughts tend to occur during the short rainy season (Lu et al. 2018).

Two meteorological indices and their interaction term are introduced as IVs: NINO3.4 and the Dipole Mode Index (DMI), which measure the intensities of El Niño and IOD events, respectively.[8] Mbigi and Xiao (2021) showed that the rainfall in Tanzania in a given year has a significantly high connection with the NINO3.4 and DMI from September in the same year to February in the following year. I sum up the two indices over this period each year as IVs. I further introduce their interaction term as another IV.[9]

These meteorological phenomena affect the potential scale of production in the short rainy season, and seem to affect farmers' welfare only in this channel. Based on the NINO3.4 index and DMI, the Tanzanian government may announce the projection of rainfall during the short rainy season, which may affect farmers' decisions to cultivate in this season. However, I find no statistical significance in the Hansen's overidentification test, which will be shown later.

Because these meteorological indices vary over time but are common to all households, I also introduce as the last IV the proportion of farmers planting annual crops in the short rainy season for each district and wave.[10] The regional prevalence of the cultivation in the short rainy season seems to be correlated with farmers' practices to cultivate in this season, but does not seem to directly determine cultivation decisions by individual farmers. Thus, this IV is unlikely to be correlated with the idiosyncratic error ϵ_{it} in Equation (4.4.1). Although this IV may be correlated with the unobserved regional potential to cultivate in the short rainy season, I control for time-constant heterogeneity in the FE model and areas of farmers' residence as covariates, as will be explained later.

Appendix 4-II shows the first-stage results in the FE-2SLS model. The four IVs tend to be individually significant and are also jointly significant.

Self-Selection to Engage in Non-Agricultural Occupations

To control for farmers' choices to hold non-agricultural occupations, I estimate Equation (4.4.2) by using an endogenous switching regression (ESR) model proposed by Murtazashvili and Wooldridge (2016). This chapter classifies farmers into two groups: farmers with and without non-agricultural occupations. In the ESR, covariates have different impacts on the outcome

[8] The NINO3.4 index, which measures the anomaly of SST in the tropical eastern Pacific Ocean, is used to define El Niño events. The Dipole Mode Index (DMI), which measures the anomaly of the SST gradient between the western and eastern parts of the tropical Indian Ocean, represents the intensity of the IOD.

[9] Instead of these meteorological indices, I attempt to use rainfall amounts during the short rainy season as an IV. However, they are not significant in the first-stage estimation at the household level. Moreover, using the two meteorological indices and their interaction term enables me to use the Hansen's overidentification test because IVs outnumber two endogenous variables, S_{it} and S_{it}^2.

[10] The largest administrative unit in Tanzania is regions, followed by districts, wards, and village.

variable, depending on which group each farmer belongs to. In Equation (4.4.2), one unit increase of S_{it} changes y_{it} by β_1 if *off-farm income*$_{it} = 0$ and by $\beta_i + \gamma_i$ if *off-farm income*$_{it} = 1$. The ESR addresses a bias caused by farmers' self-selection into each group.

Murtazashvili and Wooldridge (2016) constructed an ESR with endogenous explanatory variables in panel data by combining the control function (CF) method with the correlated random effects (CRE). The CF method controls for the farmers' selection and endogenous explanatory variables, using a two-step procedure. In the first stage, I run the probit estimation of the selection indicator, or *off-farm income*$_{it}$ in Equation (4.4.2), on covariates and IVs, and obtain generalized residuals. In the second stage, I run the 2SLS regression of the outcome variable on endogenous explanatory variables, covariates, the generalized residuals, and the interaction terms between all these variables and the selection indicator. If the generalized residuals and their interaction term with the selection indicator are jointly significant in the second stage, it shows that the self-selection is correlated with ϵ_{it} in Equation (4.4.2) and is endogenous.

To control for time-constant heterogeneity, the CF approach is combined with the CRE method, which was introduced by Mundlak (1978) and Chamberlain (1984). I add the time-averaged values of exogenous explanatory variables and IVs to the set of covariates in each first and second stage estimation. This allows household's time-constant heterogeneity (μ_i) to be correlated with covariates, unlike the traditional random effect model.

In the ESR, I must introduce IVs for *off-farm income*$_{it}$, S_{it}, and S_{it}^2. I use the proportion of *off-farm income*$_{it}$ for each district and wave, plus the four IVs aforementioned.[11] The regional prevalence of non-farm activities seems to affect the possibility for each farmer to obtain non-agricultural occupations, but does not seem to directly determine types of occupations owned by individual farmers. Thus, this IV is unlikely to be correlated with the idiosyncratic error ϵ_{it} in Equation (4.4.2). Although this IV may be correlated with the unobserved potential of each regional economy, I control for time-constant heterogeneity in the CRE model. Appendix 4-II shows the results of the probit estimation in the first stage of the ESR model. Meteorological indices and the proportion of *off-farm income*$_{it}$ tend to be individually significant. The five IVs are also jointly significant.

4.4.4 Covariates

Household characteristics $\boldsymbol{x}_{it}$ include the household head's age, gender, educational years, and years of living in a community. I include adult equivalents as an indicator of household size and the total acres of land owned by a household. To represent the household's credit constraint, I add a dummy variable that are equal to one if a household belongs to a SACCO

[11] I calculate the proportion using all households in the NPS, which include non-farmers as well as farmers in Table 4.1.

(savings and credit cooperatives). I also add to covariates a distance to the nearest market from plots owned by a household.[12]

To control for the regional potential of agricultural production, I add the rainfall amounts in the long rainy season from March to May and in the short one from September to January in each district. I also add the sum of rainfall during the long rainy season and that during the short rainy season over the past five years in each district. These rainfall variables, which have large values, are transformed into logarithmic form in the analysis.

Because food consumption levels may depend on dates of interviews to households, I introduce three dummy variables that are equal to one if households are interviewed in January-March, April-June, and October-December, respectively, with July-September as the baseline. January-March corresponds to the harvest period after the short rainy season, April-June to the cultivation period during the long rainy season, July-September to the harvest period after the long rainy season, and October-December to the cultivation period during the short rainy season. If farmers do not cultivate in the short rainy season, October-December and January-March corresponds to the lean season.

To control for the regional suitability to cultivate in the short rainy season, I introduce four dummy variables that are equal to one if households reside in Northeast, Coastal, Central, and Lakes areas, respectively, with Southern Highlands as the baseline. Finally, I introduce wave dummies, with 2008/09 as the baseline.[13]

4.4.5 Descriptive Statistics

Table 4.3 displays mean values of food consumption levels, household characteristics, and rainfall variables in four cases: whether farmers cultivate in the short rainy season and whether some members within a household have occupations except agriculture. Table 4.3 shows that daily caloric intakes per adult equivalent tend to be higher for farmers who have only agricultural occupations, whereas the HDDS tends to be higher for those who have non-agricultural occupations. Farmers without off-farm income may have higher caloric intakes through the self-consumption of staple crops with high calories (e.g., maize). In contrast, farmers with off-farm income may purchase more food, which improves their dietary diversity.[14] Although farmers with non-agricultural occupations tend to have lower caloric intakes, they tend to be wealthier than farmers without non-agricultural occupations. The average value of the total consumption including food and non-food items per 28 days is

[12] Regional grain prices may affect households' food consumption. However, I omit grain prices because there are missing values in some regions and years. For robustness, if I fill in missing values by predicting maize prices by region/year fixed-effects regression, I still obtain similar findings.

[13] Following the practice of Murtazashvili and Wooldridge (2016), I do not introduce time-averaged values of area and wave dummies in the ESR model.

[14] In my dataset, the average ratio of the purchased food consumption to the total food consumption is 43% if farmers have non-agricultural occupations, whereas it is 36% if they do not.

Table 4.3: Summary Statistics for Farmers Divided by the Cultivation in the Short Rainy Season and the Existence of Non-Agricultural Occupations

	Cultivate in Short Rain		Not Cultivate in Short Rain	
Mean Values	Own only agri. occupation	Own nonagri. occupation	Own only agri. occupation	Own nonagri. occupation
Household food consumption:				
Daily caloric intake per adult equivalent (kcal)	2963.8	2700.65	2968.03	2759.23
Weekly number of consumed food types (HDDS)	8.18	8.5	7.56	8.31
Household characteristics:				
Acres of cultivated land in short rain	2.51	2.45	0.00	0.00
Age of household head	53.86	46.98	53.28	46.61
Male-headed household (dummy)	0.73	0.79	0.70	0.76
Years of education of household head	4.07	5.08	3.74	5.12
Living years in community by household head	42.45	34.85	40.40	32.76
Adult equivalents	4.66	5.09	4.26	4.47
Acres of owned land	5.13	4.90	6.95	5.38
Belongs to SACCO (dummy)	0.03	0.07	0.01	0.05
Nearest distance to market from plots (km)	7.98	7.23	8.68	8.57
Climatic characteristics:				
Rainfall during long rainy season (mm)	438.62	434.1	385.98	399.01
Rainfall during short rainy season (mm)	483.68	504.64	449.14	442.04
Sum of long rain over past 5 years (mm)	2193.77	2203.58	1928.76	1975.65
Sum of short rain over past 5 years (mm)	2227.92	2396.02	2181.18	2156.89
Observations	979	1,798	2,760	4,818

3,595,405 Tanzanian shillings for the former, whereas it is 2,878,717 shillings for the latter, although this is not displayed in the table.

Farmers cultivate around 2.5 acres on average in the short rainy season, regardless of whether they have non-agricultural occupations, which suggests that production scales in the short rainy season are not related to the existence of off-farm income. In contrast, other household characteristics seem to be related to whether farmers have non-agricultural occupations. If household members have non-agricultural occupations, their household heads tend to be younger and more educated males who have lived in a community in shorter periods. Farmers with non-agricultural occupations tend to have more household members and own less land. They are more likely to belong to a SACCO and own plots near markets.

Farmers who cultivate in the short rainy season tend to have more rainfall during both long and short rainy seasons, which suggests that farmers can cultivate in the short rainy season on climatically favorable conditions.

4.5 Results

Using the household-level dataset in Table 4.1, Table 4.4 displays estimation results of FE and FE-2SLS models in Equation (4.4.1) and FE and ESR models in Equation (4.4.2).[15]

[15] As a caveat, I use the whole household dataset in Table 4.1 and do not restrict it to Northeast/Lakes areas in this estimation.

Estimating the FE model is helpful to consider how the endogeneity of S_{it} could affect my results, as well as to check the robustness of my results. Standard errors (SE) are clustered at the household level in the FE model; robust SE are used in the FE-2SLS model; and SE are obtained using bootstrap with repetitions 500 in the ESR model to adjust the standard errors in the second stage of ESR for the first-stage estimation (Murtazashvili and Wooldridge 2016). For brevity, Table 4.4 omits other covariates than acres of cultivated land in the short rainy season, their squared term, generalized residuals, and the interaction terms between these variables and *off-farm income*$_{it}$. For full results, see Appendix 4-II.

Columns (3) and (4) show that overidentification tests using the Hansen's J statistic are insignificant. Thus, I find no evidence that assumptions for IV including exclusion restrictions are violated in the FE-2SLS model. Columns (7) and (8) show results of the Wald test on generalized residuals and their interaction term with *off-farm income*$_{it}$. Because they are not jointly significant, I find no evidence that *off-farm income*$_{it}$ is endogenous.

Homogeneous impacts without interaction terms

Using the FE model in Equation (4.4.1), Columns (1) and (2) show that one additional acre of land cultivated during the short rainy season is significantly positively associated with farmers' caloric intakes (kcal) and HDDS by 28.57 and 0.03, respectively. Because the average level of S_{it} is around 2.50 acres in Table 4.3, the cultivation in the short rainy season seems to have limited welfare improvements for farmers suffering from large deficits in nutrition (For comparison, Dercon et al. (2009) sets 2,300 kcal per adult per day as a poverty line). Moreover, the coefficients for S_{it}^2 are insignificant in both columns, although they are negative.

In contrast, estimated impacts in the FE-2SLS model are larger for smallholders. In Columns (3) and (4), the estimated coefficients for S_{it} is 1008.41 in caloric intakes (kcal) and 0.89 in HDDS, which is even larger than those in the FE model. Moreover, the estimated coefficients for S_{it}^2 are significantly negative in both columns. Controlling for endogeneity in the FE-2SLS model may exclude the reverse causality that the estimated coefficients for S_{it} and S_{it}^2 reflect high levels of food consumption among wealthy farmers, as explained in Estimation Strategy.

To interpret the magnitudes of estimated coefficients, the upper side of Figure 4.6 shows the sample distribution of S_{it}, with their quantiles depicted by red-dotted lines (10th, 25th, 50th, 75th, and 90th). In Figure 4.6(a), S_{it} has a right-skewed distribution and concentrates from zero to two acres. Using estimated coefficients in Table 4.4, the lower side of Figure 4.6 displays estimated increases in caloric intakes and HDDS, that is, $\hat{\alpha}_1 S_{it}$ in the FE model and $\hat{\alpha}_1 S_{it} + \hat{\alpha}_2 S_{it}^2$ in the FE-2SLS model. Figures 4.6(b) and 4.6(c) show that the increase in food consumption is gradual in the FE model and tends to be larger for large-scale farmers.

Table 4.4: Estimates of Welfare Impacts of the Cultivation in the Short Rainy Season

	Homogeneous effects without interactions				Heterogeneous effects with interactions			
	FE		FE-2SLS		FE		ESR	
Explanatory Variable	(1) Kcal	(2) HDDS	(3) Kcal	(4) HDDS	(5) Kcal	(6) HDDS	(7) Kcal	(8) HDDS
Variables non-interacted with *off-farm income*$_{it}$:								
S_{it} (acres of cultivated land in short rain)	28.57*	0.03*	1008.41*	0.89**	50.08*	0.02	1058.38*	0.25
	(0.09)	(0.07)	(0.06)	(0.03)	(0.06)	(0.49)	(0.10)	(0.68)
S_{it}^2 (squared term of acres of cultivated land in short rain)	-0.69	-0.00	-97.57*	-0.07*	-1.81*	-0.00	-105.96*	-0.01
	(0.29)	(0.20)	(0.05)	(0.06)	(0.08)	(0.81)	(0.07)	(0.86)
gr_{it} (generalized residuals obtained in probit estimation)							616.22*	-0.08
							(0.07)	(0.81)
Interaction terms with *off-farm income*$_{it}$:								
S_{it} × *off-farm income*$_{it}$					-32.98	0.01	-521.96	0.87
					(0.28)	(0.71)	(0.52)	(0.31)
S_{it}^2 × *off-farm income*$_{it}$					1.56	-0.00	52.41	-0.08
					(0.19)	(0.63)	(0.47)	(0.25)
gr_{it} × *off-farm income*$_{it}$							-792.14*	0.24
							(0.06)	(0.60)
Sum of coefficients when *off-farm income*$_{it}$= 1:								
$\beta_1+\gamma_1$ (sum of coefficients for S_{it})					17.10	0.03	536.42	1.12*
					(0.37)	(0.10)	(0.26)	(0.07)
$\beta_2+\gamma_2$ (sum of coefficients for S_{it}^2)					-0.25	-0.00	-53.55	-0.09*
					(0.73)	(0.19)	(0.18)	(0.06)
Control for other covariates:								
Household characteristics	YES	YES	YES	YES	YES	YES	YES	YES
Dummies of interview period	YES	YES	YES	YES	YES	YES	YES	YES
Wave dummies	YES	YES	YES	YES	YES	YES	YES	YES
Area dummies	YES	YES	YES	YES	YES	YES	YES	YES
Overidentification test								
Hansen's J statistic (p-value in parenthesis)			0.28	3.45				
			(0.87)	(0.18)				
Wald test on generalized residuals								
Wald statistic (p-value in parenthesis)							3.67	0.31
							(0.16)	(0.86)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use standard errors clustered at the household level in the FE model, robust standard errors in the FE-2SLS model, and standard errors obtained by bootstrapping with 500 repetitions in the ESR model. For other covariates, see Appendix 4-II.

In contrast, the additional gain in the FE-2SLS model is considerably large for smallholders, shrinks as farmers cultivate more land, and reaches the maximum at 5.17 acres in caloric intakes and at 6.27 acres in the HDDS. Because S_{it} concentrates up to two acres, resource-poor farmers with less land seem to benefit more by cultivating additional land in the short rainy season. For example, the median farmer who cultivate 1.68 acres in the short rainy season (see 50th in Figure 4.6(a)) consume 1,422 more kilocalories and 1.30 more food types.

Heterogeneous impacts with interaction terms

Using the FE model in Equation (4.4.2), Column (5) in Table 4.4 shows that the coefficient for S_{it} is significantly positive (50.08) and outnumbers that obtained in the FE model without interaction terms, or 28.57 in Column (1). In contrast, the estimated coefficient for $S_{it} \times \textit{off-farm income}_{it}$ is negative. Consequently, the sum of coefficients for S_{it} when $\textit{off-farm income}_{it} = 1$, or $\beta_1 + \gamma_1$, is insignificant in the F test. Because the coefficient for S_{it}^2 is significantly negative, the inverse relationship between farmers' welfare and farm size holds for farmers without non-agricultural occupations. In Column (6), the p-value for $\beta_1 + \gamma_1$ in the F test is considerably close to 0.10 (0.1006). These results provide evidence consistent with my two hypotheses. The cultivation in the short rainy season improves caloric intakes of farmers without non-agricultural occupations and dietary diversity of farmers with non-agricultural occupations.

The estimation results of the ESR model in Equation (4.4.2) also support my two hypotheses. Column (7) shows that the coefficient for S_{it} is significantly positive and considerably outnumbers 50.08 in Column (5). The level of increase, 1058.38, is similar to that in the FE-2SLS model (1008.41). Because the estimated coefficient for S_{it}^2 is significantly negative, smallholders without non-agricultural occupations gain more calories by cultivating additional land in the short rainy season than large-scale farmers do. However, the estimated coefficient for $S_{it} \times \textit{off-farm income}_{it}$ is negative. Consequently, I find no evidence that farmers with non-agricultural occupations gain more calories by cultivating in the short rainy season, because the estimated value of $\beta_1 + \gamma_1$ is insignificant.

In contrast, Column (8) shows that the estimated coefficient for S_{it} is insignificant. However, the estimated sum of coefficients for S_{it} when $\textit{off-farm income}_{it} = 1$, or $\beta_1 + \gamma_1$, is significantly positive (1.12), and is relatively close to 0.89 in the FE-2SLS model. Because the estimated value of $\beta_2 + \gamma_2$ is significantly negative, smallholders with non-agricultural occupations realize more dietary diversity by cultivating additional land in the short rainy season than large-scale farmers do.

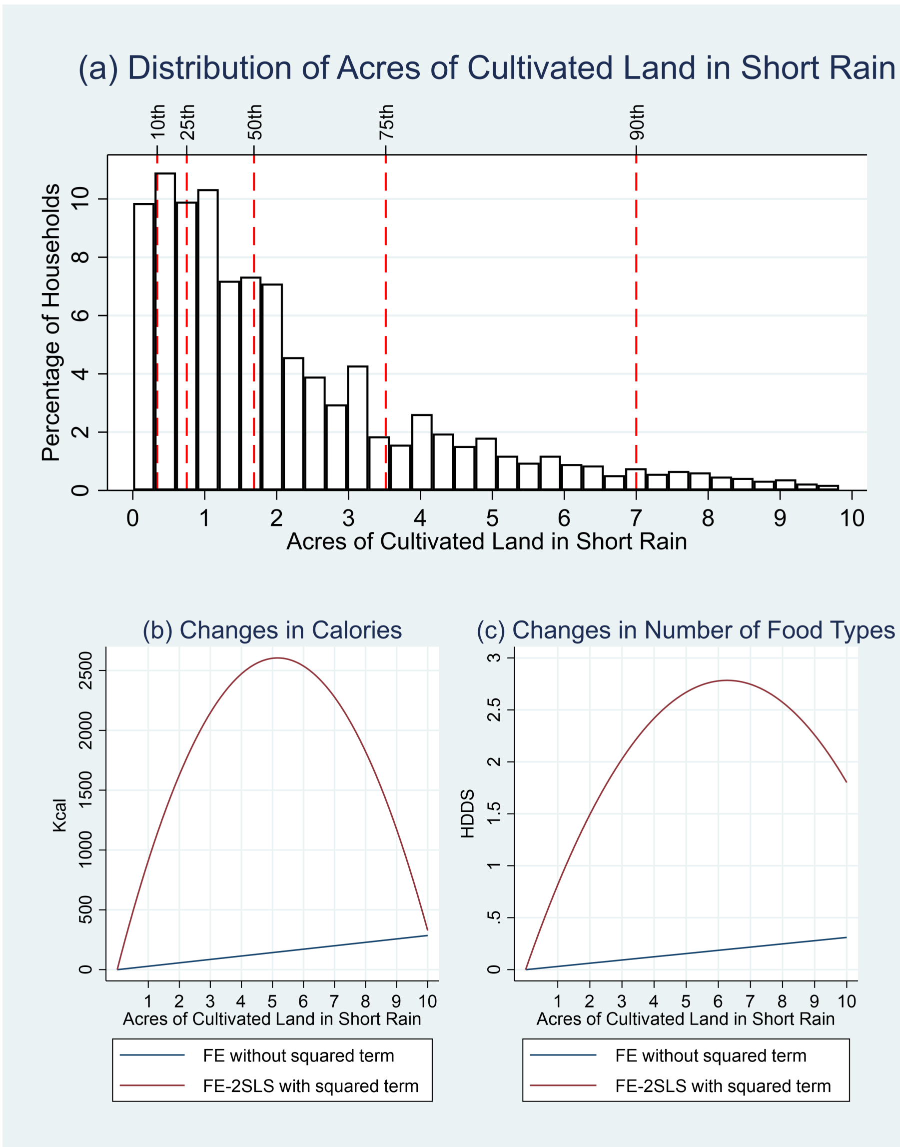


Figure 4.6: Distribution of S_{it} and Changes in Food Consumption

4.6 Discussion

4.6.1 Channels of Welfare Improvements

I then examine a path of farmers' welfare improvements in more details. Farmers who cultivate in the short rainy season may realize welfare improvements through the self-consumption of harvested crops. In contrast, welfare improvements may last over the year through the purchase of food if farmers obtain cash income by selling crops in the short rainy season.

Food Consumption Levels by Months of Interviews

Concerning the previous main analysis, Appendix 4-II shows that farmers who are interviewed in October-December and January-March tend to have less calories and HDDS than those interviewed in July-September, the harvest period after the long rainy season (see Columns (1)-(4) in Table A15). Because food consumption data in the NPS are based on the information over the past week before the interview, this result suggests that food securities tend to deteriorate from the end of the harvest period in the long rainy season to its start.

Figure 4.7 plots the average levels of farmers' caloric intakes (kcal) and HDDS by months of interviews in Northeast/Lakes areas (I restrict my attention to Northeast/Lakes areas to control for regional heterogeneity). In Figure 4.7(a), farmers who cultivate in the short rainy season tend to have more calories from January to April and have more HDDS from October to December than those who do not. Although it is unclear why farmers' caloric intakes and dietary diversity improve in different periods, the cultivation in the short rainy season seems to improve farmers' food consumption from the end of the harvest period of the long rainy season to its start, as I expected.

Crop Sales Revenue and Food Consumption from Own Production/Purchase

I then examine whether the cultivation in the short rainy season increases the sum of crop sales revenue over the two rainy seasons within a year and food consumption values per adult equivalent per week coming from own production and purchase.[16] I estimate Equation (4.4.2) by the pooled CRE Tobit model because some farmers do not sell crops or purchase food. Because crop sales revenue and food expenditure are large monetary values in Tanzanian shillings and the Tobit model is susceptible to the violation of the normality of the error term (Cameron and Trivedi, 2010), I transform these values in logarithm after I add one to the original values.

[16] To calculate food consumption values, I follow the same procedure as Dolislager et al. (2022). The NPS provides information on purchased quantities of food and their expenditure over the past week before the interview. I first calculate market prices of each food by dividing its expenditure by the purchased quantity. Then, I multiply these market prices by total quantities of consumed food coming from own production and purchases.

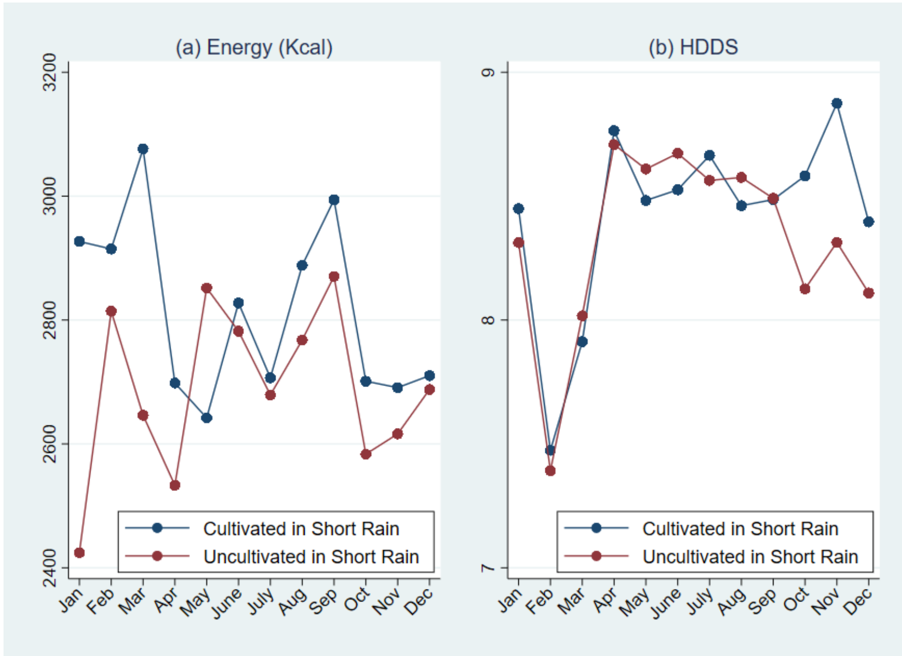


Figure 4.7: Average Calories (Kcal) and HDDS by Month of Interview in Northeast/Lakes

Table 4.5 displays average partial effects (APE) on sales revenue and food consumption values from own production and purchase. I use the same covariates as my main analysis in Equation (4.4.2), although they are omitted in the table. In Column (1), the APE for S_{it} is positive and relatively large (14%) among farmers without non-agricultural occupations, although it is insignificant. Farmers without non-agricultural occupations might obtain more revenue from crop sales by cultivating in the short rainy season. In contrast, the sum of APEs for S_{it} when *off-farm income*$_{it} = 1$ is significantly positive and almost doubles (0.27), which suggests that farmers with non-agricultural occupations gain more cash income by selling crops harvested during the short rainy season.

In Column (2), the average farmers without non-agricultural occupations consume 13% more values of food coming from own production if they cultivate one more acre of land in the short rainy season. Although farmers with non-agricultural occupations also consume more values of food from own production, its increase (6%) is low relative to farmers without non-agricultural occupations. In Column (3), both farmers with and without non-agricultural occupations purchase 3% more food by cultivating one more acre of land in the short rainy season. Farmers may spend increased revenue of crops sales on purchasing food. Importantly, the increase in the consumption from purchased food (3%) is lower than that from own production (13%) if farmers do not have non-agricultural occupations.

These results suggest that farmers without off-farm income improve their welfare mainly through the self-consumption of harvested crops, whereas farmers with off-farm income improve their welfare by using the increased revenue of crop sales. The latter farmers may

Table 4.5: Pooled CRE Tobit Estimation of the Log of Sales Revenues over the Two Rainy Seasons and Food Consumption from Own Production/Purchase in Tanzanian shillings

Explanatory Variable	(1) Sales Revenues over Two Rainy Seasons	(2) Food Consumption from Own Production	(3) Food Consumption from Purchase
Variables non-interacted with *off-farm income*$_{it}$:			
S_{it} (acres of cultivated land in short rain)	0.14	0.13^{***}	0.03^{*}
	(0.17)	(0.00)	(0.08)
S_{it}^2 (squared term of S_{it})	-0.00	-0.00^{***}	-0.00^{*}
	(0.69)	(0.00)	(0.10)
Interaction terms with *off-farm income*$_{it}$:			
$S_{it} \times$ *off-farm income*$_{it}$	0.13	-0.07^{*}	-0.01
	(0.26)	(0.10)	(0.75)
$S_{it}^2 \times$ *off-farm income*$_{it}$	-0.01	0.00^{*}	0.00
	(0.23)	(0.06)	(0.30)
Sum of coefficients when *off-farm income*$_{it}$= 1:			
$\beta_1 + \gamma_1$ (sum of coefficients for S_{it})	0.27^{***}	0.06^{*}	0.03^{**}
	(0.00)	(0.06)	(0.04)
$\beta_2 + \gamma_2$ (sum of coefficients for S_{it}^2)	-0.01^{*}	-0.00^{*}	-0.00
	(0.08)	(0.06)	(0.43)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the household level. APEs are displayed in the table. Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for areas, waves, and interview periods, holding of non-farm occupation and its interaction terms with covariates aforementioned, time-averaged values.

also improve their dietary diversity through the self-consumption, as Column (2) indicates. This observation is in line with Appendix 4-IV, which shows that farmers who have non-agricultural occupations are more likely to plant vegetables/fruits instead of planting sorghum/roots/tubers and cash crops than those who do not.

4.6.2 Determinants of the Double-Cropping over the Two Rainy Seasons

Finally, I examine farmers' decisions to cultivate their plots in the short rainy season in addition to the long one. Even if farmers can obtain more income by cultivating in both rainy seasons, they may be unable to do so owing to some factors. Clarifying promoting and constraining factors is helpful to consider developmental policies.

Using the CRE probit model at the plot-level dataset, I examine the probability that farmers plant annual crops in a plot in both two rainy seasons within a year (*double-cropping*). Because we use the plot level dataset in Table 4.1, we control for plot characteristics as well as household ones (Some plots are excluded in the analysis because of missing values in some covariates). Table 4.6 displays estimation results and covariates used in this analysis. I cluster SE at the plot level. Column (1) displays estimation results using all sample, whereas Columns (2) and (3) display results using households in Northeast/Lakes and Coastal/Central/Southern Highland areas, respectively.

Table 4.6: CRE Probit Estimates of Determinants for Planting Annual Crops in a Plot in Both Long and Short Rainy Seasons

Explanatory Variable	(1) All Sample	(2) Northeast/Lakes	(3) Coastal/Central/Southern
Household Characteristics:			
Age of household head	0.00^{*}	0.00	0.00^{***}
	(0.09)	(0.98)	(0.00)
Male-headed household (dummy)	0.02	0.04	0.01
	(0.21)	(0.32)	(0.61)
Years of education of household head	-0.00	-0.00	0.00
	(0.93)	(0.47)	(0.57)
Years of living in community by household head	-0.00^{***}	-0.00^{*}	-0.00^{**}
	(0.00)	(0.06)	(0.03)
Adult equivalents	-0.00	-0.00	0.00
	(0.61)	(0.30)	(0.38)
Belongs to SACCO (dummy)	0.02^{*}	0.04	0.01
	(0.09)	(0.12)	(0.50)
Receives subsidized fertilizer (dummy)	0.05^{***}	0.09^{*}	0.03^{***}
	(0.00)	(0.07)	(0.01)
Plot Characteristics:			
Plot is irrigated (dummy)	0.07^{***}	0.11^{***}	0.06^{***}
	(0.00)	(0.00)	(0.00)
Acres of plot	-0.00	-0.00	-0.00
	(0.22)	(0.19)	(0.76)
Erosion (dummy)	0.01	0.03^{*}	0.00
	(0.13)	(0.07)	(0.95)
Distance from plot to home (km)	-0.02^{***}	-0.04^{***}	-0.01^{*}
	(0.00)	(0.00)	(0.09)
Distance from plot to road (km)	0.00	0.00	-0.00
	(0.65)	(0.93)	(0.82)
Distance from plot to market (km)	0.01^{***}	0.00	0.01^{***}
	(0.00)	(0.65)	(0.00)
Plot is owned by a household (dummy)	0.01^{**}	0.02	0.01^{*}
	(0.05)	(0.18)	(0.10)
Soil type: sandy (dummy)	0.21^{***}	0.40^{***}	0.09^{***}
	(0.00)	(0.00)	(0.00)
Soil type: loam (dummy)	0.21^{***}	0.41^{***}	0.09^{***}
	(0.00)	(0.00)	(0.00)
Soil type: clay (dummy)	0.20^{***}	0.40^{***}	0.09^{***}
	(0.00)	(0.00)	(0.00)
Geographic Characteristics:			
Rainfall during long rainy season (mm) (log)	0.03^{**}	0.03	0.02^{*}
	(0.01)	(0.32)	(0.08)
Rainfall during short rainy season (mm) (log)	0.01	-0.01	0.01
	(0.60)	(0.66)	(0.26)
Sum of long rain over past 5 years (mm) (log)	0.11^{***}	0.23^{***}	0.08^{***}
	(0.00)	(0.01)	(0.00)
Sum of short rain over past 5 years (mm) (log)	-0.01	-0.06	-0.00
	(0.78)	(0.39)	(0.95)
Control for other covariates:			
Wave dummies	YES	YES	YES
Area dummies	YES	YES	YES
Sample Size:			
N	21,348	7,760	13,588

Note: Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the plot level. Time-averaged values are included in the analysis but omitted in the table.

In Column (1) of Table 4.6, I focus my attention on three significant results, which are suggestive of policy implications. First, farmers are more likely to cultivate a plot in both seasons by 5 percentage points (pp) if they receive subsidized fertilizer. The government of Tanzania implemented the large-scale fertilizer subsidy, called the National Agricultural Input Voucher Scheme (NAIVS), since 2008/09. Because inorganic fertilizer plays important roles in supplying nutrients to the soil, the fertilizer subsidy may enable farmers to continuously use plots without fallowing during the two rainy seasons. Second, the probability of planting in both seasons increases by 7 pp if a plot is irrigated. In the short rainy season, rainfall tends to be lower and more uncertain than in the long one, as described in Appendix 4-I. Water management by irrigation may enable farmers to address low rainfall and the interannual variability of short rains. Third, the ownership to a plot increases the probability of the double-cropping, although the increase is small (1 pp). If a plot is owned by others, farmers may not be able to freely determine a period of cultivation.

Columns (2) and (3) show that I still obtain almost the same findings even If I divide the sample to households in Northeast/Lakes and Coastal/Central/Southern Highland areas. The probability increases by the receipt of the fertilizer subsidy and irrigation in Northeast/Lakes areas (9% and 11%) outnumber those in Coastal/Central/Southern areas (3% and 6%). In contrast, the ownership in Northeast/Lakes areas is not significant unlike Coastal/Central/Southern areas. Thus, we may still be able to promote the double-cropping in Coastal/Central/Southern areas, although policy effects may be low relative to Northeast/Lakes areas. In the household-level dataset, farmers in Northeast/Lakes and Coastal/Central/Southern areas cultivate 2.66 and 1.86 acres on average, respectively, if they cultivate in the short rainy season. Thus, the scale of the expansion of the double-cropping may not be small when farmers who are encouraged to cultivate in the short rainy season by policies start cultivating in this season.

4.7 Conclusion

This chapter has examined welfare impacts of the cultivation in the short rainy season, with a focus on farmers' food securities. In Background Information, I showed that the cultivation in the short rainy season is not a minor or special economic activity. More than half of the Tanzanian farmers cultivate in the short rainy season in the Northeastern part and regions around Lake Victoria. The composition of planted crops and agricultural productivity are also similar between the two rainy seasons. These observations suggest that the cultivation in the short rainy season plays important roles in farmers' livelihoods. In Conceptual Framework, I introduced the subsistence constraint into the customary consumer problem to analyze welfare impacts of additional agricultural income in the lean season. Farmers who have no off-farm income are expected to spend additional income on increasing their caloric intakes for survival. In contrast, farmers who have off-farm income are expected to spend additional income on increasing their dietary diversity.

My empirical results showed that farmers can obtain more caloric intakes and realize more dietary diversity by cultivating in the short rainy season. I also found consistent results with my theoretical hypotheses. Farmers who have no occupation except agriculture obtain more calories by cultivating in the short rainy season, whereas farmers who have non-agricultural occupations obtain more dietary diversity by the cultivation. Importantly, resource-poor farmers with small farm size benefit more per acre of planted land in the short rainy season than resource-rich farmers with large farm size do.

My determinant analysis provided three policy implications for the promotion of the double-cropping over the two rainy seasons. First, policies to maintain soil nutrients may enable farmers to cultivate without fallowing in the short rainy season, because beneficiaries for the Tanzanian fertilizer subsidy, NAIVS, are more likely to cultivate in both rainy seasons within a year. Second, water management during the short rainy season plays crucial roles in addressing low and uncertain rainfall in this season. The probability of the double cropping increased in an irrigated plot by 7 pp. Third, policies to guarantee the land ownership may promote the double cropping. It is known that land rights affect farmers' decisions to cultivate or fallow their land (e.g., Goldstein and Udry 2008). In my analysis, farmers are more likely to cultivate owned plots over the two rainy seasons.

There are two limitations in this study. First, the cultivation in the short rainy season is strongly constrained by bimodal rainfall patterns. Thus, policies relying on the short rainy season may be applicable only to environmentally favorable areas. Further research is required to clarify the range of applicability of this study and examine how to expand the double-cropping in environmentally disadvantaged areas. Second, fertilizer subsidies and irrigation schemes, which were suggested as policy implications, may be costly and difficult to introduce in Sub-Saharan Africa. Future research should explore farm management practices, in which farmers can maintain nutrients in the soil and efficiently use low rainfall by using local resources.

Chapter 5

Farmers' Livelihood Strategies to Cultivate in a Secondary Cropping Season for the Schooling of Children: Concepts and Evidence from Tanzania

5.1 Introduction

Promoting the schooling of children has been a key issue in rural areas of low-income countries. In the face of climatic and economic shocks, rural households often increase labor of their children at the expense of their schooling (Jensen 2000; Amin et al. 2006; Duryea et al. 2007; Gitter and Barham 2007; Dammert 2008; Alvi and Dendir 2011; Cogneau and Jedweb 2012; Dillon 2013; Bandara et al. 2015; Emerson et al. 2017; Martey et al. 2023). Moreover, poor households suffer from the lack of income and credit for educational expenses (Thomas et al. 2004; Ersado 2005; Beegle et al. 2006; Hazarika and Sarangi 2008; Dillon 2021; Martey et al. 2021).

Is it possible for rural households to address the lack of educational expenses by themselves? Previous studies have considered policy tools such as cash transfer and the provision of loan and information (Schultz 2004; de Janvry et al. 2006; Edmonds 2006; Shimamura and Lastarria-Cornhiel 2010; Benhassine et al. 2015; Benedetti et al. 2016; Muralidharan and Prakash 2017; Barrera-Osorio et al. 2019; Baird et al. 2019; Millán et al. 2020; Premand and Barry 2022; de Walque and Valente 2023; Mendola and Negasi 2023; Gazeaud and Ricard 2024). However, if rural households have local ways to address the lack of cash, we may be able to expand the schooling of children more effectively by supporting the adoption of such livelihood strategies.

To explore this possibility, this chapter aims to clarify that the double-cropping enables the schooling of children among Tanzanian farmers. In East Africa, the bimodal rainfall pattern is prevalent and offers two cropping seasons, the long rainy season as a main cultivation period and the short rainy season as a secondary cultivation period (Mutai and Ward 2000; Waha et al. 2020). In Tanzania, most farmers cultivate in the long rainy season (*masika*) from March to May and harvest crops in July and August. The short rainy season begins in September/October and continues through December/January. The short rainy season provides an additional opportunity for farmers to produce food before the start of the long rainy season (Oestigaard 2016). Thus, farmers who cultivate in the short rainy season may be able to cover educational expenses for the schooling of their children, using additional crop sales. In fact, Ikeno (2011) reported through his field observations that Tanzanian farmers attempted to cultivate by one more season to cover new expenses for public services under the Structural Adjustment Policy.

This chapter has two main objectives. First, I examine whether the double-cropping, or the cultivation in both long and short rainy seasons, promotes the schooling of children among Tanzanian farmers. To estimate impacts on education, this chapter focuses on children's enrollment in pre-primary, primary, and lower-secondary schools, which are popular ways of schooling in Tanzania. As another educational indicator, I also focus on the number of children who continue to attend these schools without dropping out. Second, I examine whether the double-cropping has heterogeneous impacts by economic levels. In low-income countries, educational attainments of children are strongly restricted by households' economic levels (Jacoby 1994; Mani et al. 2013; Martey et al. 2021). If the double-cropping improves the children's education stronger for poor farmers than for wealthy farmers, the double-cropping works as a useful tool to reduce educational inequality. Thus, I estimate how the effects of the double-cropping change over households' wealth levels.

My dataset comes from six waves from 2008/09 to 2020/21 in the Tanzania National Panel Survey (NPS). Combining correlated random effects (CRE) probit/Tobit models with the control function (CF) approach, I find that the double-cropping increases the probability of children's enrollment in primary and lower-secondary schools. This effect is stronger for asset-poor farmers than for asset-rich farmers. Moreover, I find gender-differential effects; farmers who cultivate in both rainy seasons make their girls, rather than boys, enroll in primary and lower-secondary schools. In contrast, I find that the double-cropping promotes boys to continue to attend primary schools without dropping out.

This chapter makes three contributions to the literature. First, this study explores a new policy tool to promote children's education in low-income countries. To my knowledge, this paper is the first study to empirically link the double-cropping to the schooling of children.

Second, this study adds new evidence on economic roles of the short rainy season in

East Africa. To my knowledge, there are few empirical studies focusing on the short rainy season because previous studies often concentrated on analyzing farmers' activities in the long rainy season (Bevis et al 2019; Bhargava et al. 2019). A few studies, as parts of their main analyses, investigated welfare impacts of seasonal droughts in the short rainy season and farmers' perceptions on its long-run rainfall variability in Ethiopia (Holden and Shiferaw 2004; Dimitrova 2021; Kassie et al. 2013). However, most these studies neither focused on the short rainy season as a main research topic nor examined its economic roles. In Chapter 4, I showed that Tanzanian farmers improved their food security by cultivating in the short rainy season. This study adds new evidence by illuminating another economic role of the short rainy season: the improvement of children's education.

Third, this chapter suggests new evidence that the double-cropping may be favorable for vulnerable people such as poor households and females. Chapter 4 also showed that resource-poor farmers with small farm size improved their food consumption per acre of cultivated land in the short rainy season more than resource-rich farmers do. This chapter shows that the double-cropping also reduces educational inequality and promotes female education.

The rest of this chapter is organized as follows. Section 5.2 provides the background information on the children's education and short rainy season in Tanzania. Section 5.3 describes the conceptual framework. Section 5.4 explains my dataset and estimation strategy. Section 5.5 presents estimation results. Section 5.6 discusses my analytical results. Finally, Section 5.7 concludes this chapter.

5.2 Background Information

5.2.1 Educational System and its Transformation in Tanzania

Figure 5.1 depicts the formal structure of education in Tanzania. Until 2015, the Tanzanian formal education followed the 2-7-4-6-3+ system, with each number denoting years of schooling required for completing pre-primary schools, primary schools, lower-secondary schools, upper-secondary schools, and universities, respectively (3+ means three or more years) (Ministry of Education, Science and Technology (MoEST) 2017). Primary schools for children aged 7-13 years were compulsory and fee-free, as indicated by the red-colored side in the "Primary" box in Figure 5.1. To enroll in lower-/upper-secondary schools and universities, students are required to pass the Primary School Leaving Examination (PSLE), the O+ exam, and the A+ exam, respectively.[1] Students can enter the MS+ course (vocational courses such as carpentry) after primary schools (NBS 2022). After the completion of lower and upper secondary schools, there are also the Certificate course and the Diploma course,

[1] However, students who do not pass the PSLE can enter lower-secondary schools if they are private schools.

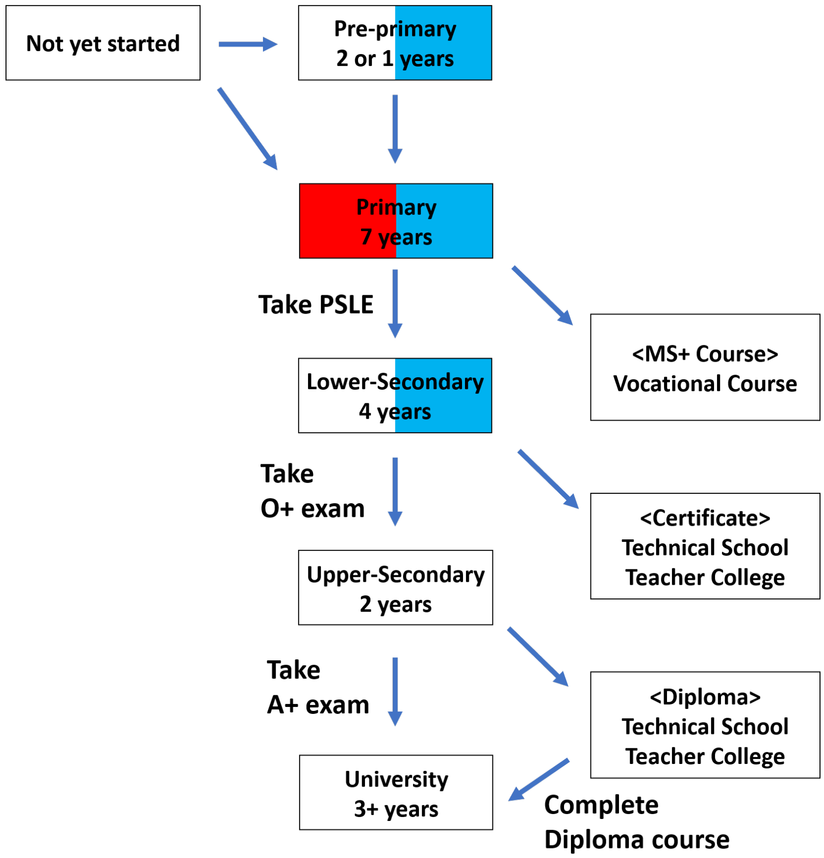


Figure 5.1: Formal Educational Structure in Tanzania. Schooling was compulsory and fee-free before 2015 if a box is red-colored and after 2016 if it is blue-colored.

respectively, which comprise technical schools and teacher colleges aimed at training teachers at primary and lower-secondary levels. Students who complete the Diploma course are allowed to enter a university.

After the "Education and Training Policy" was introduced in 2014, pre-primary and lower-secondary schools were made compulsory and fee-free in 2016, as indicated by the blue-colored sides in the three boxes in Figure 5.1. Schooling years for pre-primary schools were changed from two to one year.

Expanding fee-free education enabled high enrollments in pre-primary schools. The gross enrollment rate (GER) for pre-primary schools increased from 39.5% in 2010 to 102.6% in 2016 (MoEST 2017). [2] Although the GER for primary schools was 96.9% in 2017, the impact of expanding fee-free education was limited in lower-secondary school: its GER was around 40% in 2017. The GERs were even smaller for non-compulsory education. The GRE was less than 10% for upper-secondary schools in 2017.

[2]The GER is the ratio of the number of students enrolling in schools to the number of children in the corresponding age group of the school enrollment. The GER can be more than 100% if many students younger or older than specified school age enroll in schools.

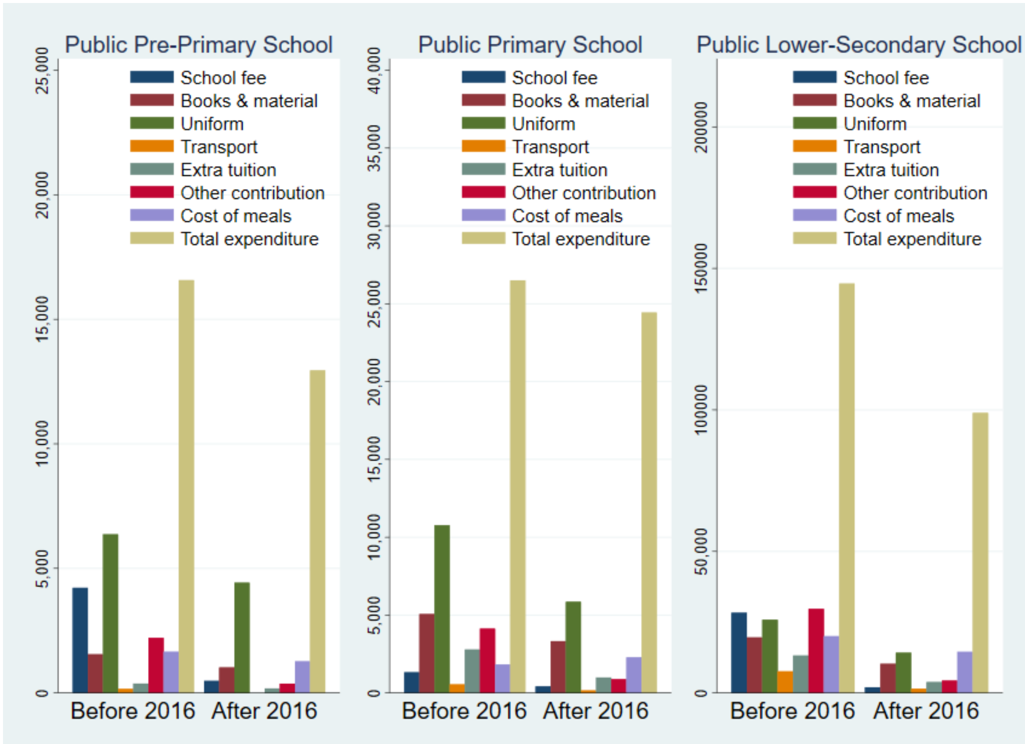


Figure 5.2: Average Educational Expenditure per Student in Tanzanian shillings

5.2.2 Educational Expenditure

Using six waves from 2008/09 to 2020/21 in the NPS, I then examine educational costs incurred by Tanzanian farmers. Out of 11,031 farmers in my dataset described later, I construct the dataset of children within these farmers. Figure 5.2 displays the average educational expenditure per student per year in pre-primary, primary, and lower-secondary schools, which are popular ways of schooling in Tanzania, as will be described later. In Figure 5.2, educational costs are separated into multiple components (school fees, books and material, uniform, transport, extra tuition, other contribution, cost of meals, and total expenditure including in-kind values). Educational costs are also separated by two phases before and after 2016 because pre-primary and lower-secondary schools were made compulsory since 2016. I restrict my attention to public schools because most Tanzanian farmers do not attend private schools.[3] Value scales in the vertical axes of Figure 5.2 are different by educational stages for visibility.[4]

Figure 5.2 shows that educational expenses are necessary for the schooling of children even in the compulsory education. Although school fees are close to zero in the compulsory stage, farmers incur other educational costs such as books and materials and uniforms. Farmers also

[3] In my dataset, 16,379 children attend public schools, whereas 1,243 children attend private schools.

[4] In Figure 5.2, I do not adjust for temporal price differences across waves, because they are small in the NPS and the values are almost unchanged (NBS 2022).

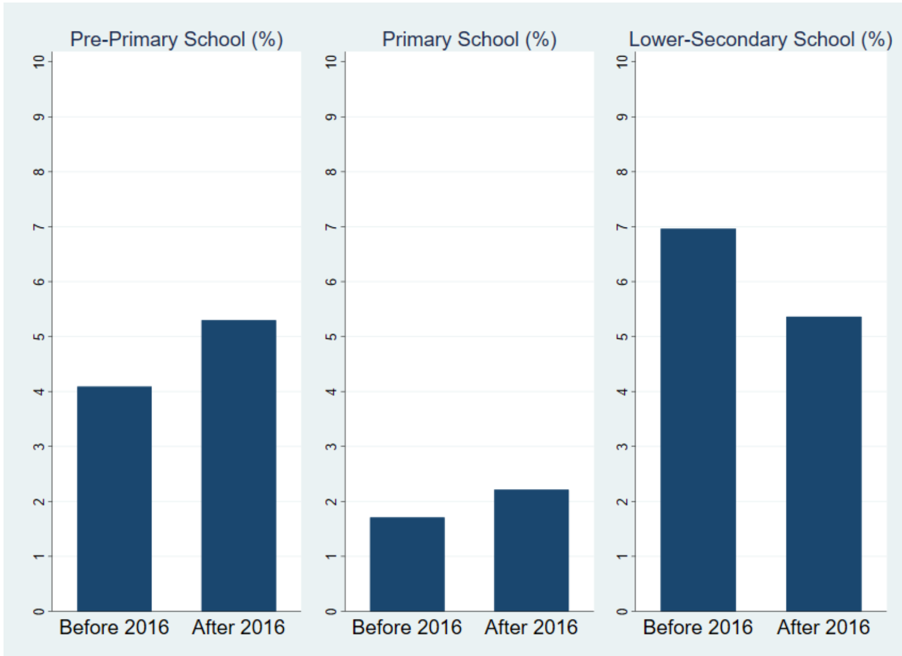


Figure 5.3: Percentage of Educational Expenses to Total Consumption Expenditure within a Household

incurred other contributions to schools before 2016. Consequently, the total expenditures are far from free even in the compulsory education, which may impose an economic burden on Tanzanian farmers.[5] These observations are in line with Dillon (2020), who observed that students' families in Malawi paid for uniforms, books, stationery supplies, and contributions toward the maintenance and construction of school buildings, although primary school fees had been already abolished.

Figure 5.2 also shows that farmers must incur considerably large expenses if their children attend lower-secondary schools. The total expenses in lower-secondary schools, 100,000 Tanzanian shillings (Tsh) ($50) after 2016, are even larger than those in pre-primary schools (below 15,000 Tsh, or $7.5, after 2016) and primary schools (around 25,000 Tsh, or $12.5).

Figure 5.3 shows proportions of educational expenses to the total consumption expenditure within a household for pre-primary, primary, and lower-secondary schools. The proportion in primary schools is the lowest and is followed by pre-primary schools, which is opposite to the order of educational expenditure in Figure 5.2. As I will see, primary schools are the most popular way of schooling in Tanzania, and may tend to impose a less economic burden on households than pre-primary schools. The proportion in lower-secondary schools is the highest but decreases by around two percents after the fee-free policy.

[5]This observation is consistent with descriptions on educational costs in the Tanzania's school system at `https://africaid.org/tanzanias-school-system-an-overview/`.

5.2.3 Distribution of Students in Tanzania

Using the same dataset of children, I then see the distribution of students in each educational level among Tanzanian farmers. Table 5.1 displays the numbers of children attending schools separately by actual ages of children. The prefixes, PP, D, F, and U denote the educational stages of pre-primary schools, primary schools, secondary schools, and universities, respectively. Ages of children are categorized by official school ages of each educational stage. Table 5.1 also displays in parenthesis the numbers of children enrolling in schools over the past year before the interviews. The last two columns in Table 5.1 displays the number of children attending schools and the number of non-schooling children for each age group (schooling includes all educational types such as vocational and teaching schools, although they are not displayed here).

As a caveat, children in a certain grade are not necessarily at the corresponding school age. For example, suppose that a boy aged 6 enrolled in a primary school in January. If he was interviewed in the NPS before his birthday, his age is still 6 (not 7), although his grade is D1. Even if children continue to attend schools consistently with official school ages, the lag of one age may exist in Table 5.1. As another caveat, there are some students whose ages are far apart from official school ages in classes. These students are still included because judging the truth of the data is difficult. Thus, Table 5.1 may roughly sketch the distribution of students.

Table 5.1 shows that primary schools, which are compulsory and fee-free over all waves of the NPS, are the most prevalent way of schooling among Tanzanian farmers. Many children also attend pre-primary and lower-secondary schools, although they are non-compulsory before 2016. In contrast, only a small fraction of children attend upper-secondary schools and universities. Appendix 5-I shows that the numbers of children attending MS+ course, Certificate course, Diploma course, and non-formal education are also limited.

In Table 5.1, children tend to attend a class whose official school age is close to their actual ages. On the other hand, it is widely observed that many children enroll in and attend schools even if their actual ages are above official school ages of the attending classes. Moreover, the last column shows that, in all age groups, there are large numbers of children not attending schools, which even applies to official ages of primary schools. As was explained previously, farmers incur various educational costs even if tuition is free, which may be an obstacle for the schooling of children. Many farmers still fail in making their children enroll in and attend schools.

Table 5.1: Number of Students in Each Educational Level

Age of	Pre-Primary	Primary							Lower-Secondary				Upper-Secondary		University			Schooling	Non-Schooling
Children	PP	D1	D2	D3	D4	D5	D6	D7	F1	F2	F3	F4	F5	F6	U1	U2	U3+	children	children
Official School Age of Pre-Primary School:																			
5 (PP)	425 (303)	71 (71)	2				1											499	1,437
6 (PP)	546 (381)	428 (424)	41	2	1			1										1,019	1,018
Official School Age of Primary School:																			
7 (D1)	293 (176)	703 (655)	330	35				1										1,362	598
8 (D2)	143 (86)	559 (500)	585	315	23	1		1										1,629	380
9 (D3)	41 (22)	270 (227)	492	524	256	31		1										1,615	261
10 (D4)	10 (6)	139 (116)	335	500	514	222	17	1										1,738	228
11 (D5)	11 (7)	58 (46)	146	296	441	449	191	16										1,609	211
12 (D6)	3	17 (12)	69	174	298	484	421	134	11 (11)									1,615	239
13 (D7)	1 (1)	12 (9)	27	66	159	260	432	271	90 (89)	7								1,334	365
Official School Age of Lower-Secondary School:																			
14 (F1)	5 (3)	1 (1)	10	39	95	152	300	303	214 (206)	84	9							1,216	595
15 (F2)	1	2 (1)	5	13	32	75	189	219	174 (172)	191	53	3						964	680
16 (F3)		2	3	12	11	45	93	146	143 (133)	206	154	53						879	810
17 (F4)				7	9	10	49	54	65 (58)	130	145	102	8 (7)	1				601	888
Official School Age of Upper-Secondary School:																			
18 (F5)		2 (1)		3	2	6	23	31	43 (39)	86	119	104	24 (22)	4	2 (1)			478	1,063
19 (F6)					1	5	6	16	25 (25)	50	59	83	19 (15)	7	6 (2)	1		308	946
Official School Age of University:																			
20 (U1)			2	1		5	2	4	13 (13)	19	46	50	14 (12)	16	7 (5)	6	1	239	1,092
21 (U2)				1		1	1	3	4 (3)	11	27	28	7 (7)	11	14 (11)	6	6	155	983
22 (U3)						1	1		3 (1)	8	8	22	8 (6)	8	6 (6)	12	6	116	958
Total	1,479	2,264	2,047	1,988	1,842	1,747	1,726	1,202	785	793	620	445	80	47	35	25	13	17,376	12,752

Note: The table displays the number of students for each age of children attending schools. In parenthesis, I show the number of children enrolling in schools over the past year before the interview. In the last two column, schooling means all educational types including vocational and teaching schools.

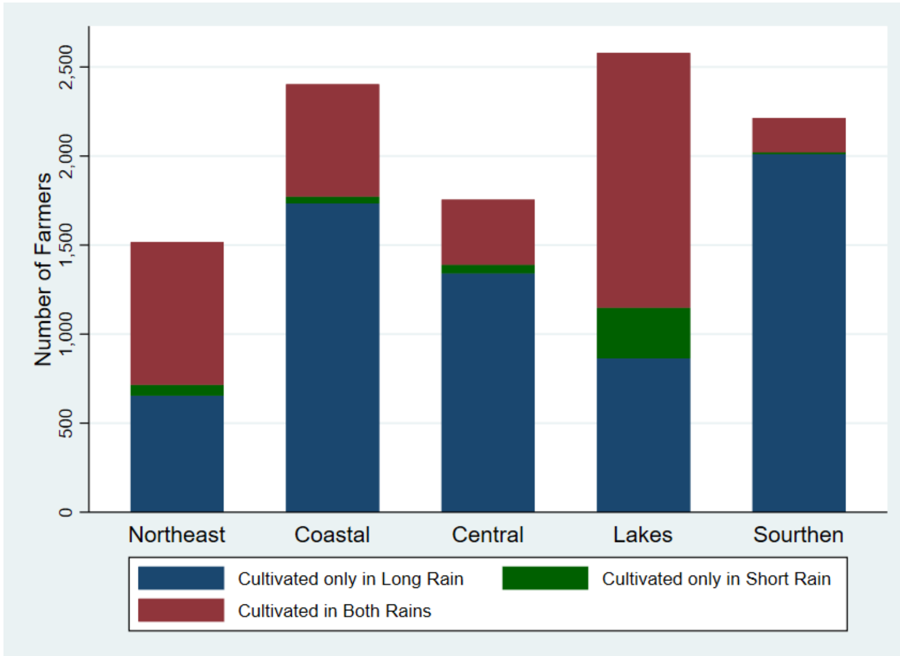


Figure 5.4: Number of Farmers Cultivating in Each Rainy Season

5.2.4 Short Rainy Season in Tanzania

The previous observations suggest that Tanzanian farmers struggle to cover necessary expenses for the schooling of their children. I then examine whether the cultivation in the short rainy season has a potential to promote the schooling of children among farmers.

Using the household dataset described later, Figure 5.4 displays the regional distribution of farmers who cultivate in (i) only the long rainy season, (ii) only the short rainy season, and (iii) both two rainy seasons, as in the previous chapter. Figure 5.4 divides the Tanzanian mainland into five geographical areas: Northeast, Coastal, Central, Lakes, and Southern Highlands.[6] Figure 5.4 indicates that the cultivation in the short rainy season is active in the Northeast area (e.g., Kilimanjaro region) and Lakes area (regions around Lake Victoria), which are typical areas with bimodal rainfall patterns. There are also some farmers cultivating in the short rainy season in other regions such as Coastal and Central areas. The cultivation in the short rainy season is widely observed in Tanzania.

As the previous chapter indicates, the composition of planted crops and maize productivity are considerably similar between the long and short rainy seasons. Moreover, farmers tend to cultivate in the short rainy season, regardless of whether they have non-agricultural occupations. These observations suggest that the cultivation in the short rainy season is not a minor or special economic activity in Tanzania.

I then examine whether the cultivation in the short rainy season is favorable income sources for farmers struggling to cover educational expenses. Figure 5.5 shows the propor-

[6] The five geographical areas include the same Tanzanian regions as those in the previous chapter.

tion of farmers selling crops and the average sales revenues conditional on farmers' selling, separately in each of the two rainy seasons. To control for regional heterogeneity, Figure 5.5 focuses its attention to Northeast and Lakes areas.[7] In Figure 5.5, crops are classified into five types: staples (e.g., maize, rice), sorghum/tubers/roots, legumes/oil, fruits/vegetables, and cash crops.

Figure 5.5 shows that sales activities are more active in the long rainy season than in the short one. In Figure 5.5(a), the proportion of farmers selling crops is nearly 60% in the long rainy season, whereas it is 40% in the short rainy season. Moreover, Figures 5.5(c) and 5.5(d) show that, once farmers sell crops, the average farmers obtain 500,000 Tsh ($250) of sales revenue in the long rainy season, whereas they obtain 300,000 Tsh ($150) in the short rainy season.

However, farmers' sales activities have similarities in the two rainy seasons because main crops for sale are staples and legumes/oil in both rainy seasons. Farmers' sales activities in the short rainy season are not special relative to the long rainy season, and may be useful income sources for wide ranges of farmers. Importantly, the average sales revenue in the short rainy season (300,000 Tsh = $150) appears to be sufficient for covering educational expenses per student per year in pre-primary schools (less than 15,000 Tsh = $7.5 after 2016) and primary schools (around 25,000 Tsh = $12.5), and even in lower-secondary schools (100,000 Tsh = $50 after 2016) (see Figure 5.2). I also calculate the average profits (sales revenue net of input costs), which are not displayed in Figure 5.5.[8] I find that the profits in the short rainy season are 134,711 Tsh ($67) vis-a-vis 314,611 Tsh ($157) in the long rainy season, and can cover expenses in lower-secondary schools.

Observations in the field also support the idea that farmers cover educational expenses by cultivating in the short rainy season in addition to the long one. Ikeno (2011), through his field survey to Kirisi hamlet in Kilimanjaro region, reported that cost-sharing policies for education and health sectors under the Structural Adjustment Policy triggered irrigation farming using ponds in the long dry season in 1990's, so that farmers could cultivate by one more season to cover new expenses for public services. Appendix 5-I shows that the double-cropping is positively associated with more annual revenue of crop sales. Appendix 5-I also shows that farmers are more likely to sell crops in the short rainy season when the number of children enrolling in schools and that of non-schooling children increase within a household, which is unobserved for sales in the long rainy season.

[7] In the NPS, a part of sales of fruits and permanent cash crops are not differentiated by rainy seasons. Thus, Figure 5.5 does not incorporate these data.

[8] I calculate input costs by summing up purchasing and rental costs of inorganic/organic fertilizer, seed, pesticide/herbicide, hired labor, and farm implements.

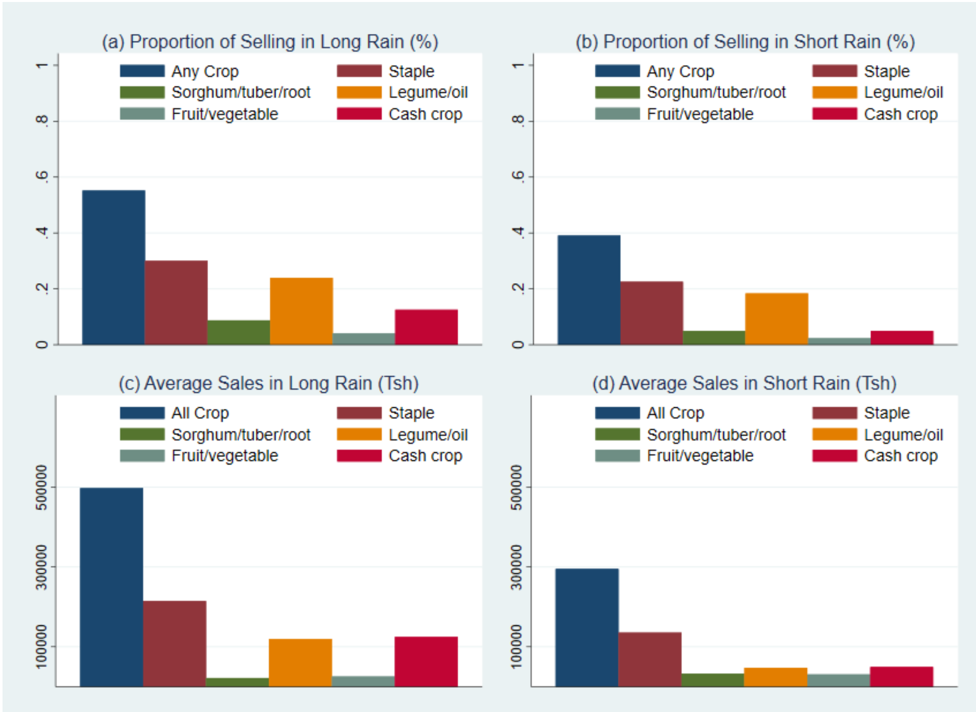


Figure 5.5: Proportions of Farmers Selling Crops and Average Sales Revenues by Five Crop Types in Each Rainy Season in Northeast and Lakes

5.3 Conceptual Framework

The previous descriptive statistics indicate that farmers may be able to cover necessary expenses for the schooling of children by cultivating in the short rainy season. Thus, I firstly present the following hypothesis:

Hypothesis 3. *If farmers cultivate in the short rainy season in addition to the long one, they can make their children enroll in schools and continue to attend them without dropping out.*

However, the cultivation in the short rainy season may have heterogeneous impacts by farmers' economic levels. Previous studies observed that policy changes and economic shocks tend to have smaller impacts on children's education among wealthy farmers than poor farmers (Beegle et al. 2006; Dillon 2021). Wealthy farmers tend to be well-educated and engage in high-return non-farm activities, which mitigate damages by agricultural shocks (Lay et al. 2008; Loison et al. 2015). Wealthy farmers can also rely on financial markets, which buffer against policy changes and economic shocks (Dillon 2021).

To conceptually illustrate heterogeneous impacts on education by economic levels, I rely on the subsistence constraint. Consider two goods, food c and education e, and express a farmer's utility function by $u(c, e)$. One interpretation of this utility function is altruism (Baland and Robinson 2000). Farmers feel satisfaction by making their children receive

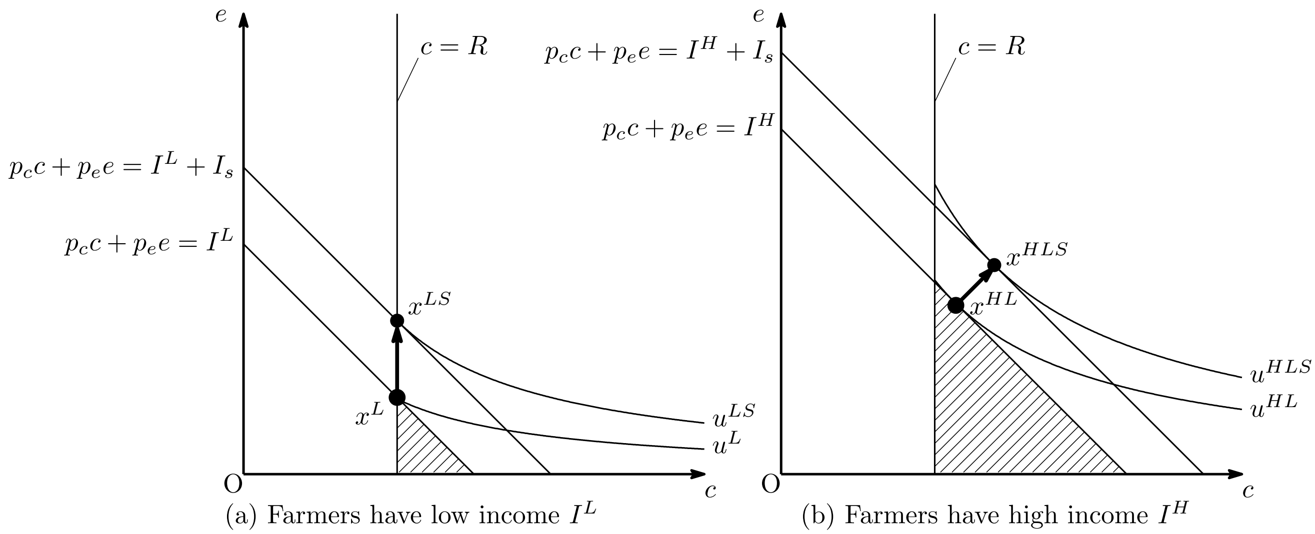


Figure 5.6: Impacts of additional agricultural income in the short rainy season. The shaded parts represent feasible consumption bundles that satisfy both the budget and subsistence constraints under the initial income I^L and I^H

education. I denote market prices for c and e by p_c and p_e, respectively. Then, the budget constraint is $p_c c + p_e e \leq I$, where $I > 0$ is income. I further introduce the subsistence constraint $c \geq R$, where $R \geq 0$ is the subsistence requirement (Basu and Van 1998). Farmers must consume the minimum level of food to meet basic needs in their livelihoods. In the subsistence constraint, food consumption c is a necessity, whereas education e is a non-necessity.

I consider poor farmers who have low income I^L and wealthy farmers who have high income I^H (both on-farm and off-farm incomes are included together in I^L and I^H for a simple notation). Figure 5.6(a) displays how the consumption by poor farmers moves if they obtain additional agricultural income I_s in the short rainy season. Because the initial income I^L is low, poor farmers must prioritize consuming c relative e to satisfy the subsistence constraint $c \geq R$. Then, the solution in I^L is given at the intersection between the budget and subsistence lines, x^L. If poor farmers obtain additional agricultural income I_s in the short rainy season, the solution moves along the subsistence line and reaches x^{LS}. Poor farmers, whose free consumption is restricted by the subsistence constraint, spend all additional income I_s on increasing the educational consumption e.

In contrast, Figure 5.6(b) displays how the consumption choice by wealthy farmers moves by I_s. Because the initial income I^H is high, wealthy farmers do not have to care about the violation of the subsistence constraint. Thus, the solution under I^H, or x^{HL}, coincides with the solution in the customary consumer problem without the subsistence constraint, where the gradient of the indifference curve equals that of the budget line. In x^{HL}, wealthy farmers already realize the balanced consumption between c and e in line with their preference. If wealthy farmers obtain additional agricultural income I_s in the short rainy season, the solution moves from x^{HL} to x^{HLS}. Wealthy farmers spend the additional income I_s on both c and e.

Poor farmers prioritize consuming necessities relative to children's education, which means that there is more room for poor farmers to increase educational consumption than wealthy farmers. Thus, the additional agricultural income in the short rainy season contributes to more education of children among poor farmers than among wealthy farmers. Therefore, this chapter secondly presents the following hypothesis:

Hypothesis 4. *The positive impacts of the double-cropping are stronger for poor farmers than for wealthy farmers.*

5.4 Dataset and Estimation Strategy

5.4.1 Dataset

My dataset comes from six waves from 2008/09 to 2020/21 in the the Tanzanian NPS. The NPS comprises nationally representative households and provides socioeconomic and agricultural information at household and plot levels.

Table 5.2 displays the number of households in each wave in my dataset. My dataset is unbalanced and comprises farmers who have agricultural plots, reside in the Tanzanian mainland, and have at least two waves of data for panel construction. In 2008/09-2012/13, the composition of households is basically the same in the NPS. However, the NPS refreshed composition of households in 2014/15. Out of households who were interviewed before 2012/13, 547 households were interviewed again in 2014/15. They were also interviewed in 2018/19. In contrast, the NPS added new 1,774 households in 2014/15. They were interviewed again in 2020/21.

I construct my dataset regardless of whether they have children or not, because drawing a line between "children" and "adults" is difficult. As Table 5.1 shows, Tanzanian people often attend schools even if their actual ages are far above official school ages of attending classes. In my dataset, 8,466 out of 11,031 households have children at official school ages from pre-primary to lower-secondary schools (5-17). To check the robustness, I will restrict my dataset to these households after the main analysis.

Table 5.3 roughly sketches the timeline of agricultural activities and children's schooling in the NPS, taking a case of Wave 2012/13. In each wave, all farmers were asked of agricultural activities in the long rainy season of the same period (March-May in 2012). In contrast, farmers were asked of agricultural activities in the most *recent* short rainy season from the date of interview. Thus, farmers provided information on the short rainy season *before* the long rainy season (October-December in 2011) if they were interviewed early and that *after*

Table 5.2: Number of Households in Each Wave

	2008/09	2010/11	2012/13	2014/15	2018/19	2020/21	Total Sample Size
Number of Households	1,853	2,247	2,296	2,321	540	1,774	11,031

Table 5.3: Timeline of NPS in Wave 2012/13

	2011			2012												2013								
	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Short rainy season (*before*)	Cultivate			Harvest																				
Long rainy season						Cultivate			Harvest															
Short rainy season (*after*)													Cultivate			Harvest								
Start of schooling				×												×								
Month of interview in NPS													Interview											

Note: "×" indicates the starting month of schooling in pre-primary, primary, and lower-secondary schools.

the long rainy season (October-December in 2012) if they were interviewed late. All waves in the NPS have the same chronological structure. This chronological difference will be utilized in constructing instrumental variables and rainfall variables later.

As Table 5.3 shows, pre-primary, primary, and lower-secondary schools, which are the focus of my analysis, start from January. Thus, the start of schooling quickly follows the cultivation in the short rainy season in October-December. Farmers may decide to cultivate in the short rainy season to prepare for expenses of children's enrollment, which can be a source of endogeneity, as I will describe.

Because my dataset is an unbalanced panel, I test for the sample selection bias by adding a lagged selection indicator to covariates in my main analysis and examining its statistical significance (see pages 832-833 in Wooldridge (2010) for details). Because the lagged selection indicator is insignificant in most dependent variables, I proceed without correcting for the sample selection bias.[9]

5.4.2 Estimation Strategy

My purpose is to examine whether the double-cropping over the two rainy seasons promotes farmers to make their children enroll in schools and continue to attend them.

Enrollment in School

As a dependent variable, I create a dummy variable that is equal to one if any child within a household enrolls in a school. For each child in a household, the NPS provides the information

[9]Exceptionally, I found statistical significance when a dependent variable is the number of children who continue to attend primary schools. In this case, I attempted to correct for the bias by adding to covariates the inverse Mills ratio (page 835 in Wooldridge (2010)) and still found similar results.

on children's grades in schools both at the interview time and one year before. By comparing contemporary and preceding grades of each child, we can discern whether any child in a household enrolled in schools over the past year. Because pre-primary, primary, and lower-secondary schools are popular ways of schooling among Tanzanian farmers, I examine the probability of children's enrollment in each of these schools, using the probit model.

Moreover, children must pass the PSLE to go on to public lower-secondary schools. Out of 18,664 children who took the PSLE, only 38% passed it in my dataset. Thus, even if farmers cultivate in both rainy seasons to cover educational expenses for lower-secondary schools, their children may not pass the PSLE. In order to accurately capture the impacts of the double-cropping, I also examine the probability that children within a household take the PSLE, using the probit model. Unfortunately, the NPS does not provide the information on the PSLE in 2008/09. Thus, the sample size reduces from 11,031 to 9,178 in the analysis of the PSLE.

Continuation of School without Dropping Out

Out of children who attended schools one year before the interview time, I also construct the number of children who continue to attend the same schools at the interview time without dropping out. I do not count children who are at the highest grades of primary and lower-secondary schools (D7 and F4 in Table 5.1), so that I can exclude from the variable's definition children's enrollment in the next level of school. For example, the transition from D7 to F1 in Table 5.1 is not counted as the continuation of primary schools. Unfortunately, the NPS classifies all pre-primary students as one group, regardless of different grades therein. Thus, we do not exclude from the variable's definition children at the top grade of pre-primary schools. I construct three numbers of children who continue to attend pre-primary, primary, and lower-secondary schools, respectively.

In my dataset, 6,438 out of 11,031 households have children who continue to attend pre-primary, primary, or lower-secondary schools from one year before the interview time, which means that a moderate portion of households take zero values in the school continuation (corner solutions). Therefore, I rely on the Tobit model to estimate the increase in the number of children continuing to attend schools in each of pre-primary, primary, and lower-secondary schools.

Statistical Model

To test Hypothesis 3, I estimate the following statistical model for farmer i and year t, separately by enrollment in and continuation of school j:

$$y^*_{ijt} = \alpha_{1j} D_{it} + \alpha_{2j} \ln(Asset_{it} + 1) + \boldsymbol{\gamma_j'} \boldsymbol{x_{it}} + \mu_{ij} + \epsilon_{ijt}, \tag{5.4.1}$$

where y^*_{ijt} is the latent variable for the dependent variable y_{ijt}. When I examine children's enrollment in schools, I consider four dummy variables; we have $y_{ijt} = 1$ if any child within a household enrolls in a (i) pre-primary school, (ii) primary school, (iii) lower-secondary school, or (iv) takes the PSLE. In contrast, I also consider as y_{ijt} three numbers of children who continue to attend (i) pre-primary schools, (ii) primary schools, or (iii) lower-secondary schools.

In Equation (5.4.1), D_{it} is a dummy variable that is equal to one if farmer i cultivates in the two rainy seasons within a year (double-cropping).[10] I also introduce to Equation (5.4.1) $\ln(Asset_{it} + 1)$, the natural logarithm for the sum of all asset values per adult equivalent (e.g., televisions, motor vehicles, and houses) plus one. To obtain $Asset_{it}$, I calculate predicted prices of each asset, multiply them by the number of asset holdings within a household, and divide the sum of predicted asset values by adult equivalents. Because asset values take large monetary values, I take their natural logarithm to prevent outlier effects. I add one to the original asset values because they are zero for a few farmers.

If Hypothesis 3 is true, the double-cropping over the two rainy seasons promotes children's school enrollment and continuation. Thus, I expect $\alpha_{1j} > 0$ in Equation (5.4.1).

To test heterogeneous effects by farmers' wealth levels in Hypothesis 4, I then estimate the following equation:

$$y^*_{ijt} = \beta_{1j} D_{it} + \beta_{2j} D_{it} \times \ln(Asset_{it} + 1) + \beta_{3j} \ln(Asset_{it} + 1) + \boldsymbol{\delta_j'} \boldsymbol{x_{it}} + \mu_{ij} + \epsilon_{ijt}, \quad (5.4.2)$$

In Equation (5.4.2), I introduce the interaction term between D_{it} and $\ln(Asset_{it} + 1)$, as previous studies did (Beegle et al. 2006; Kruger 2007; Bandara et al. 2015; Dillon 2021). The effect of the double-cropping for the poorest farmers is β_{1j}, which is obtained when $Asset_{it} = 0$, and changes to $\beta_{1j} + \beta_{2j} \ln(Asset_{it} + 1)$ when $Asset_{it} > 0$. If Hypothesis 4 is true, I expect $\beta_{1j} > 0$ and $\beta_{2j} < 0$.

In Equations (5.4.1) and (5.4.2), $\boldsymbol{x_{it}}$ is a set of household characteristics. Individual-specific effects and idiosyncratic errors are denoted by μ_{ij} and ϵ_{ijt}, respectively.

The heterogeneous impacts of the double-cropping through asset values may take a form of the inverted U-shaped curve because poorest households with no asset may still struggle to incur educational expenses even if they obtain additional income by the double-cropping. In Appendix 5-II, I add to the estimated equation the interaction term between D_{it} and the squared term of asset values and estimate it. However, I find no statistical significance for its interaction term.

[10] In the robustness check later, I will change D_{it} by acres of planted land during the short rainy season.

Controlling for Time-Constant Heterogeneity

As Figure 5.4 suggests, farmers' decisions to cultivate in the short rainy season depend on regional environmental conditions, which are time-invariant but may be unobserved from the dataset. To control for household time-constant heterogeneity, I combine non-linear models (probit and Tobit models) with the correlated random effects (CRE) method.[11]

The CRE method introduced by Mundlak (1978) and Chamberlain (1984) adds the time-averaged values of all time-varying covariates to the set of covariates. This allows us to assume that time-varying covariates are uncorrelated with individual-specific effects μ_{ij} (Wooldridge 2010).[12]

Controlling for Time-Variant Heterogeneity

The variables of my interest, D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$, may still be endogenous because of the reverse causality. Farmers must make an early decision on whether they make their children enroll in and attend schools or not, because procedures for enrollment and starting of a new grade, preparation for entrance examination, and sizable cash outflows are necessary early in advance. Because of predictable properties of education, my estimation may reflect the reverse causality that expected events of children's enrollment and attendance at schools in the future ($y_{ijt} > 0$) promote farmers to cultivate in the short rainy season ($D_{it} = 1$) (I suggested this possibility while citing Ikeno (2011) in the Background Information). As Table 5.3 shows, the start of schooling quickly follows the short rainy season, which may also strengthen the reverse causality. Thus, the coefficient for D_{it} would be overestimated if this reverse causality existed.

To control for time-variant heterogeneity ϵ_{ijt} in Equations (5.4.1) and (5.4.2), I further combine the CRE method with the control function (CF) approach. The CF approach controls for endogenous variables through a two-step procedure, which is similar to two-stage least squares (2SLS) regression (Wooldridge 2015). In the first step, I first regress an endogenous variable on instrumental variables (IV) and covariates, and then obtain the generalized residuals (Wooldridge 2014). In the second step, I regress a dependent variable on the endogenous variables, residuals, and covariates. The CF approach provides a simple Hausman test. If the residuals significantly differ from zero in the second step estimation, an explanatory variable that is considered endogenous is actually endogenous.

To control for endogeneity of D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ in each of Equations (5.4.1) and (5.4.2), I introduce four IVs that are closely related to the short rainy season. As Figure 5.4

[11] The fixed effects approach should not be used in my nonlinear models, because estimated coefficients are inconsistent unless the number of time periods is large (incidental parameters problem) (Wooldridge, 2010).

[12] I adopt the pooled CRE probit/Tobit models because their marginal effects are obtained straightforward in Stata. The pooled probit model also allows for serial dependence (Wooldridge 2010).

shows, most farmers cultivate in the long rainy season. Thus, the double-cropping over the two rainy seasons is almost determined by whether farmers cultivate in the short rainy season.

The first three IVs are constructed from the NINO3.4 index and Dipole Mode Index (DMI), which are meteorological indices for El Niño and the Indian Ocean Dipole (IOD) events, respectively. It is known that El Niño and IOD events, anomalies of the sea surface temperatures (SST) in the tropical eastern Pacific Ocean and in the tropical Indian Ocean, respectively, are closely linked to the rainfall amount in the short rainy season and hence to farmer's cultivation decisions in this season (Mutai and Ward 2000; Lu et al. 2018). Because the rainfall variability in the short rainy season has strong relations to El Niño and IOD events from September to February, as I explained in Chapter 4. Thus, I sum up NINO3.4 and DMI indices over this period, create their interaction term, and introduce these three variables as IVs.

Because these meteorological indices vary by year but take the same values among all Tanzanian farmers, I further add to IVs the proportion of farmers cultivating in both two rainy seasons for each district (*districts* are the second largest administrative unit following *regions*). The regional prevalence of the double-cropping seems to correlated with its practices by farmers, but does not seem to directly determine cultivation decisions by individual farmers. Thus, this IV is unlikely to be correlated with the idiosyncratic error ϵ_{ijt}. Although this IV may be correlated with the unobserved regional potential for agricultural production, I control for time-constant heterogeneity in the CRE model.[13]

Appendix 5-II shows the estimation results in the first step of the CF approach. The four IVs tend to be individually significant and are also jointly significant.

Covariates

Household characteristics $\boldsymbol{x_{it}}$ include age, gender, and years of education of a household head. Because the composition of a household is related to children's enrollment and attendance at schools, I include adult equivalents as a household-size indicator and the number of children with official ages from pre-primary schools to lower-secondary schools (5-17). I also include the number of adults aged from 18 to 60 (except for students and disabled) because the working force within a household may affect the necessity for child labor, which may discourage the schooling of children. I add to $\boldsymbol{x_{ijt}}$ the total acres of land owned by a

[13] Although household asset values are related to broad ranges of farmers' economic characteristics, I do not address the possible endogeneity for $\ln(Asset_{it}+1)$. This is because it is difficult to introduce an appropriate IV for household asset values. However, I conducted a robustness check by introducing the average value of $Asset_{it}$ for each district as an IV for $\ln(Asset_{it}+1)$ (I calculate the average using all households in the NPS, which include non-farmers as well as farmers in Table 5.2). The regional abundance of assets seems to be correlated with the household's asset levels, but does not seem to directly determine the asset formation of individual households. Thus, this IV is unlikely to be correlated with the idiosyncratic error ϵ_{ijt}. The results are still similar to my main results.

household. Because households may be able to cover educational expenses by receiving loans, I add a dummy variable that is equal to one if a household belongs to a SACCO (savings and credit cooperatives). Because urban households may have a better access to schools, I also add a dummy variable that is equal to one if a household resides in urban areas. Schooling of children may be easier if a farmer has other income sources than agriculture. Thus, I include a dummy variable that is equal to one if any member within a household has non-agricultural occupations.

To control for regional climatic tendencies, I add to $\boldsymbol{x_{ijt}}$ the rainfall amounts in the long rainy season from March to May and those in the short rainy season from September to January in each district. I also add the sum of rainfall during the long rainy season and that during the short rainy season over the past five years in each district. These rainfall variables, which have large values, are transformed into logarithmic form in the analysis.

To control for geographical characteristics, four area dummies, which are shown in Figure 5.4, are introduced, with Southern Highlands the base area. Finally, five wave dummies are included, with 2008/09 the base period.[14]

5.4.3 Descriptive Statistics

Table 5.4 shows mean values of variables, separately by whether farmers cultivate in both two rainy seasons within a year or not. Mean differences are tested between the two farmer groups. Farmers who cultivate in both rainy seasons are more likely to make their children enroll in all levels of schools and take the PSLE. Moreover, farmers make more children continue to attend primary and lower-secondary schools without dropping out if they cultivate in both rainy seasons.

I find significant differences for all household and climatic characteristics except for household asset values and number of working adults. If households cultivate in both rainy seasons, their heads tend to be older and well-educated males. Farmers who cultivate in both rainy seasons tend to have more household members (adult equivalents) and more children, but own less acres of land. They are also more likely to belong to SACCO, reside in urban areas, and have other occupations than agriculture. Farmers who cultivate in both seasons tend to have more rainfall during both long and short rainy seasons, which suggests that farmers can cultivate in the short rainy season on climatically favorable conditions.

[14]In the analysis of the PSLE, I set the base period as 2010/11 because the data in 2008/09 are not used.

Table 5.4: Summary Statistics of Household and Climatic Characteristics

Mean Values of Variables	Cultivated in Both Rains	Uncultivated in Both Rains	P-Value
Educational Variables:			
Whether any children within a household:			
· enroll in pre-primary schools (dummy)	0.10	0.07	0.00^{***}
· enroll in primary schools (dummy)	0.18	0.16	0.00^{***}
· enroll in lower-secondary schools (dummy)	0.07	0.06	0.06^{**}
· take PSLE (dummy)	0.12	0.09	0.00^{***}
Number of children who :			
· continue to attend pre-primary schools	0.04	0.03	0.68
· continue to attend primary schools	1.01	0.89	0.00^{***}
· continue to attend lower-secondary schools	0.18	0.15	0.01^{***}
Household characteristics:			
Asset values (Tsh)	3,273,761	3,479,559	0.23
Age of household head	49.94	48.79	0.00^{***}
Male-headed household (dummy)	0.76	0.74	0.01^{**}
Years of education of household head	4.82	4.59	0.00^{***}
Adult equivalents	4.72	4.43	0.00^{***}
Number of children aged 7-17	2.21	1.99	0.00^{***}
Number of working adults	2.15	2.13	0.68
Acres of owned land	5.84	6.91	0.00^{***}
Belongs to SACCO (dummy)	0.06	0.03	0.00^{***}
Resides in urban areas (dummy)	0.11	0.19	0.00^{***}
Has non-agricultural occupations (dummy)	0.66	0.63	0.01^{***}
Climatic characteristics:			
Rainfall during long rainy season (mm)	445.56	389.43	0.00^{***}
Rainfall during short rainy season (mm)	483.30	450.66	0.00^{***}
Sum of long rain over past 5 years (mm)	2210.56	1948.21	0.00^{***}
Sum of short rain over past 5 years (mm)	2298.71	2171.83	0.00^{***}
Observations	3,419	7,612	

Note: Mean Differences are tested between two groups at significance levels of * p<0.1, ** p<0.05, *** p<0.01.

5.5 Results

5.5.1 Children's Enrollment in Schools

Using the pooled CRE probit model, Table 5.5 displays the estimation results of children's enrollment in schools; Table 5.5(a) estimates Equation (5.4.1) and does not include the interaction term between D_{it} and $\ln(Asset_{it}+1)$ in covariates, whereas Table 5.5(b) includes it to estimate Equation (5.4.2). In Columns (1)-(4), the pooled CRE probit model is not combined with the CF approach, whereas the two methods are combined in Columns (5)-(8). Comparing the two types of results is helpful to consider how the endogeneity of D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ could affect my results, as well as to check their robustness. In Table 5.5, household and regional characteristics, area and wave dummies, the constant, and the time-averaged values are included as covariates in the estimation but excluded from the table for brevity. Appendix 5-II shows full results of Table 5.5(a). Because the coefficients in the probit model do not represent marginal effects, Table 5.5 indicates average partial effects (APE). In Columns (1)-(4), standard errors (SE) are clustered at the household level. In Columns (5)-(8), I obtain SE by bootstrapping with repetitions 200 to adjust SE for the two-step estimation (Wooldridge 2015).

In Columns (1)-(4) of Table 5.5(a), I find no evidence that the double-cropping is associated with the probability of children's school enrollment. If I control for time-variant heterogeneity by the CF approach, the generalized residuals obtained in the first-step estimation for D_{it} is significant in Column (6), which implies that D_{it} is endogenous by the Hausman test. Then, Column (6) shows that the double-cropping significantly promotes children's enrollment on primary schools by 7 percentage points (pp), which supports Hypothesis 3.

I then examine heterogeneous effects by households' asset levels in Table 5.5(b). In Columns (1) and (2), the double-cropping is positively associated with the probabilities that children within the poorest households with no asset enroll in pre-primary schools and primary schools by 17 pp and 32 pp, respectively. However, the probabilities of children's enrollment in pre-primary and primary schools tend to decrease by 1 pp and 2 pp, respectively, as household asset levels increase by 1%, which supports Hypothesis 4. In contrast, I find no statistical significance for the probabilities that children enroll in lower-secondary schools and take the PSLE in Columns (3) and (4).

If the CRE probit model is combined with the CF approach, generalized residuals obtained in the first-step estimation for D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are not individually significant in Columns (5)-(8). Thus, I find no evidence that D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are endogenous. However, the sizes of estimated APEs for D_{it} in Columns (5) and (6) become small (0.08 and 0.20) relative to those in Columns (1) and (2) (0.16 and 0.30). Without controlling for time-variant heterogeneity, the estimated coefficients for D_{it} could reflect the reverse causality.

Table 5.5: CRE Probit Estimates of APEs for Children's Enrollment in Schools

	CRE model uncombined with CF				CRE model combined with CF			
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary	(8) Take PSLE
(a) Exclude the interaction term $D_{it} \times \ln(Asset_{it} + 1)$								
Double-cropping and assets:								
D_{it} (double-cropping) (dummy)	0.00	0.02	-0.00	0.01	-0.02	0.07**	-0.01	0.01
	(0.58)	(0.17)	(0.86)	(0.50)	(0.43)	(0.01)	(0.61)	(0.77)
$\ln(Asset_{it} + 1)$ (log of household asset values in Tsh)	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00
	(0.34)	(0.87)	(0.52)	(0.88)	(0.46)	(0.96)	(0.53)	(0.89)
Generalized residuals:								
Generalized residuals for D_{it}					0.02	-0.04**	0.01	0.01
					(0.17)	(0.02)	(0.45)	(0.71)
(b) Include the interaction term $D_{it} \times \ln(Asset_{it} + 1)$								
Double-cropping and assets:								
D_{it} (double-cropping) (dummy)	0.16*	0.30**	-0.02	-0.01	0.08	0.20*	0.10*	0.12
	(0.06)	(0.02)	(0.81)	(0.95)	(0.27)	(0.06)	(0.09)	(0.26)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.01*	-0.02**	0.00	0.00	-0.01	-0.01	-0.01*	-0.01
	(0.07)	(0.02)	(0.82)	(0.89)	(0.16)	(0.28)	(0.08)	(0.26)
$\ln(Asset_{it} + 1)$ (log of household asset values in Tsh)	0.00	0.01	0.00	-0.00	-0.00	0.00	0.00	0.00
	(0.92)	(0.20)	(0.79)	(0.82)	(0.99)	(0.62)	(0.24)	(0.68)
Generalized residuals:								
Generalized residuals for D_{it}					0.00	-0.03	0.00	-0.01
					(0.99)	(0.28)	(0.96)	(0.66)
Generalized residuals for $D_{it} \times \ln(Asset_{it} + 1)$					0.06	-0.02	0.03	0.07
					(0.37)	(0.85)	(0.68)	(0.44)
F statistic (p-values in parenthesis)								
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it}) + 1$)	3.61	6.99**	0.09	0.47	2.50	7.84**	3.02	1.27
	(0.16)	(0.03)	(0.96)	(0.79)	(0.29)	(0.02)	(0.22)	(0.53)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use SE clustered at the household level in Columns (1)-(4) and SE by bootstrapping with 200 repetitions in Columns (5)-(8). For other covariates, see full results in Table A22 of Appendix 5-II.

That is, expectations for children's school enrollment promote farmers to cultivate in the short rainy season, regardless of whether the double-cropping really enables their children to enroll in schools, which leads to the overestimation in Columns (1) and (2).

However, I still find that D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ are jointly significant in Column (6), as the F test for $\beta_{1j} = \beta_{2j} = 0$ in Equation (5.4.2) shows. Thus, the benefit of the double-cropping is the strongest for the poorest households with no asset and shrinks as households' wealth levels increase.

In Columns (7) and (8) of Table 5.5(b), I find that estimated APEs for D_{it} are positive and those for $D_{it} \times \ln(Asset_{it} + 1)$ are negative, which are different from estimated signs in Columns (3) and (4). In particular, Column (7) shows that the probability of children's enrollment in lower-secondary schools significantly increases by 10 pp if the poorest households cultivate in both rainy seasons, and decreases by 1 pp per 1% increase of household asset levels. Although educational expenses for lower-secondary schools are even higher than those for pre-primary and primary schools (see Figure 5.2), the additional cultivation in the short rainy season enables farmers to make their children enroll in lower-secondary schools.

The reverse causality, which was observed for pre-primary and primary schools, does not seem to work for lower-secondary schools. This may be attributed to the discontinuity between primary and lower-secondary schools. Farmers may be weakly motivated to make their children go on to lower-secondary schools, compared with primary schools. Then, farmers may not be inclined to utilize the short rainy season to prepare for children's enrollment in lower-secondary schools.

Children's Continuation of Schools without Dropping Out

Using the pooled CRE Tobit model, Table 5.6 displays the estimation results on the number of children who continue to attend schools since one year before the interview time. Table 5.6(a) does not include the interaction term $D_{it} \times \ln(Asset_{it} + 1)$ in covariates, whereas Table 5.6(b) includes it. The pooled CRE Tobit model and the CF approach are not combined in Columns (1)-(3), but they are combined in Columns (4)-(6). Household and regional characteristics, area and wave dummies, and time-averaged-values are included in the analysis but omitted for brevity For full results of Table 5.6(a), see Appendix 5-II.

Contrary to my expectation, Table 5.6 shows that the double-cropping has no significant impacts on the children's school continuation in all columns, regardless of whether the interaction term $D_{it} \times \ln(Asset_{it} + 1)$ is included or not. In Column (5), the two generalized residuals are significant, which implies that D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ are endogenous by the Hausman test. Then, the coefficient for D_{it} becomes positive in Column (5), compared with the negative coefficient in Column (2). However, it is still insignificant.

I provide three possible explanations for different results between children's school en-

Table 5.6: CRE Tobit Estimates of APEs for the Number of Children Continuing to Attend Schools

	CRE model uncombined with CF			CRE model combined with CF		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Pre-Primary	(5) Primary	(6) Lower-Secondary
(a) Exclude the interaction term $D_{it} \times \ln(Asset_{it}+1)$						
Double-cropping and assets:						
D_{it} (double-cropping) (dummy)	-0.01	0.01	-0.01	0.00	-0.00	-0.01
	(0.26)	(0.81)	(0.52)	(0.91)	(0.98)	(0.84)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.63)	(0.71)	(0.59)	(0.62)	(0.72)	(0.60)
Generalized residuals:						
Generalized residuals for D_{it}				-0.00	0.02	0.01
				(0.70)	(0.49)	(0.64)
(b) Include the interaction term $D_{it} \times \ln(Asset_{it}+1)$						
Double-cropping and assets:						
D_{it} (double-cropping) (dummy)	0.01	-0.22	-0.06	-0.04	0.10	-0.01
	(0.88)	(0.35)	(0.56)	(0.39)	(0.65)	(0.95)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.00	0.02	0.00	0.00	-0.01	0.00
	(0.75)	(0.33)	(0.62)	(0.47)	(0.35)	(0.82)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	-0.01	-0.00	-0.00	0.00	-0.00
	(0.80)	(0.41)	(0.46)	(0.50)	(0.76)	(0.47)
Generalized residuals:						
Generalized residuals for D_{it}				-0.01	-0.12**	0.05
				(0.46)	(0.03)	(0.19)
Generalized residuals for $D_{it} \times \ln(Asset_{it}+1)$				0.04	0.59***	-0.16
				(0.49)	(0.00)	(0.17)
F statistic (p-values in parenthesis)						
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it}+1)$)	0.77	0.55	0.31	0.84	3.23	0.39
	(0.46)	(0.58)	(0.73)	(0.66)	(0.20)	(0.82)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use SE clustered at the household level in Columns (1)-(3) and SE by bootstrapping with 200 repetitions in Columns (4)-(6).
For other covariates, see full results in Table A23 of Appendix 5-II.

rollment and continuation. First, the start of schooling, or children's enrollment, quickly follows the short rainy season, as Table 5.3 shows. Thus, the additional agricultural income in the short rainy season may work effectively for children's school enrollment, rather than their school continuation thereafter. Second, the cultivation in the short rainy season may work as a one-time positive income shock. Appendix 5-II examines the frequency of farmers' double-cropping and indicates that farmers do not always cultivate in both rainy seasons. Through the determinant analysis at the plot level, the previous chapter suggested that the maintenance of soil nutrients is an important factor for the double-cropping. If maintaining soil nutrients is difficult, farmers may temporarily cultivate in the short rainy season to cover expenses for new events such as school enrollment, rather than continual expenses for daily activities such as school continuation. Third, farmers who obtain additional agricultural income in the short rainy season may increase the schooling of children disproportionately between girls and boys, which may obscure the impacts of the double-cropping. Previous studies found different income effects on the schooling of children by gender (Gubert and Robilliard 2008; Björkman-Nyqvist 2013; Mani et al. 2013). I will discuss gender-differential effects later.

5.5.2 Different Impacts by Wealth Levels

The previous results on children's enrollment indicate that the impacts of the double-cropping on children's enrollment is the strongest for the poorest households with no asset and shrink as household asset levels increase. However, it is still unclear whether the benefits of the double-cropping last over a wide range of asset levels, or disappear at the moderate asset levels.

Figure 5.7 shows how the predicted probabilities of children's enrollment in primary and lower-secondary schools change by asset levels, provided that farmers cultivate in both rainy seasons ($D_{it} = 1$). The log of household asset values per adult equivalent is placed in the horizontal axis. The 95% confidence intervals (CI) are depicted together at 10th, 25th, 50th, 75th, and 90th quantiles of the asset values (Appendix 5-II depicts the histogram of original asset values).

Figures 5.7(a) and 5.7(b) show that the predicted probabilities of children's enrollment in primary and lower-secondary schools decrease by around 2 pp from 10th to 90th quantiles, which is even lower than the increased probabilities of enrollment caused by the double-cropping itself (recall that the APEs for D_{it} are 0.20 in Column (6) and 0.10 in Column (7) in Table 5.5).

Therefore, the effects of the double-cropping on children's enrollment remain over wide ranges of asset levels. This suggests that the additional cultivation in the short rainy season can be a useful strategy for both poor and wealthy farmers to cover educational expenses.

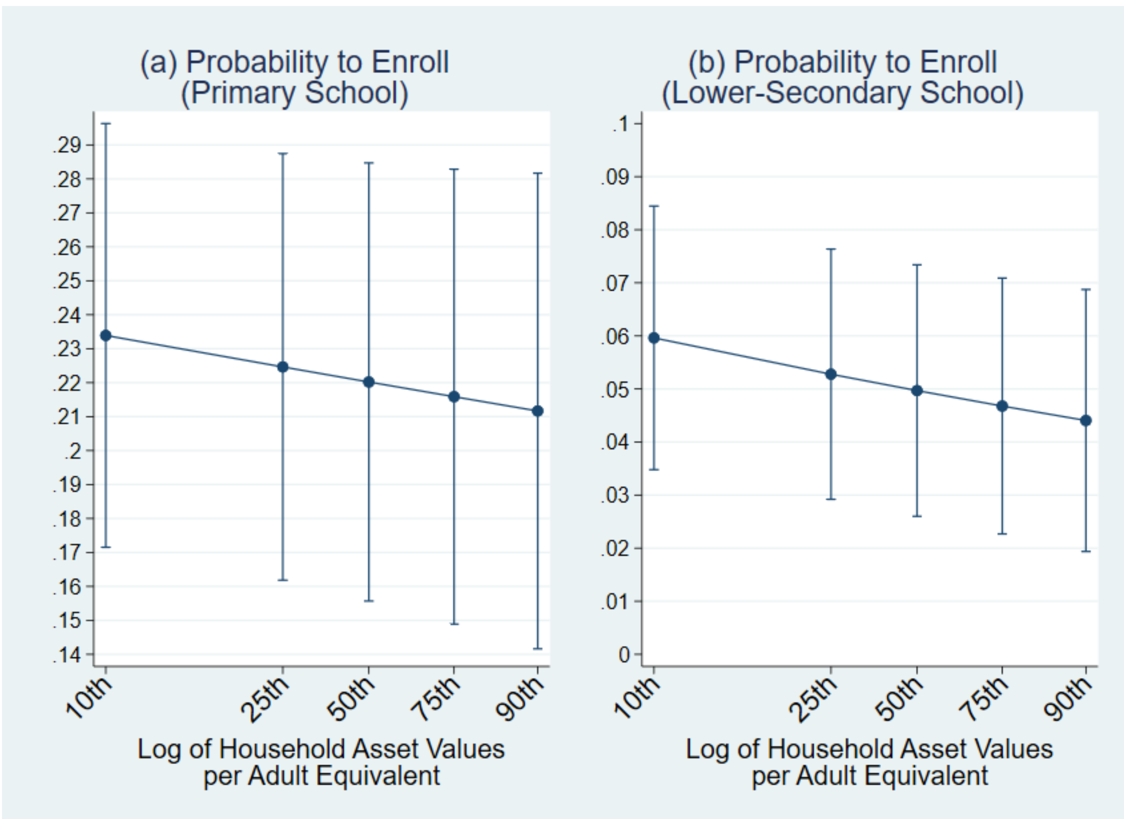


Figure 5.7: Predicted Probability of Children's Enrollment with 95% Confidence Intervals at 10th, 25th, 50th, 75th, 90th Quantiles of Asset Values

5.5.3 Robustness Check

For the robustness check, I restrict my dataset to households with children at the official ages from pre-primary to lower-secondary schools (5-17). Using the CF approach, Appendix 5-II shows results including the interaction term $D_{it} \times \ln(Asset_{it} + 1)$. I still obtain the same findings.

For another robustness check in Appendix 5-II, I change the dummy variable for the double-cropping (D_{it}) by acres of land planted during the short rainy season. I find that acres of planted land are positively associated with the probability for children to enroll in lower-secondary schools and take the PSLE. Moreover, these probability increases are weaken as households' asset levels increase.

5.6 Discussions

5.6.1 Impacts on Household Educational Expenditure

I examine whether and how the double-cropping enables farmers to incur more educational expenses. Combining the pooled CRE Tobit model with the CF approach, Table 5.7 displays estimation results of household's educational expenses (Tsh) for school fees, books and material, uniform, transport, extra tuition, other contribution, and cost of meals, separately by (a) pre-primary schools, (b) primary schools, and (c) lower-secondary schools. In the last column, I estimate total educational expenses in each school. In Table 5.7, I include the interaction term $D_{it} \times \ln(Asset_{it} + 1)$ in covariates, because the heterogeneity by economic levels plays an important role in my previous results. I transform these educational expenses in logarithm after adding one to original values because educational expenditures are large monetary values and the Tobit model crucially relies on the normality assumption of the idiosyncratic error (Cameron & Trivedi 2010). I use the same covariates as my main analysis, except that I add to covariates the number of children attending private schools. I obtain SE by bootstrapping with repetitions 200.

In Columns (2) and (3), educational expenses for books and materials and uniforms increase in pri-primary and primary schools if farmers cultivate in both rainy seasons, and decrease as asset levels increase. In Column (6), expenses for other contributions in primary schools also increase by the double-cropping, and decrease as asset levels increase. These results are in line with the observation in the Background Information that farmers must incur educational costs such as books and materials, uniform, and informal contributions even in the compulsory education. Therefore, the additional agricultural income in the short rainy season enables farmers to address educational expenses uncovered in public educational services. Although books, materials, uniforms, and contributions may not be essential for the

Table 5.7: CF-CRE Tobit Estimates of Educational Expenses

Explanatory Variable	(1) School Fee	(2) Books & Material	(3) Uniform	(4) Transport	(5) Extra Tuition	(6) Other Contrib.	(7) Cost of Meals	(8) Total Exp.
			(a) Educational Expenses in Pre-Primary Schools					
D_{it} (double-cropping)	1.21	2.65**	3.37**	0.34	-1.97**	1.82	0.01	3.50***
(dummy)	(0.22)	(0.03)	(0.02)	(0.55)	(0.03)	(0.11)	(0.99)	(0.01)
$D_{it} \times \ln(Asset_{it}+1)$ (inter-	-0.10	-0.20**	-0.25**	-0.03	0.10	-0.12	-0.01	-0.27***
action term)	(0.16)	(0.02)	(0.01)	(0.44)	(0.11)	(0.16)	(0.84)	(0.00)
$\ln(Asset_{it}+1)$ (log of house-	0.03	0.07	0.08	0.01	-0.01	0.04	-0.02	-0.03
hold asset values in Tsh)	(0.43)	(0.17)	(0.19)	(0.64)	(0.78)	(0.45)	(0.68)	(0.62)
			(b) Educational Expenses in Primary Schools					
D_{it} (double-cropping)	0.26	3.13***	3.66***	0.02	-1.55**	3.18***	-0.40	3.32***
(dummy)	(0.47)	(0.01)	(0.01)	(0.94)	(0.02)	(0.00)	(0.60)	(0.01)
$D_{it} \times \ln(Asset_{it}+1)$ (inter-	-0.02	-0.21***	-0.26***	0.00	0.08*	-0.19**	0.00	-0.24***
action term)	(0.56)	(0.01)	(0.01)	(0.91)	(0.08)	(0.02)	(0.93)	(0.01)
$\ln(Asset_{it}+1)$ (log of house-	0.02	0.08	0.08	0.00	-0.00	0.05	-0.02	-0.01
hold asset values in Tsh)	(0.13)	(0.10)	(0.13)	(0.75)	(0.99)	(0.24)	(0.51)	(0.93)
			(c) Educational Expenses in Lower-Secondary Schools					
D_{it} (double-cropping)	0.02	-0.16	0.13	-0.04	-0.48	-1.09	-0.32	-0.22
(dummy)	(0.98)	(0.86)	(0.89)	(0.92)	(0.44)	(0.18)	(0.64)	(0.85)
$D_{it} \times \ln(Asset_{it}+1)$ (inter-	-0.01	0.00	-0.02	0.00	0.01	0.06	0.02	-0.01
action term)	(0.83)	(0.98)	(0.75)	(0.89)	(0.81)	(0.28)	(0.63)	(0.93)
$\ln(Asset_{it}+1)$ (log of house-	0.00	0.00	0.02	-0.01	0.01	-0.00	-0.03	-0.03
hold asset values in Tsh)	(0.99)	(0.93)	(0.65)	(0.65)	(0.58)	(0.95)	(0.18)	(0.45)

Note: Sample size: $N = 11,031$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 with p-values. I obtain SE by bootstrapping with repetitions 200.
Other covariates omitted from the table: age, gender, and educational years by household head, adult equivalents, numbers of children aged 5-17 and adults aged 18-60, acres of owned land, membership to SACCO, residence in urban areas, holding of non-agricultural occupations, rainfall in cultivation period and sum of rainfall over five years in each long and short rainy season, area and wave dummies, number of children attending private schools, time-averaged values

schooling of children, these results suggest that farmers who obtain additional agricultural income can afford to allocate more income to the children's education. In Column (8), the total educational expenditures in pre-primary and primary schools are also significantly increased if farmers cultivate in both rainy seasons, although extra tuition decreases in these schools.

In contrast, I find no statistical significance in lower-secondary schools. This insignificance may be attributed to the selective nature of lower-secondary schools. Pre-primary and primary schools are open to wide ranges of children in Tanzania. Children may be able to attend pre-primary and primary schools even if they have no books, materials, and uniforms. However, children must pass the PSLE to go on to lower-secondary schools (the pass rate is 38%, as was explained). Households who intend to make their children go on to lower-secondary schools need to be well-motivated and finish preparing for their schooling from the period of enrollment. Thus, the double-cropping may not increase educational expenditure for specific items in lower-secondary schools.

5.6.2 Gender-Differential Impacts

I further examine whether the double-cropping affects the schooling of children differently by girls and boys. Combining the pooled CRE probit/Tobit models with the CF approach, Table 5.8 displays estimation results for the probability of children's enrollment and the number of children continuing schools separately by (a) girls and (b) boys. The interaction term $D_{it} \times \ln(Asset_{it} + 1)$ is included in covariates.

In Columns (2) and (3) of Table 5.8(a), the double-cropping increases the probability of girls' enrollment in primary and lower-secondary schools, while these effects decrease as asset levels increase (both F tests are significant). Moreover, the estimated coefficients for D_{it} in primary and lower-secondary schools, 0.18 and 0.07, are similar to the estimated coefficients in Columns (6) and (7) of Table 5.5(b) (0.20 and 0.10). In contrast, I find no statistical significance for boys' enrollment in these schools, although the coefficients for D_{it} are positive in all columns. Therefore, the previous results in Table 5.5(b) strongly reflect girls' school enrollment.

In Columns (5)-(7), I find no evidence that the double-cropping promotes girls to continue to attend schools. In contrast, I find some significant results for boys. In Column (5), the double-cropping discourages boys from continuing to attend pre-primary schools, while this negative effect becomes weaker as asset levels rise. This odd result may be attributed to my definition of the school continuation, which does not count children's enrollment in primary schools after graduating from pre-primary schools (i.e., the transition from PP to D1 in Table 5.1). Farmers who obtain additional income by the double-cropping may make their boys enroll in primary schools (not significant though) rather than stay in pre-primary

Table 5.8: CF-CRE Probit/Tobit Estimates of APEs for Girls' and Boys' School Enrollment and Continuation

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
(a) Schooling of girls							
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.08	0.18**	0.07	0.11	0.03	-0.18	-0.08
	(0.18)	(0.05)	(0.17)	(0.18)	(0.40)	(0.30)	(0.33)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.01	-0.01	-0.01*	-0.01*	-0.00	0.01	0.01
	(0.21)	(0.15)	(0.06)	(0.09)	(0.23)	(0.26)	(0.24)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	0.00	-0.00	0.00	0.00	-0.00	-0.01*	-0.01**
	(1.00)	(0.77)	(0.30)	(0.38)	(0.97)	(0.08)	(0.01)
F statistic (p-values in parenthesis							
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it})+1$)	1.78	5.86*	5.23*	4.40	2.25	1.29	1.55
	(0.41)	(0.05)	(0.07)	(0.11)	(0.32)	(0.52)	(0.46)
(b) Schooling of boys							
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.02	0.04	0.05	0.02	-0.07**	0.28	0.07
	(0.72)	(0.67)	(0.33)	(0.75)	(0.04)	(0.18)	(0.35)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.00	0.00	-0.00	0.00	0.01**	-0.03**	-0.00
	(0.43)	(0.91)	(0.54)	(1.00)	(0.03)	(0.05)	(0.39)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	0.00	0.00	-0.00	-0.00	0.01*	0.00
	(0.91)	(0.35)	(0.24)	(0.75)	(0.36)	(0.05)	(0.14)
F statistic (p-values in parenthesis)							
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it})+1$)	2.09	3.33	1.91	1.06	4.79*	8.33**	0.89
	(0.35)	(0.19)	(0.39)	(0.59)	(0.09)	(0.02)	(0.64)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I obtain SE by bootstrapping with 200 repetitions in all columns.

Other covariates omitted from the table: age, gender, and educational years by household head, adult equivalents, numbers of children aged 5-17 and adults aged 18-60, acres of owned land, membership to SACCO, residence in urban areas, holding of non-agricultural occupations, rainfall in cultivation period and sum of rainfall over five years in each long and short rainy season, area and wave dummies, time-averaged values schools, which may decrease boy's continuation of pre-primary schools. This problem does not occur in primary and lower-secondary schools because children at the top grades of these schools can be distinguished in the NPS.

However, D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are jointly significant in Column (6), suggesting that the double-cropping promotes boys to continue to attend primary schools, while its effects decrease by asset levels. The positive coefficient for D_{it} (0.28) is in contrast with the negative coefficient in girls (-0.18), which may obscure the impacts of the double-cropping in Column (5) of Table 5.6(b). Some studies found that households tend to protect boys' education against negative shocks at the expense of education of older girls, who are responsible for household chores (Gubert and Robilliard 2008; Björkman-Nyqvist 2013). Farmers may seek to protect investments in boys' education by preventing their dropping-out (Thomas et al. 2004).

5.6.3 Child Labor in the Short Rainy Season

I examine child labor in the short rainy season because children's working on the farm may discourage their schooling (Amin et al. 2006; Emerson et al. 2017; Dunne and Humphrey 2022). Figure 5.8 compares child and adult labor for (i) land preparation and planting, (ii) weeding, ridging, and fertilizing, and (iii) harvesting and threshing in each of the long and short rainy seasons (Value scales are adjusted between children and adults for viewability). In all farm activities, the amounts of child labor are smaller in the short rainy season by around two days than in the long rainy season. Moreover, days of child labor are even smaller than days of adult labor in both rainy seasons except for harvesting and threshing, which suggests that Tanzanian farmers do not heavily rely on child labor in agricultural activities, in particular during the short rainy season.

Summing up days of child labor over all agricultural activities, I estimate the impacts of the double-cropping on total days of child labor during the short rainy season. Combining the pooled Tobit model with the CRE and CF approach under the same covariates as the main analysis, Table 5.9 displays estimation results. Table 5.9 shows that the double-cropping increases total days of child labor by around 18 days among the poorest households with no asset, which is consistent with the observation in Figure 5.8 that the average days of child labor are around 6 days in each of the three agricultural activities. Because the F statistic is significant, total days of child labor decrease by 0.80 days if household asset values increase by 1%.

I then examine whether increased child labor during the short rainy season discourages the schooling of children. Combining the pooled probit/Tobit models with the CRE and CF approach, Table 5.10 shows estimation results of children's school enrollment and continuation in each school. I replace D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ by total days of child labor

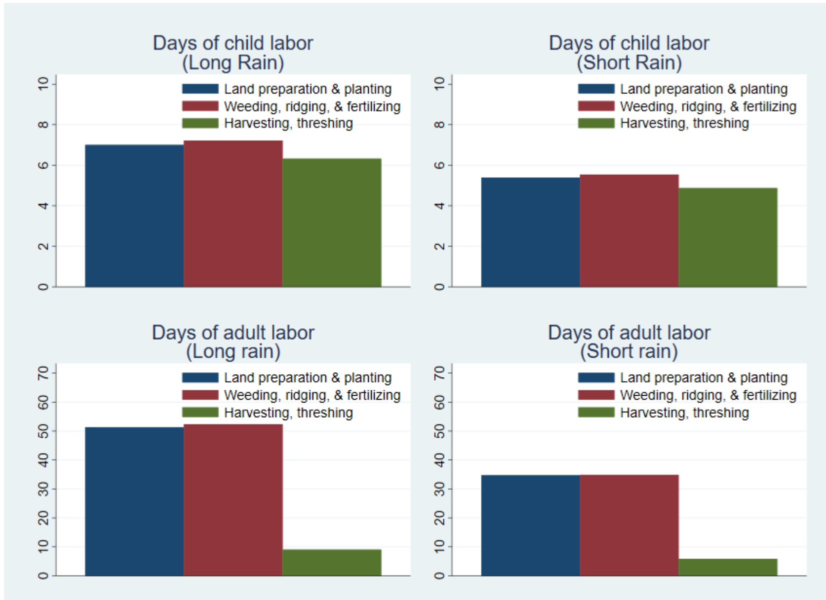


Figure 5.8: Days of Child and Adult Labor for the Cultivation in Long and Short Rainy Seasons

Table 5.9: Pooled CF-CRE Tobit Estimates of APEs for Total Days of Child Labor in the Short Rainy Season

Explanatory Variables	Total Days of Child Labor in the Short Rainy Season
D_{it} (double-cropping) (dummy)	18.17**
	(0.02)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.80
	(0.41)
$\ln(Asset_{it}+1)$ (log of asset values)	0.59***
	(0.01)
F statistic (p-values in parenthesis): $\beta_{1j} = \beta_{2j} = 0$	16.89***
	(0.00)

Note: Sample size: $N = 11,031$. Significance levels: * p<0.1, ** p<0.05, *** p<0.01 in parenthesis. I obtain SE by bootstrapping with 200 repetitions.
I use the same covariates as Table A22, though they are omitted from the table.

Table 5.10: CF-CRE Probit/Tobit Estimates of APEs for Children's Enrollment in and Continuation of School on Total Days of Child Labor during the Short Rainy Season

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
Total days of child labor in short rain	-0.00	-0.00	-0.00	0.00	-0.00**	-0.00	-0.00
	(0.38)	(0.23)	(0.79)	(0.82)	(0.03)	(0.61)	(0.66)
Interaction term between child labor days in short rain and $\ln(Asset_{it}+1)$	0.00*	0.00	0.00	0.00	0.00	0.00	0.00
	(0.09)	(0.11)	(0.49)	(0.64)	(0.17)	(0.83)	(0.87)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.01	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.17)	(0.80)	(0.68)	(0.64)	(0.63)	(0.81)	(0.67)
F statistic (p-values in parenthesis							
$\beta_{1j} = \beta_{2j} = 0$	3.16	2.64	0.67	1.57	4.97*	0.54	0.32
	(0.21)	(0.27)	(0.72)	(0.46)	(0.08)	(0.76)	(0.85)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I obtain SE by bootstrapping with 200 repetitions in all columns.
Other covariates omitted from the table: age, gender, and educational years by household head, adult equivalents, numbers of children aged 5-17 and adults aged 18-60, acres of owned land, membership to SACCO, residence in urban areas, holding of non-agricultural occupations, rainfall in cultivation period and sum of rainfall over five years in each long and short rainy season, area and wave dummies, time-averaged values.

during the short rainy season and their interaction term with $\ln(Asset_{it}+1)$, although other covariates are the same as the main analysis. In the first stage of the CF approach, I use the same IVs because child labor in the short rainy season is necessary only when farmers cultivate in this season. I find that the proportion of the double-cropping for each district and NINO3.4 are individually significant in the two estimations of child labor days and their interaction term, although the results are not displayed here.

In Column (1), the interaction term is significantly positive, which suggests that households with more assets can make their children enroll in pre-primary schools. In Column (5), child labor in the short rainy season discourages children from continuing to attend pre-primary schools. Because the F statistic is significant, households with more assets can make their children attend pre-primary schools without dropping out.

However, I find no statistical significance in other types of schools, although the coefficient for child labor days in the short rainy season is negative in all columns. Moreover, all the coefficients for child labor days are tiny. In Column (5), if total days of child labor during the short rainy season are 18 days, the number of children who continue to attend pre-primary schools declines by only around 0.06 among the poorest households with no asset. Although child labor during the short rainy season may interrupt daily school activities by children, it does not have clear adverse effects on school enrollment and continuation by children.

5.6.4 Impacts of the Educational Reform in 2016

Finally, I examine how the impacts of the double-cropping were changed by the fee-free education. I introduce a period dummy equaling one after the reform and create its interaction terms with D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$. Appendix 5-II shows that the double-cropping after the educational reform is negatively associated with the number of non-schooling children at official ages of pre-primary schools (5-6). This result coincides with the explanation in the Background Information that the GER for pre-primary schools considerably increased in 2016. The educational reform may contribute to early education of children through enhancing the potential role of the short rainy season in Tanzania.[15]

5.7 Conclusion

Using the six waves from 2008/09 to 2020/21 of the Tanzanian NPS, I investigated whether farmers' double-cropping over the long and short rainy seasons promotes children's enrollment and attendance at schools. Controlling for unobserved time-constant and time-variant heterogeneity, I found that the double-cropping increases the probability of children's enrollment in primary and lower-secondary schools. Moreover, farmers who cultivate in both rainy seasons make their girls, rather than boys, enroll in these schools. In contrast, the double-cropping promotes boys to continue to attend primary schools without dropping out. These benefits of the double cropping weaken as household asset levels rise, but remain over wide ranges of asset levels. I further showed that the additional agricultural income in the short rainy season enables farmers to address educational expenses uncovered in public educational services. Although the double-cropping increases child labor during the short rainy season, I find no clear adverse effects on school enrollment and continuation by children.

There are two policy implications. First, we should design policies to support the double-cropping by Tanzanian farmers in order to reduce educational inequality and promote female education. For example, fertilizer subsidies, irrigation, and land tenure may be desirable because the maintenance of soil nutrients, water management, and land ownership are important factors for the double-cropping, as shown in the previous chapter. Second, the cultivation in the short rainy season can be a strategic option for farmers to address wide economic issues. Farmers can address a lack of food and cash income by cultivating in the short rainy season. Therefore, the cultivation in the short rainy season may work as farmers' livelihood strategies to address the lack of cash necessary for life events such as marriage and moving. Further possibilities for the double-cropping should be explored in the future.

[15] I find no evidence on gender heterogeneity in this analysis.

Chapter 6

Conclusion

6.1 Findings of Each Chapter

This thesis has explored potential sources of heterogeneity of livelihood strategies by farmers' economic levels through crop diversification and cultivation in a secondary cropping season. In Chapter 2, I first constructed the economic framework of the subsistence constraint as a conceptual foundation of subsequent chapters. In each of Chapters 3-5, this thesis used this framework to conceptually explain how the adoption of crop diversification and welfare impacts of the double-cropping are different by farmers' economic levels. Each chapter conducted empirical analyses to provide supporting evidence for the conceptual framework, using the Tanzanian NPS.

The second chapter introduced a generalized functional form of the subsistence constraint ($f(x) \geq R$). I firstly showed that the subsistence constraint is binding for a solution ($f(x) = R$) when the MRS there does not coincide with the price ratio. This binding condition depicts a situation where the consumer prefers balanced consumption but has to specialize in certain goods for survival. If the budget constraint is also binding ($p \cdot x = I$), solutions under the subsistence constraint are given at the intersection between the budget and subsistence lines. This property was frequently used to depict solutions in the conceptual frameworks of subsequent chapters. I also showed that the price ratio lies between two marginal rates of substitution for a utility function u and subsistence function f at a solution. Finally, I propose a hypothesis for subsequent chapters on farmers' livelihood strategies. That is, farmers' production decisions and welfare impacts are heterogeneous between poor and non-poor farmers, based on their different purposes: to preserve the subsistence requirement or to improve the quality of livelihoods.

The third chapter showed that poor and non-poor farmers are differently motivated for crop diversification. Poor farmers, who fear the possibility that the subsistence requirement is not attained and who aspire to enjoy balanced dietary intakes, change crop diversity for robust food security. In contrast, non-poor farmers, who can afford to sell harvests beyond self-consumption, adopt crop diversification to stabilize market income. Therefore, different motivations for crop diversification between poor and non-poor result from different livelihood priorities, that is, food security v.s. market income.

The fourth chapter showed that the cultivation in a secondary cropping season improves farmers' food security during the lean season in Tanzania. Farmers who cultivate in the short rainy season gain more caloric intakes and realize more dietary diversity. These welfare improvements are stronger for smallholders with small farm size. In particular, farmers who have no occupations except agriculture gain more caloric intakes by cultivating in the short rainy season, whereas farmers have non-farm occupations realize more dietary diversity by the cultivation. The cultivation in the short rainy season may have different channels of welfare improvements because farmers' livelihood priorities are different by the existence of off-farm income sources. If farmers have no income sources during the lean season, they suffer from extreme poverty and first and foremost prioritize consuming high-caloric food to ease hunger. Thus, these farmers spend additional agricultural income in the short rainy season on improving caloric intakes. In contrast, if farmers have off-farm income during the lean season, they meet basic caloric needs but still aspire to enjoy consuming a variety of foods in severe conditions of food markets during the lean season. Thus, they spend additional agricultural income in the short rainy season on improving dietary diversity.

The fifth chapter showed that the double-cropping in a secondary cropping season after the main one improves the schooling of children among Tanzanian farmers. If farmers cultivate in both two rainy seasons, their children, in particular girls, are more likely to enroll in primary and lower-secondary schools. In contrast, farmers who cultivate in both two rainy seasons make boys continue to attend primary schools without dropping out from there. Moreover, these effects are stronger for asset-poor farmers than for asset-rich farmers. The effects of the double-cropping on the schooling of children may be different in magnitude because of different priorities of consumption between necessities and non-necessities. Asset-poor farmers prioritize consuming necessities in their livelihoods (e.g., food) and restrict their consumption of non-necessities (education). Thus, the additional agricultural income in the short rainy season enables them to consume more amounts of non-necessities relative to necessities. In contrast, asset-rich farmers, whose consumption of non-necessities are not restricted, do not change their consumption patterns even if they obtain additional agricultural income in the short rainy season.

The fourth and fifth chapters showed that the double-cropping is more effective for poor farmers with small farm size and low levels of assets than for wealthy farmers. Because the double-cropping has a potential to reduce farmers' economic inequality, we should target poor farmers when promoting the implementation of the double-cropping.

6.2 Policy Implication

The findings in the preceding three chapters support the hypothesis in Chapter 2. I found that the adoption of crop diversification and the welfare impacts of the cultivation in the secondary cropping season are heterogeneous by farmers' economic levels. As shown in Chapter 2, the subsistence constraint becomes binding when income is low or prices of necessities are high. Thus, poor farmers care about whether they fall into low income or when prices of necessities become high. In contrast, non-poor consumers does not need to care about the subsistence constraint and attempt to improve the quality of livelihood (e.g., dietary diversity, non-food consumption), rather than to preserve livelihood requirement (e.g., caloric intakes). The two livelihood strategies in this thesis reduce the possibility that farmers fall into low income and improve their food security when food prices tend to be high, which leads to the heterogeneity by farmers' economic levels.

The findings in the preceding three chapters suggest that the heterogeneity of farmers' livelihood strategies comes from different livelihood priorities by economic levels. Poor farmers have to preserve the minimum requirements of livelihoods. Thus, poor farmers' motivations to adopt livelihood strategies and their welfare impacts are closely linked to types of livelihood requirement behind (e.g., basic caloric needs). In contrast, non-poor farmers can afford to pursue livelihood strategies without caring about the minimum requirement of livelihoods. To clarify sources of heterogeneity in farmers' livelihood strategies, we must consider how consumption patterns of poor farmers are restricted for survival, or put it in another way, a form of the subsistence constraint.

Considering different livelihood priorities by farmers' economic levels is useful to build policy tools to promote livelihood strategies. For example, consider a subsidy program that provides farmers with seeds. If we aim to promote crop diversification for poverty reduction, we should supply seeds of drought-resistant crops (e.g., sorghum, legumes) to prevent devastating climatic damages to the food security of poor farmers. In contrast, if we aim to promote crop diversification for risk reduction in markets, we should supply non-poor farmers with seeds that are expected to produce stable market income if they are planted.

On the other hand, if we aim to promote the double-cropping by providing a seed subsidy during the short rainy season, we should target poor farmers with small farm size and low levels of assets. Then, we should supply seeds of staple crops with high caloric contents (e.g., maize) to farmers who cannot rely on non-agricultural income sources and who suffer from extreme poverty during the lean season. In contrast, we should supply seeds of non-staple crops (e.g., legumes, vegetables) to farmers who can rely on non-agricultural income sources but who still aspires to enjoy consuming a variety of food.

Therefore, when we construct a policy of livelihood strategies, we must consider whether farmers' priorities in their livelihoods are different by economic levels, and if so, how they are different. Then, we should determine the purpose of the policy in advance to target specific farmer groups. For example, we must build policies of crop diversification after determining whether to secure the food security of poor farmers or stabilize the market income of non-poor farmers. These clarifications enable us to build policies of livelihood strategies which are suitable for farmers' livings. The findings in this thesis provide theoretical backgrounds and empirical evidence to make better agricultural support programs for farmers.

6.3 Future Work

Finally, I state two possibilities of future research. The first possibility is to explore an empirical foundation for the subsistence constraint, although this thesis assumed a specific form of the subsistence constraint depending on each research topic. Importantly, this thesis derived microeconomic properties without considering specific utility functions. As Chapter 2 shows, consumer solutions under the subsistence constraint are given at the intersection between the budget and subsistence lines. My theoretical framework relies on the identification of the budget constraint and a form of the subsistence constraint. Thus, my future research aims to consider various forms of the subsistence constraint and test the validity for each constraint form, using data on individuals' income and food choices in low-income countries. If we find a form of the subsistence constraint which is consistent with real consumption behaviors in low-income countries, its applicability will be enhanced in development economics.

The second possibility of future research is to build policies to promote farmers' cultivation in the short rainy season in Tanzania. Through determinant analyses in Chapter 4, I showed that the maintenance of soil nutrients, water management, and land ownership are important factors for the double cropping. To suggest the double-cropping as a policy tool, my future research aims to conduct a randomized control trial providing Tanzanian farmers with fertilizer subsidies during the short rainy season, for example, and rigorously estimate the impacts on farmers' practices of the double-cropping and on their welfare.

Supplementary Appendices

Appendices for Chapter 2

Appendix 2-I. Supplementary Information on Theorem 2.2.3

Counterexample of Theorem 2.2.3 at a solution in the boundary of $\mathbb{R}^N_+$

Figure A1 provides a counterexample where Theorem 2.2.3 does not hold when x^{sub} is in the boundary of $\mathbb{R}^N_+$. The functional forms of u and f are given by a linear form with $u_i(x) = c_i > 0$ for all $x \in V$ and $f_i(x) = \epsilon_i > 0$ for all $x \in \mathbb{R}^N_+$. There is only one feasible consumption bundle x^{sub}, which lies at the intersection between the budget and subsistence lines and is the solution. In Figure A1, $\dfrac{p_1}{p_2} < \dfrac{f_1(x^{sub})}{f_2(x^{sub})} < \dfrac{u_1(x^{sub})}{u_2(x^{sub})}$, where the price ratio does not lies between the MRS and the subsistence line.

Ordering between $u_i(x^{sub})/u_j(x^{sub})$ and p_i/p_j when "$f_i(x^{sub})/f_j(x^{sub}) = 0/0$" in Theorem 2.2.3

In Figure A2, there is only one feasible consumption bundle x^{sub}, which is also the solution. Moreover, we have $f_1(x^{sub}) = f_2(x^{sub}) = 0$. Two utility functions in Figure A2, u and v, have indifference curves with different slopes at x^{sub}, that is, $\dfrac{u_1(x^{sub})}{u_2(x^{sub})} < \dfrac{p_1}{p_2} < \dfrac{v_1(x^{sub})}{v_2(x^{sub})}$. Therefore, any ordering between the MRS and price ratio is allowed when "$f_i(x^{sub})/f_j(x^{sub}) = 0/0$."

Proofs on Theorem 2.2.5 and Theorem 2.2.6

Lemma A1. *Take any x, $x' \in \mathbb{R}^N_+$ such that $x \neq x'$ and $p \cdot x = p \cdot x' = I$. Let $h_{ij} \in \mathbb{R}^N_+$ be such that the i-th element is $\dfrac{p_j}{p_i}$, the j-th element is -1, and the other elements are zero. Let $\mathcal{I} = \{(i,j) \in \mathbb{N} \times \mathbb{N} \mid 1 \leq i,\, j \leq N,\, x_i - x_i^{'} > 0,\, x_j - x_j^{'} < 0\}$. Then,*

$$x - x' = \sum_{(i,j)\in\mathcal{I}} c_{ij} h_{ij}$$

for some $c = (c_{ij})_{(i,j)}$ such that $c_{ij} > 0$ for all i, $j \in \mathcal{I}$.

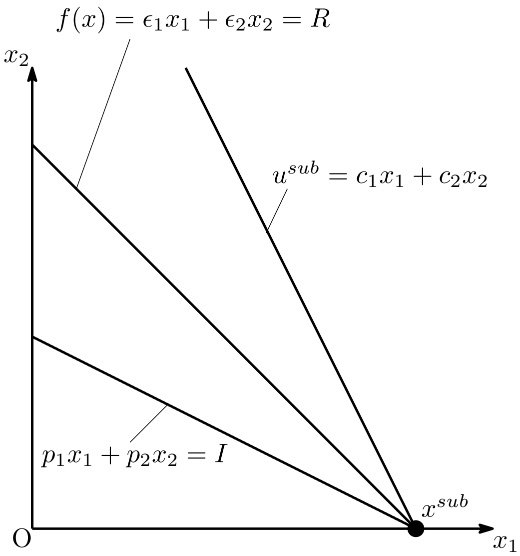


Figure A1: Counterexample of Theorem 2.2.3.

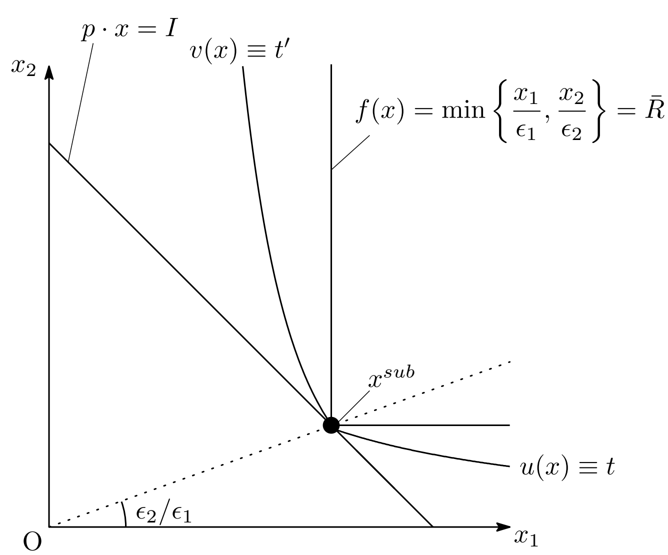


Figure A2: Example of "$f_i(x^{sub})/f_j(x^{sub}) = 0/0$."

Proof. Let $z := x - x'$ and z_i be its i-th element. Denote by n the number of nonzero elements in z ($1 \leq n \leq N$). We use induction on n.

Consider $n = 2$ first. Without loss of generality, suppose $z_i \neq 0$ for $i = 1, 2$ and $z_j = 0$ for $j = 3, \ldots, N$. Note $\dfrac{p_2}{p_1} = -\left(\dfrac{z_1}{z_2}\right)$ because $p_1 z_1 + p_2 z_2 = 0$. If $z_1 > 0$ and $z_2 < 0$, then $z = c_{12} h_{12}$, where $c_{12} = -z_2 > 0$. If $z_1 < 0$ and $z_2 > 0$, then $z = c_{21} h_{21}$, where $c_{21} = -z_1 > 0$. Thus, the claim is true for $n = 2$.

Assume that the claim is true for $n = k$. Consider $n = k + 1$. Because $p \cdot z = 0$, we must have both positive and negative elements in z. Without loss of generality, suppose $z_i \neq 0$ for $i = 1, \ldots, k$, $z_{k+1} > 0$, and $z_j = 0$ for $j = k + 2, \ldots, N$. Because $p \cdot z = 0$, we have $z_{k+1} = -\displaystyle\sum_{i=1}^{k} \left(\frac{p_i}{p_{k+1}}\right) z_i$.

If $z_i < 0$ for all $1 \leq i \leq k$, we have $z = \sum_{i=1}^{k} c_{k+1,i} h_{k+1,i}$, where $c_{k+1,i} = -z_i > 0$. Thus, suppose $z_i > 0$ for some $1 \leq i \leq k$. For simplicity, suppose that $z_i > 0$ for all $1 \leq i \leq l$ and $z_j < 0$ for all $l + 1 \leq j \leq k$. Note $z_{k+1} = v_{k+1} + w_{k+1}$, where $v_{k+1} = -\displaystyle\sum_{i=1}^{l} \left(\frac{p_i}{p_{k+1}}\right) z_i$ and $w_{k+1} = -\displaystyle\sum_{j=l+1}^{k} \left(\frac{p_j}{p_{k+1}}\right) z_j$. Thus, z can be separated to two parts as follows:

$$z = (\underbrace{z_1, \ldots, z_l}_{l\,elements}, \underbrace{0, \ldots, 0}_{(k-l)\,elements}, \underbrace{v_{k+1}}_{(k+1)th\,element}, \underbrace{0, \ldots, 0}_{n-(k+1)\,elements})$$
$$+ (\underbrace{0, \ldots, 0}_{l\,elements}, \underbrace{z_{l+1}, \ldots, z_k}_{(k-l)\,elements}, \underbrace{w_{k+1}}_{(k+1)th\,element}, \underbrace{0, \ldots, 0}_{n-(k+1)\,elements}).$$

In each of the first and second terms above, there are k nonzero elements at most. Recall that $z_1, \ldots, z_l > 0$ and $v_{k+1} < 0$, whereas $z_{l+1}, \ldots, z_k < 0$ and $w_{k+1} > 0$. Because the claim is true for k by the assumption of induction, z can be written as:

$$z = \sum_{i=1}^{l} c_{i,k+1} h_{i,k+1} + \sum_{j=l+1}^{k} c_{k+1,j} h_{k+1,j}$$

for some $c_{i,k+1} > 0\,(i = 1, \ldots, l)$ and $c_{k+1,j} > 0\,(j = l + 1, \ldots, k)$. Thus, the claim is true for $k + 1$. By the principle of induction, we have completed the proof. □

Theorem 2.2.5. *Let $N = 2$. Take any $(p, I) \in \mathbb{R}^3_{++}$ and any $\bar{x} \in B(p, I) \cap \mathbb{R}^2_{++}$ such that $f(\bar{x}) = R$ and $p \cdot \bar{x} = I$. Suppose that u is strictly quasi-concave, strictly increasing, and continuously differentiable in the neighborhood of $\bar{x}$. Suppose also that f is quasi-concave, nondecreasing, and continuously differentiable in the neighborhood of $\bar{x}$, Finally, suppose*

$u_i(\bar{x}) > 0$ and $f_i(\bar{x}) > 0$ $(i = 1, 2)$. If $\dfrac{u_1(\bar{x})}{u_2(\bar{x})} < \dfrac{p_1}{p_2} < \dfrac{f_1(\bar{x})}{f_2(\bar{x})}$ or $\dfrac{f_1(\bar{x})}{f_2(\bar{x})} < \dfrac{p_1}{p_2} < \dfrac{u_1(\bar{x})}{u_2(\bar{x})}$ holds, then $\bar{x}$ is the unique solution, or $X^{sub}(p, I) = \{x^{sub}\}$.

Proof. Because u is strictly quasi-concave and $B(p, I)$ is a convex set by the quasi-concavity of f, $\bar{x}$ is the unique global maximizer if it is a local maximizer. Thus, we will show that $\bar{x}$ is a local maximizer.

Because u is strictly increasing and f is nondecreasing, $p \cdot x = I$ must hold at a solution. Thus, take any $\epsilon > 0$ and any $v \in \mathbb{R}^N_+$ such that $||v|| = 1$ and $p \cdot v = 0$. Let $z = \epsilon v$.

If $v_1 > 0$ and $v_2 < 0$, then $v = c_{12}h_{12} = c_{12}\left(\dfrac{p_2}{p_1}, -1\right)$ $(c_{12} > 0)$ by Lemma A1. By the first-order Taylor expansion, we have

$$f(\bar{x}+z) = f(\bar{x})+\nabla f(\bar{x})\cdot z+o(||z||) = f(\bar{x})+\epsilon\times(\nabla f(\bar{x})\cdot v)+o(\epsilon) = f(\bar{x})+c_{12}\epsilon\times\left(f_1(\bar{x})\frac{p_2}{p_1} - f_2(\bar{x})\right)+o(\epsilon).$$

If $\dfrac{f_1(\bar{x})}{f_2(\bar{x})} < \dfrac{p_1}{p_2}$, we would have $f(\bar{x} + z) < f(\bar{x}) = R$ for all sufficiently small $\epsilon > 0$. Thus, $\dfrac{u_1(\bar{x})}{u_2(\bar{x})} < \dfrac{p_1}{p_2} < \dfrac{f_1(\bar{x})}{f_2(\bar{x})}$ must hold. Then,

$$u(\bar{x} + z) = u(\bar{x}) + \nabla u(\bar{x}) \cdot z + o(||z||) = u(\bar{x}) + c_{12}\epsilon \times \left(u_1(\bar{x})\frac{p_2}{p_1} - u_2(\bar{x})\right) + o(\epsilon),$$

which implies $u(\bar{x} + z) < u(\bar{x})$ for all sufficiently small $\epsilon > 0$.

If $v_1 < 0$ and $v_2 > 0$, then $v = c_{21}h_{21} = c_{21}\left(-1, \dfrac{p_1}{p_2}\right)$ $(c_{21} > 0)$ by Lemma A1. By the first-order Taylor expansion, we have

$$f(\bar{x} + z) = f(\bar{x}) + c_{21}\epsilon \times \left(f_2(\bar{x})\frac{p_1}{p_2} - f_1(\bar{x})\right) + o(\epsilon).$$

If $\dfrac{p_1}{p_2} < \dfrac{f_1(\bar{x})}{f_2(\bar{x})}$, we would have $f(\bar{x} + z) < f(\bar{x}) = R$ for all sufficiently small $\epsilon > 0$. Thus, $\dfrac{f_1(\bar{x})}{f_2(\bar{x})} < \dfrac{p_1}{p_2} < \dfrac{u_1(\bar{x})}{u_2(\bar{x})}$ must hold. Then,

$$u(\bar{x} + z) = u(\bar{x}) + c_{21}\epsilon \times \left(u_2(\bar{x})\frac{p_1}{p_2} - u_1(\bar{x})\right) + o(\epsilon),$$

which implies $u(\bar{x} + z) < u(\bar{x})$ for all sufficiently small $\epsilon > 0$.

Because $\bar{x}$ is a local maximizer, we have completed the proof. □

Theorem 2.2.6. *Let $N = 2$. Take any $(p, I) \in \mathbb{R}^3_{++}$ and any $\bar{x} \in B(p, I) \cap \mathbb{R}^2_{++}$ such that $f(\bar{x}) = R$ and $p \cdot \bar{x} = I$. Suppose that u is strictly quasi-concave, strictly increasing, and* *continuously differentiable in the neighborhood of $\bar{x}$. Suppose also that f is quasi-concave, nondecreasing, and continuously differentiable in the neighborhood of $\bar{x}$, Finally, suppose $u_k(\bar{x}) > 0\,(k = 1, 2)$, $f_i(\bar{x}) > 0$, and $f_j(\bar{x}) = 0$. If $\dfrac{u_i(\bar{x})}{u_j(\bar{x})} < \dfrac{p_i}{p_j}$ holds, then $\bar{x}$ is the unique solution, or $X^{sub}(p, I) = \{x^{sub}\}$.*

Proof. As in the proof of Theorem 2.2.5, we will show that $\bar{x}$ is a local maximizer. Because u is strictly increasing and f is nondecreasing, $p \cdot x = I$ must hold at a solution. Thus, take any $\epsilon > 0$ and any $v \in \mathbb{R}^N_+$ such that $||v|| = 1$ and $p \cdot v = 0$. Let $z = \epsilon v$.

If $v_i < 0$ and $v_j > 0$, then $v = c_{ji}h_{ji}\,(c_{ji} > 0)$ by Lemma A1. By the first-order Taylor expansion, we have

$$f(\bar{x} + z) = f(\bar{x}) + c_{ji}\epsilon \times \left(f_j(\bar{x})\frac{p_i}{p_j} - f_i(\bar{x}) \right) + o(\epsilon).$$

Because $f_i(\bar{x}) > 0$ and $f_j(\bar{x}) = 0$, we have $f(\bar{x} + z) < f(\bar{x}) = R$ for all sufficiently small $\epsilon > 0$. Thus, $v_i > 0$ and $v_j < 0$ must hold. Then, $v = c_{ij}h_{ij}\,(c_{ij} > 0)$ by Lemma A1. By the first-order Taylor expansion, we have

$$f(\bar{x} + z) = f(\bar{x}) + c_{ij}\epsilon \times \left(f_i(\bar{x})\frac{p_j}{p_i} - f_j(\bar{x}) \right) + o(\epsilon).$$

Thus, we have $f(\bar{x} + z) > f(\bar{x}) = R$ for all sufficiently small $\epsilon > 0$. Because $\dfrac{u_i(\bar{x})}{u_j(\bar{x})} < \dfrac{p_i}{p_j}$, we have

$$u(\bar{x} + z) = u(\bar{x}) + c_{ij}\epsilon \times \left(u_i(\bar{x})\frac{p_j}{p_i} - u_j(\bar{x}) \right) + o(\epsilon),$$

which implies $u(\bar{x} + z) < u(\bar{x})$ for all sufficiently small $\epsilon > 0$. Because $\bar{x}$ is a local maximizer, we have proved the theorem.

$\square$

Appendices for Chapter 3

Appendix 3-I. Mathematical Details in the Conceptual Framework

Mathematical Definitions of Concavity and Quasi-Concavity and their implications

In the conceptual framework, the utility function in the second stage, u, is assumed to be strictly concave. Mathematically, we say u is *strictly concave* in $(c,q) \in \mathbb{R}_+^{J+K}$ if $u\big(\alpha(c,q)+(1-\alpha)(c',q')\big) > \alpha u(c,q)+(1-\alpha)u(c',q')$ for all $\alpha \in (0,1)$ and all $(c,q),(c',q') \in \mathbb{R}_+^{J+K}$ with $(c,q) \neq (c',q')$. Importantly, u is strictly quasi-concave if it is strictly concave. We say u is *strictly quasi-concave* in $(c,q) \in \mathbb{R}_+^{J+K}$ if $u\big(\alpha(c,q)+(1-\alpha)(c',q')\big) > \min\{u(c,q),u(c',q')\}$ for all $\alpha \in (0,1)$ and all $(c,q),(c',q') \in \mathbb{R}_+^{J+K}$ with $(c,q) \neq (c',q')$.

To illustrate farmers' desire to achieve balanced consumption, take two consumption bundles (c_1,q_1) and (c_2,q_2) with $(c_1,q_1) \neq (c_2,q_2)$ from the same indifference curve $\overline{u} \equiv u(c,q)$ $(\overline{u} \in \mathbb{R})$. By the definition of quasi-concavity, we have $u(\alpha(c_1,q_1)+(1-\alpha)(c_2,q_2)) > u(c_1,q_1) = u(c_2,q_2)$ for all $\alpha \in (0,1)$. Thus, the balanced consumption $\alpha(c_1,q_1)+(1-\alpha)(c_2,q_2)$ is preferable to (c_1,q_1) and (c_2,q_2).

To illustrate farmers' risk-averse preferences, consider a land allocation with high crop diversity E^H and a land allocation with low crop diversity E^L. Higher degrees of crop diversification guarantee secure consumption by stabilizing agricultural income $I_f(\omega,E)$. For simplicity, suppose that the farmer who adopts E^H realizes the expected consumption under E^L certainly for any state, that is,

$$c(\omega,E^H) = \sum_{\omega' \in \Omega} \pi(\omega';h)c(\omega',E^L), \quad q(\omega,E^H) = \sum_{\omega' \in \Omega} \pi(\omega';h)q(\omega',E^L) \tag{6.3.1}$$

for all $\omega \in \Omega$ (h is fixed for the farmer). Then, it follows from Equation (3.2.1) that

$$\begin{aligned}
U(E^H) &= \sum_{\omega \in \Omega} \pi(\omega;h)u\left(c(\omega,E^H),q(\omega,E^H)\right) \\
&= \sum_{\omega \in \Omega} \pi(\omega;h)u\left(\sum_{\omega' \in \Omega} \pi(\omega';h)c(\omega',E^L), \sum_{\omega' \in \Omega} \pi(\omega';h)q(\omega',E^L)\right) \quad (\because \text{Equation (6.3.1)}) \\
&> \sum_{\omega \in \Omega} \pi(\omega;h) \sum_{\omega' \in \Omega} \pi(\omega';h)u\left(c(\omega',E^L),q(\omega',E^L)\right) \quad (\because \text{strict concavity of } u) \\
&= \sum_{\omega' \in \Omega} \pi(\omega';h)u\left(c(\omega',E^L),q(\omega',E^L)\right) \quad (\because \sum_{\omega \in \Omega} \pi(\omega;h) = 1) \\
&= U(E^L).
\end{aligned}$$

Thus, the risk-averse farmer prefers E^H to E^L.

Brief Discussion of Three Potential Factors in Imperfect Markets

Throughout the conceptual framework, I assumed perfect neoclassical markets to simplify the argument. However, the essence of my argument on three potential factors is the same even in imperfect markets.

For simplicity, I first consider the extreme case where markets do not exist; farmers consume all of the crops produced by themselves, without consuming any product by others. If farmers produce good j linearly using an endowment e_j, the level of farmers' consumption realized in state ω becomes

$$(c_1(\omega), \ldots, c_J(\omega)) = (\alpha_1(\omega)e_1, \ldots, \alpha_J(\omega)e_J),$$

where $\alpha_j(\omega)$ is a productivity shock (note that market shocks, or price fluctuations, do not exist). Then, poor farmers prioritize satisfying the subsistence constraint for all states ω:

$$\epsilon_1\alpha_1(\omega)e_1 + \cdots + \epsilon_J\alpha_J(\omega)e_J \geq R$$

Because the level of consumption still depends on productivity shocks in each state, poor farmers specialize in producing staple food if high crop diversity involves risks of insufficient energy intakes (poverty traps). Conversely, they diversify crop portfolios if specialization in a certain crop involves risks of heavy damages (risk dispersion).

If agronomic synergies between crops exist and crop diversification improve productivity, the realized consumption becomes

$$(c_1(\omega), \ldots, c_J(\omega)) = \left(\alpha_1(\omega)e_1 + \sum_{k\neq 1} d_{1k}(\omega)e_1e_k, \ldots, \alpha_J(\omega)e_J + \sum_{k\neq J} d_{Jk}(\omega)e_Je_k\right),$$

where $d_{jk}(\omega) \geq 0$. Because of complementary effects between crops, $d_{jk}(\omega) \geq 0$, the subsistence constraint becomes

$$\epsilon_1\alpha_1(\omega)e_1 + \cdots + \epsilon_J\alpha_J(\omega)e_J + \sum_{j\neq k} \epsilon_j d_{jk}(\omega)e_je_k \geq R.$$

Then, the subsistence constraint is relaxed by $\sum_{j\neq k} \epsilon_j d_{jk}(\omega)e_je_k$. Thus, poor farmers whose choice is restricted by the subsistence constraint can realize more balanced consumption by crop diversification (balanced dietary intakes).

Finally, because non-poor farmers can afford to satisfy the subsistence constraint in all states, their main motivation for crop diversification is to stabilize consumption (not market income in this case).

Similar arguments still hold in an intermediate case between perfect markets and autarky.

Suppose that farmers must incur transaction costs $\tau > 1$ for market sales. Then, the farmers' budget constraint becomes

$$y_j(\omega) = \alpha_j(\omega)e_j + \sum_{j \neq k} d_{jk}(\omega)e_j e_k,$$
$$\sum_{j=1}^{J} p_{c_j}(\omega)c_j + \tau \left(\sum_{j=1}^{J} p_{c_j}(\omega)s_j \right) \leq \sum_{j=1}^{J} p_{c_j}(\omega)y_j + I_n,$$
$$\sum_{k=1}^{Q} p_{q_k}(\omega)q_k \leq \sum_{j=1}^{J} p_{c_j}(\omega)s_j + I_n,$$

where $y_j(\omega)$ is the amount of production for crop j and s_j is its amount of sales. To consume a non-food good q_k, farmers must obtain crop sales $\sum_{j=1}^{J} p_{c_j}(\omega)s_j$.

The same argument holds for the three potential factors, as I explain when markets do not exist. If poor farmers predict insufficient energy intakes at some states under diversified crop portfolios, they specialize in growing staple crops with high-caloric intakes to satisfy the subsistence constraint (poverty traps). Conversely, if specialization in a certain crop involves the risk of heavy damages, subsistence requirement induces poor farmers to disperse risks by growing multiple crops (risk dispersion). Poor farmers who realize complementary effects between crops can relax the choice restriction imposed by the subsistence constraint (balanced dietary intakes). In contrast, non-poor farmers can afford to consume non-food goods q_k relative to poor farmers, and hence want to stably consume them by stabilizing crop sales (stabilization of market income).

Note that neither transaction costs nor crop sales are irrelevant to the existence of the three potential factors. The existence of markets or transaction costs affect farmers' choices on self-consumption and crop sales. However, the prioritization toward subsistence requirement, the desire for balanced dietary intakes, and the incentive to stabilize market income still work because these factors are characterized by the subsistence constraint, rather than the assumption of perfect markets.

Appendix 3-II. Two Indicators of Crop Diversification

To assess the diversity of cultivated crops in farmland, there are several indicators of crop diversification, whose developments depend on how we interpret "diversity" (Duelli and Obrist 2003).

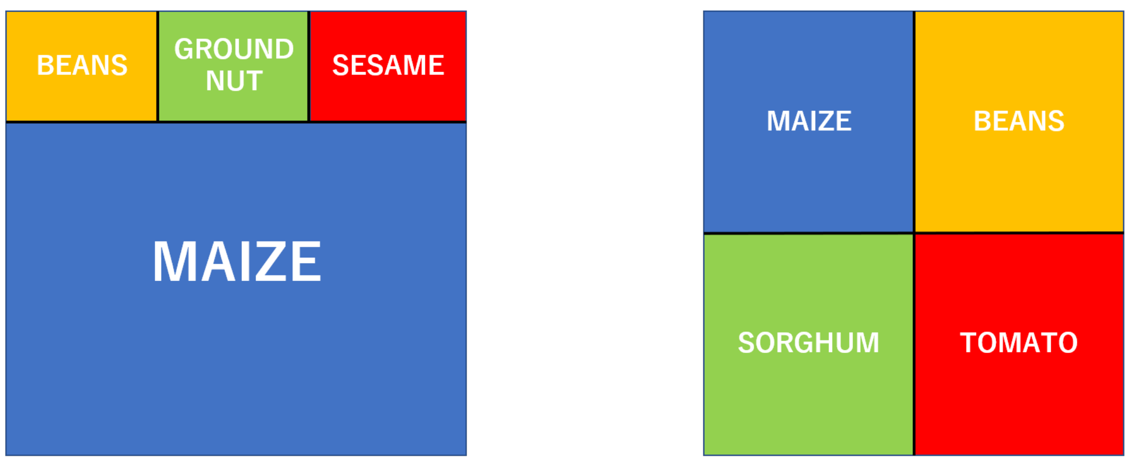
The simplest indicator is to count the number of crops cultivated in a farmland. Another frequently used indicator is the Simpson index (SI), which is defined as follows:

$$SI = 1 - \sum_c s_c^2 \tag{6.3.2}$$

where s_c is the area share of crop c in total farmland. The SI is large if a farmer plants many crops evenly on the farmland. Conversely, we have $SI = 0$, or complete specialization, if only one crop is planted on the whole land.

The choice of indicators is important because each indicator is closely connected to why farmers grow multiple crops. Figure A3 displays the hypothetical land allocations of crops by two farmers. The farmer in the left farmland allocates a large land share to maize, and the remaining marginal share to beans, groundnuts, and sesame. In this farmland, the remaining three crops serve as side crops to support the growth of the main crop (maize). Conversely, the farmer in the right-hand farmland cultivates maize, beans, sorghum, and tomatoes in even land allocations. Because both farmers cultivate four crops, the indicator of counting crop numbers is equal for the two farmers. However, the right-hand farmer outnumbers the left-hand one in the SI because the right-hand farmland is diversified in terms of evenness in land allocations. The SI values the evenness of land shares of planted crops as diversity, whereas the simple indicator of counting crop numbers allows the dominance of a main crop in farmland if side crops are planted.

My analysis relies on both indicators of crop diversification rather than on a single indicator. The use of more than one indicator sheds light on farmers' various motivations for crop diversification. In Figure A3, the left-hand farmer who encountered a rise in fertilizer price may have started planting new legumes as a nitrogenous substitute for fertilizer. The right-hand farmer who lacked the necessary cash for purchasing food may have produced each crop in equal land proportions to obtain balanced nutrition for self-consumption. If only one indicator were used, farmers' various motivations for crop diversification would not be observed.



Dominance of a main crop over the other side crops

Even land allocation of planted crops

Figure A3: Hypothetical Land Distributions of Crops Planted by Two Farmers

Appendix 3-III. Items of Five Broad Crop Types

In Descriptive Statistics, I classified all crops broadly into five types: staples, sorghum/tubers/roots, legumes/oil, fruits/vegetables, and cash crops. To construct these five crop types, I refer to the category in the NPS questionnaire. The questionnaire is available at `https://microdata.worldbank.org/index.php/catalog/2252`. I enumerate all items included in each crop type below:

- **Staples:** maize paddy, wheat, barley
- **Sorghum/roots/tubers:** sorghum, bulrush millet, finger millet, cassava, sweet potatoes, Irish potatoes, yams, cocoyams, onions, ginger
- **Legumes/oil:** beans, cowpeas, green gram, chickpeas, Bambara nuts, field peas, sunflower, sesame, groundnut, soyabeans, caster seed
- **Fruits/vegetables:** passion fruit, banana, avocado, mango, papaw, orange, grapefruit, grapes, mandarin, guava, plums, apples, pears, peaches, lime, lemon, pomelo, jack fruit, durian, bilimbi, rambutan, bread fruit, Malay apple, star fruit, custard apple, God fruit, mitobo, plum, peaches, pomegranate, date, tungamaa, vanilla, cabbage, tomatoes, spinach, carrot, chilies, amaranths, pumpkins, cucumber, eggplant, watermelon, cauliflower, okra, fiwi
- **Cash crops:** cotton, tobacco, pyrethrum, jute, seaweed, sisal, coffee, tea, cocoa, rubber, wattle, kapok, sugar cane, cardamom, tamarind, cinnamon, nutmeg, clove, black pepper, pigeon pea, pineapple, palm oil, coconut, cashew nut, green tomato, monkeybread, bamboo, firewood/fodder, timber, medicinal plant, fence tree

Appendix 3-IV. Effects of Household and Regional Characteristics

In Appendix 3-IV, I examine the household and regional characteristics that are estimated to affect crop diversification differently between the two regimes in Table 3.4. Old household heads in the lower regime tend to be more conservative in increasing crop numbers in farmlands than young ones. Lower-regime farmers with more educated household heads tend to increase both crop numbers and the SI. Because cultivation ways are specific to each crop, some educational levels may be helpful for growing multiple crops. Farmers with more household members, which are proxied by adult equivalents, pursue more specialized crop portfolios in both crop number and SI. This specialization affects lower-regime farmers more strongly than upper-regime farmers. Lower-regime farmers with limited agricultural resources may assign higher priority to cultivating staple food crops to feed many household members. Larger areas of owned land provide lower-regime farmers with more capability to increase the SI. Each of these household characteristics tends to have a larger coefficient in absolute values and show statistical significance in the lower regime, compared with the upper regime. Poverty may make farmers' decisions on crop diversification more sensitive in the lower regime.

In contrast, if household characteristics are related to market activities, they tend to affect crop diversification by upper-regime farmers more strongly than that by lower-regime farmers. If farmers' residences are far from agricultural markets, upper-regime farmers significantly increase the number of crops. Although upper-regime farmers relatively afford to purchase agricultural inputs for their cultivation, high transaction costs may induce farmers to substitute crop diversification for inputs. Upper-regime farmers also significantly increase crop numbers if they belong to a SACCO. Sufficient cash sources may enable upper-regime farmers to prepare the necessary inputs for planting new crops. In market-related characteristics, the availability of inputs may determine the adoption of crop diversification by upper-regime farmers.

Past rainfall amounts are positively associated with the increases in crop numbers and the SI in both the two regimes. Farmers may be able to adopt crop diversification more easily in agriculturally advantaged areas.

Appendix 3-V. Estimation Using Indicators of Crop Diversification Based on Five Crop Types

This appendix analyzes whether my main results are changed if the two indicators of crop diversification are constructed using five crop types (staples, sorghum/tubers/roots, legumes/oil, fruits/vegetables, and cash crops). The crop number is redefined so that it equals the number of cultivated crop types in each farmer. Similarly, the SI is redefined so that s_c in Equation (3.3.1) is the area share of each crop type in total farmland. Then, the two indicators are standardized in each region and year.

Table A1 shows estimation results of the Hansen's fixed-effects threshold model for each indicator. Covariates unrelated to past agricultural shocks are included in the analysis but omitted from the table. Estimated threshold parameters, 206,276 in numbers of crop types and 204,938 in the SI, are the same as those in my main analysis (see Table 3.2). F-statistics are significant in both the indicators.

Columns (1)-(2) show results concerning the number of cultivated crop types. Column (1) indicates that lower-regime farmers who experienced a large rise in food prices increase numbers of cultivated crop types by 0.36 times the SD of numbers of cultivated crop types. Column (2) indicates that upper-regime farmers who experienced a large rise in input prices increase numbers of crop types by 0.14 times the SD. Upper-regime farmers also increase numbers of crop types in response to crop diseases/pest outbreaks by 0.15 times the SD, which was unobserved in my main results. Because damages by crop diseases/pest outbreaks tend to spread widely over specific crops, even upper-regime farmers may have incentives to grow different crop types.

Column (3) shows that lower-regime farmers increase the SI by 0.26 and 0.35 times the SD of the SI in response to drought/flood and a rise in food prices, respectively. However, a past rise in input prices decreases the SI by 0.29 times the SD, which was not found in my main results. As Figure 3.9 shows, a relatively large proportion of lower-regime farmers purchase agricultural inputs. Some crops in vegetables/fruits and cash crops may require farmers to purchase seeds and seedlings. These crops may also require farmers to supply large amounts and specific types of nutrients by fertilizer, which cannot be covered by crop diversification. Because lower-regime farmers have insufficient cash income to purchase higher-priced inputs, they may be unable to maintain crop diversification by past rises in input price. In Column (4), I find no evidence that upper-regime farmers change the SI in response to past shocks.

The findings in Table A1 are similar to results in my main analysis of Table 3.4. However, in Table A1, I find no evidence that crop diversification is discouraged by past livestock loss among lower-regime farmers and that it is encouraged by a large fall in crop prices among upper-regime farmers.

Table A1: Estimates of Indicators of Crop Diversification Based on Five Crop Types

	Simple Counting		Simpson Index	
Explanatory Variable	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.28	0.01	0.26^{*}	-0.00
	(0.18)	(0.07)	(0.15)	(0.07)
Experienced crop disease/pests (dummy)	-0.09	0.15^{*}	-0.24	0.09
	(0.21)	(0.08)	(0.20)	(0.07)
Experienced large fall in sale prices for crops (dummy)	-0.06	0.11	-0.02	0.04
	(0.17)	(0.07)	(0.18)	(0.07)
Experienced large rise in price of food (dummy)	0.36^{**}	0.07	0.35^{**}	-0.01
	(0.17)	(0.06)	(0.16)	(0.06)
Experienced large rise in agricultural input prices (dummy)	-0.30	0.14^{*}	-0.29^{*}	0.08
	(0.19)	(0.08)	(0.16)	(0.07)
Experienced livestock death or theft (dummy)	0.08	0.01	-0.11	-0.02
	(0.25)	(0.09)	(0.23)	(0.09)
Estimated threshold	206,276		204,938	
Linearity test				
F statistic (p-value in parenthesis)	79.11^{*}		75.08^{*}	
	(0.06)		(0.08)	

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to agricultural shocks are omitted from the table.

Appendix 3-VI. Other Results Observed in Table 3.5

Table 3.5 provides some results unobserved in my main analysis. In Column (1), lower-regime farmers who experienced a fall in crop price decrease the probability of planting only staples. As illustrated in Figure 3.9, a relatively large proportion of lower-regime farmers also participate in output markets. Thus, lower-regime farmers may also have some incentives to stabilize crop sales.

In Column (2), upper-regime farmers who experienced a rise in food price also decrease the probability of the mono-cropping of staple crops, as lower-regime farmers do. However, the coefficient of the latter greatly outnumbers that of the former (-0.22 against -0.04).

In Column (3), lower-regime farmers are less likely to plant staples plus sorghum/roots/tubers in response to a rise in input prices. Sorghum/roots/tubers may be replaced by other climatically resistant and nutrient-complementary crops such as legumes.

In Column (8), upper-regime farmers increase the probability of cultivating staples plus vegetables/fruits in response to a rise in food prices. Because cultivation ways of vegetables/fruits are specific to each crop, upper-regime farmers may afford to cultivate it, relative to lower-regime farmers, when their food securities are threatened.

In Column (9), lower-regime farmers are more likely to cultivate staples plus cash crops if they experienced a rise in food prices. Lower-regime farmers may also rely on crop diversification to compensate for income loss caused by market shocks, as upper-regime farmers do. By cultivating cash crops, lower-regime farmers may ensure cash inflows to compensate for increased cash outflows owing to a rise in food price. Conversely, lower-regime farmers who experienced a rise in input price are less likely to plant cash crops. Some cash crops may require farmers to purchase seeds and seedlings or to supply large amounts of and specific types of nutrients by fertilizer. Lower-regime farmers with limited income may not afford to purchase higher-priced inputs to plant cash crops.

Table A2: Estimates of Indicators of Crop Diversification Based on Annual Crops

Explanatory Variable	Simple Counting		Simpson Index	
	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.02	-0.21**	0.38***	0.00
	(0.05)	(0.10)	(0.15)	(0.07)
Experienced crop disease/pests (dummy)	0.06	0.08	-0.20	0.06
	(0.07)	(0.14)	(0.17)	(0.08)
Experienced large fall in sale prices for crops (dummy)	0.03	0.08	0.07	0.05
	(0.06)	(0.12)	(0.18)	(0.08)
Experienced large rise in price of food (dummy)	0.05	-0.04	0.36**	0.02
	(0.05)	(0.11)	(0.16)	(0.06)
Experienced large rise in agricultural input prices (dummy)	0.06	-0.03	-0.18	0.08
	(0.06)	(0.14)	(0.16)	(0.07)
Experienced livestock death or theft (dummy)	-0.02	0.60***	-0.38*	0.04
	(0.07)	(0.21)	(0.21)	(0.09)
Estimated threshold	559,750		204,938	
Linearity test				
F statistic (p-value in parenthesis)	80.31*		76.70*	
	(0.06)		(0.07)	

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Covariates unrelated to agricultural shocks are omitted from the table.

Appendix 3-VII. Redefining indicators of crop diversification and crop types

In Appendix 3-VII, I reconstruct crop numbers and the SI by excluding fruit trees and permanent cash crops because the cultivation of permanent crops may be unsusceptible to shocks in the short run once farmers plant them.

Using the Hansen's fixed-effects threshold model, Table A2 displays estimation results of the simple count of crop numbers and the SI, which are calculated based on annual crops (pigeon pea is included in this calculation because of the importance as legumes). The estimated threshold in crop numbers, 559,750 Tsh, is even larger than that in Table 3.2 (206,276 Tsh). Thus, estimation results in Columns (1) and (2) are not similar to main results in Table 3.4. In contrast, the estimated threshold in the SI is the same as that in the main analysis (204,938). Hence, estimation results are similar to main results in Table 3.4. In Column (3), experiences of drought/flood and rises in food prices increase the SI and experiences of livestock loss decrease it among lower-regime farmers. These results coincide with Column (4) in Table 3.4.

We also separately estimate farmers' decisions to plant vegetables, fruits, annual cash crops, and permanent cash crops plus staples. In the crop-specific analyses, vegetables and fruits are classified in the same crop type as dietary sources of minerals and vitamins. Similarly, annual and permanent cash crops are also classified in the same type as cash income sources. However, farmers may plant annual and permanent crops differently in response to shocks. Moreover, the classification of pigeon pea is moved from cash crops to legumes. Although pigeon pea is classified as permanent cash crops in the NPS (see Appendix 3-III), it also has crop characteristics as legumes.

Using the Hansen's fixed-effects threshold model, Table A3 displays estimation results on the probability that farmers cultivate each of legumes, vegetables, fruits, annual cash crops, and permanent cash crops in addition to staples. Estimated thresholds in legumes, vegetables, and permanent cash crops are close to estimated thresholds in Table 3.2 (around 200,000 Tsh). Then, results are similar to main ones in these crop types. For example, Column (1) shows that lower-regime farmers plant legumes in addition to staples in response to a rise in food prices, whereas upper-regime farmers plant it in response to a fall in crop prices. Legumes still play important roles in crop diversification when pigeon peas are included in its category. Column (4) shows that upper-regime farmers plant vegetables plus staples in response to a rise in food price. Columns (9) shows that lower-regime farmers plant permanent cash crops in response to rises in food price and input price. These results are also observed in Table 3.5.

In contrast, estimated thresholds in fruits and annual cash crops are far above the estimated thresholds in Table 3.2. Thus, the results in Columns (5)-(8) do not coincide with those in Table 3.5.

Table A3: Estimates of Probability of Cultivating Each Crop Type on Shock Variables under Redefined Crop Types

Explanatory Variable	Staple & Legume		Staple & Vegetable		Staple & Fruit		Staple & Annual cash crop		Staple & Permanent cash crop	
	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime	(5) Lower Regime	(6) Upper Regime	(7) Lower Regime	(8) Upper Regime	(9) Lower Regime	(10) Upper Regime
Shock Variables:										
Experienced drought/flood (dummy)	0.02	-0.03*	0.01	0.02	-0.00	0.02	-0.00	0.01	-0.05	0.01
	(0.03)	(0.02)	(0.03)	(0.02)	(0.00)	(0.01)	(0.02)	(0.01)	(0.06)	(0.02)
Experienced crop disease/pests (dummy)	-0.01	-0.01	0.02	-0.00	0.01	0.00	0.00	0.03**	-0.07*	0.00
	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)
Experienced large fall in sale prices for crops (dummy)	-0.04	0.04*	-0.00	0.03	-0.01	0.02	-0.00	-0.00	-0.01	-0.00
	(0.03)	(0.02)	(0.04)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)	(0.07)	(0.02)
Experienced large rise in price of food (dummy)	0.05*	-0.02	0.03	0.03**	0.01	-0.02	-0.00	0.00	0.15***	-0.00
	(0.03)	(0.02)	(0.03)	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.05)	(0.02)
Experienced large rise in agricultural input prices (dummy)	-0.02	-0.01	0.03	0.02	-0.00	0.04*	-0.01	0.02	-0.10*	-0.02
	(0.03)	(0.02)	(0.03)	(0.02)	(0.00)	(0.02)	(0.01)	(0.02)	(0.05)	(0.02)
Experienced livestock death or theft (dummy)	-0.05	0.07**	-0.06	0.02	-0.02	-0.02	-0.01	0.00	0.05	0.01
	(0.03)	(0.03)	(0.04)	(0.03)	(0.01)	(0.03)	(0.02)	(0.03)	(0.07)	(0.02)
Estimated threshold	262,468		227,130		547,071		467,022		186,306	
Linearity test										
F statistic (p-value in parenthesis)	53.84		50.26		90.47		84.25*		45.21	
	(0.55)		(0.70)		(0.22)		(0.09)		(0.78)	

Note: $N = 2,664$, Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level. Other covariates are omitted from the table.

Appendix 3-VIII. Incorporating Crop Diversification in the Short Rainy Season

As explained in the dataset construction, my main analysis focuses on farmers' activities in the long rainy season (*masika*), the main cultivation season from March to May in Tanzania. However, the bimodal rainfall pattern is prevalent in some regions of Tanzania and offers the secondary cropping season, that is, the short rainy season (*vuli*) from October to December. Dividing the Tanzanian mainland into the five agro-ecological zones (Northeast, Coastal, Lakes, Central, and Southern Highlands), Figure A4 compares the number of farmers cultivating in both the long and short rainy seasons with that of farmers cultivating only in the long rainy season in my dataset.[1] In Figure A4, farmers' cultivation in the short rainy season is widely observed in Tanzania, though its numbers differ by regional rainfall patterns.

This appendix examines whether my results are changed if I incorporate the cultivation in the short rainy season. Indicators of crop diversification are newly constructed as aggregate information in the two rainy seasons. I calculate the crop number as the sum of crops cultivated in the two rainy seasons. In contrast, I calculate the SI by regarding the total farmland as the sum of land cultivated over the two rainy seasons. These two indicators are calculated based on the difference of crops, regardless of in which season they are cultivated. For example, if farmers cultivate maize in long rains and banana in short rains, they are regarded as different two crops. In contrast, even if farmers cultivate maize in both two rainy seasons, it is regarded as the same one crop. Finally, the two indicators are standardized in each region and year.

Table A4 shows estimation results of the Hansen's fixed-effects threshold model for each indicator. The set of covariates are the same as my main analysis, except that I add a dummy variable that is equal to one if farmers cultivated in the short rainy season. Estimated thresholds are 180,108 in the number of planted crops and 190,013 in the SI, both of which are similar to those in my main analysis. The linearity test significantly supports the threshold model in the number of planted crops, whereas it does not in the SI.

Using the threshold for crop numbers, Figure A5 illustrates the planting proportions and average land shares of each crop type among lower-/upper-regime farmers cultivating in the short rainy season (I do not display aggregated proportions and land shares over the two rainy seasons). Figure A5 shows that crop diversity is relatively similar in both regimes, overall. However, the planting proportions and land shares in the upper regime seem to be slightly higher in staples and legumes/oil and smaller in sorghum/tubers/roots than in the lower regime. The short rainy season tends to have lower and more uncertain rainfall than the long rainy season (Mutai and Ward 2000). Therefore, lower-regime farmers with low income

[1] Each agro-ecological zone contains the following administrative regions: Northeast for Arusha, Kilimanjaro, Manyara, and Tanga regions; Coastal for Dar es Salaam, Lindi, Morogoro, Mtwara, and Pwani regions; Central for Dodoma, Kigoma, Singida, and Tabora regions; Lakes for Kagera, Mara, Mwanza, and Shinyanga regions; Southern Highlands for Iringa, Mbeya, Rukwa, and Ruvuma regions.

Table A4: Estimates of Aggregate Indicators of Crop Diversification over the Two Rainy Seasons

	Crop Number		Simpson Index	
	(1)	(2)	(3)	(4)
Explanatory Variable	Lower Regime	Upper Regime	Lower Regime	Upper Regime
Household and regional characteristics:				
Age of household head	-0.04*	-0.04*	-0.03	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Years of education of household head	0.08**	0.03	0.09***	0.03
	(0.03)	(0.02)	(0.03)	(0.02)
Adult equivalents	-0.02	-0.04	-0.05	-0.03
	(0.04)	(0.03)	(0.03)	(0.02)
Acres of owned land	0.01	-0.00**	0.01	-0.00**
	(0.01)	(0.00)	(0.01)	(0.00)
Distance to nearest agricultural market (km) (log)	0.11	0.32**	-0.03	0.11
	(0.17)	(0.14)	(0.18)	(0.16)
Distance to district headquarter (km) (log)	0.06	0.01	0.03	-0.00
	(0.08)	(0.03)	(0.08)	(0.03)
Belongs to cooperative (dummy)	0.34*	0.06	0.19	0.08
	(0.18)	(0.07)	(0.21)	(0.07)
Belongs to SACCO (dummy)	1.08**	0.24**	-0.00	0.15
	(0.46)	(0.12)	(0.40)	(0.10)
Avg. *masika*-rainfall over past five years (mm) (log)	1.09**	1.06**	0.74	0.78
	(0.45)	(0.45)	(0.54)	(0.54)
In Wave 3 (2012/13) (dummy)	0.36**	0.06	0.38*	0.05
	(0.17)	(0.07)	(0.19)	(0.07)
Cultivated in short rains (dummy)	0.17	0.21***	0.10	0.23***
	(0.26)	(0.08)	(0.22)	(0.08)
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.59***	-0.09	0.54***	-0.10
	(0.17)	(0.07)	(0.16)	(0.07)
Experienced crop disease/pests (dummy)	-0.13	0.09	-0.15	0.04
	(0.22)	(0.07)	(0.18)	(0.07)
Experienced large fall in sale prices for crops (dummy)	0.04	0.15**	0.19	0.06
	(0.16)	(0.07)	(0.20)	(0.07)
Experienced large rise in price of food (dummy)	0.16	0.09*	0.18	0.09
	(0.16)	(0.06)	(0.20)	(0.06)
Experienced large rise in agricultural input prices (dummy)	-0.25	0.19***	-0.23	0.08
	(0.18)	(0.07)	(0.18)	(0.07)
Experienced livestock death or theft (dummy)	-0.71**	0.14	-0.63***	0.03
	(0.30)	(0.09)	(0.23)	(0.09)
Estimated threshold	180,108		190,013	
Linearity test				
F statistic (p-value in parenthesis)	91.25**		72.96	
	(0.01)		(0.16)	

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level.

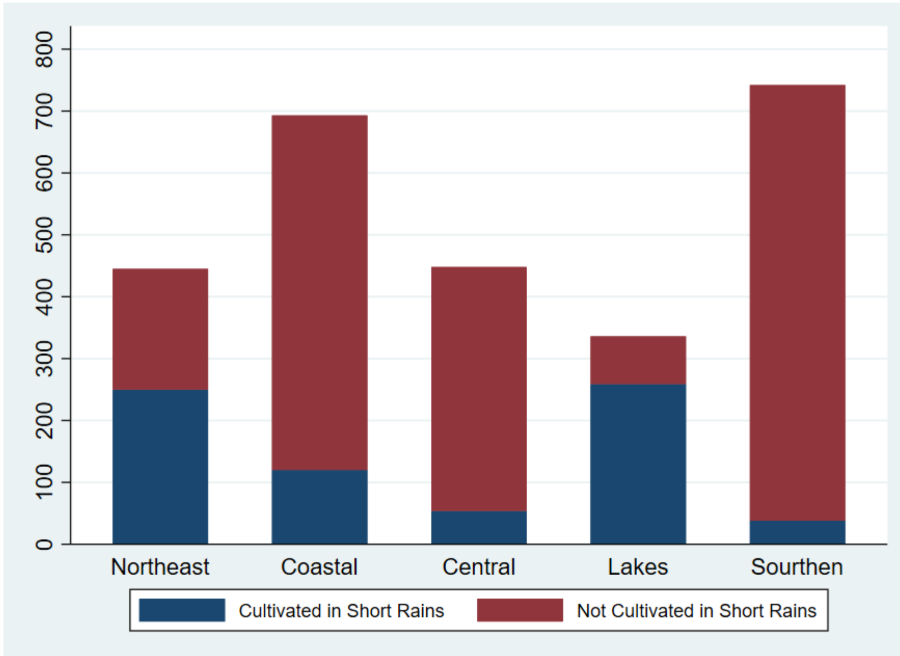


Figure A4: Number of Farmers who Cultivate in the Short Rainy Season or Not as well as the Long One

may have to cultivate low-valued but drought-resistant crops (sorghum/tubers/roots).

Columns (1) and (3) in Table A4 show that crop numbers and the SI significantly increase if lower-regime farmers experienced drought/flood, and conversely decrease if they experienced livestock loss. In Column (2), upper-regime farmers who experienced a large price fall in output markets and a large price rise in input markets significantly increase crop numbers. Therefore, I still obtain similar findings, even if I extend indicators of crop diversification over different rainy seasons within a year.

However, Column (2) indicates that the experience of a large fall in food prices significantly increases crop numbers among upper-regime farmers, not among lower-regime farmers, which is in conflict with results in Table 3.4. Interestingly, Columns (2) and (4) show that the cultivation in the short rainy season is significantly positively associated with crop diversity among upper-regime farmers. Upper-regime farmers may afford to pursue crop diversification in the short rainy season, and may cultivate different crops in short rains than those in long rains. Conversely, lower-regime farmers may prioritize producing certain drought-resistant crops under the volatile and unreliable rainfall in the short rainy season, as Figure A5 suggests.

In Table A5, aggregate indicators of crop diversification over the two rainy seasons are constructed using five crop types, as in Appendix 3-V. Estimated thresholds are 180,108 in both two indicators. The linearity test significantly supports the threshold model in the number of planted crop types, whereas it does not in the SI. Column (1) shows that the number of planted crop types significantly increases if lower-regime farmers experienced drought/flood

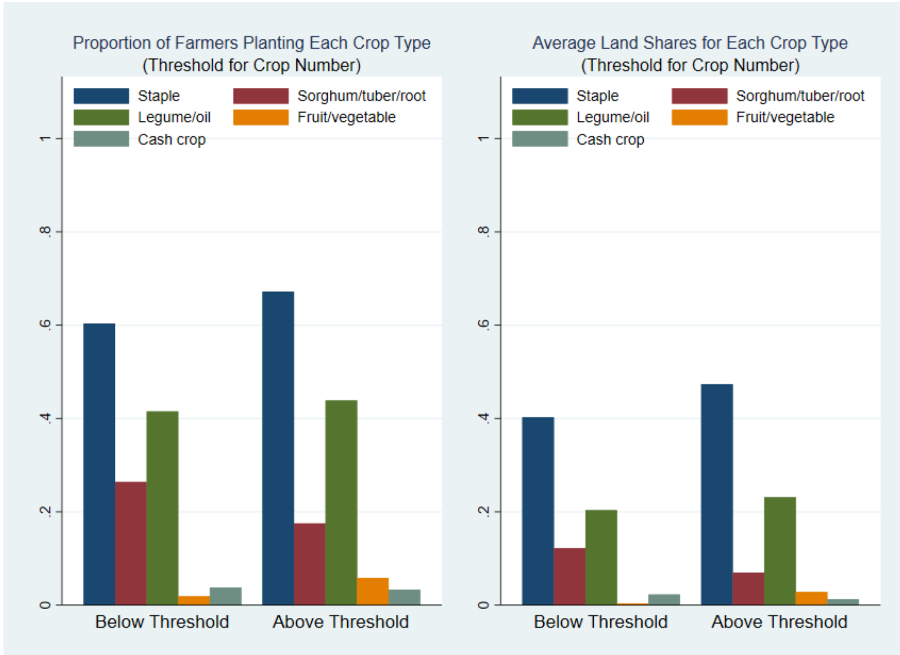


Figure A5: Crop Diversity in the Lower and Upper Regimes in the Short Rainy Season

and a rise in food prices. Conversely, the number of planted crop types significantly decreases if they experienced a rise in input prices and livestock loss. In Column (2), the experiences of crop diseases/pests, a fall in crop prices, a rise in food prices, and livestock loss increase the number of planted crop types among upper-regime farmers. Column (3) indicates that lower-regime farmers increase the SI in response to drought/flood and decrease it in response to crop diseases/pests, a fall in crop prices, a rise in input prices, and livestock loss. Many of these results were already observed in either Table 3.4 or Table A1.

Table A5: Estimates of Aggregate Indicators of Crop Diversification over the Two Rainy Seasons, Based on Five Crop Types

	Crop Type Number		Simpson Index	
Explanatory Variable	(1) Lower Regime	(2) Upper Regime	(3) Lower Regime	(4) Upper Regime
Household and regional characteristics:				
Age of household head	-0.03	-0.03	-0.02	-0.03
	(0.02)	(0.02)	(0.03)	(0.02)
Years of education of household head	0.07^{**}	0.02	0.07^{**}	0.03
	(0.03)	(0.02)	(0.03)	(0.02)
Adult equivalents	-0.01	-0.05^{**}	-0.00	-0.05^{**}
	(0.04)	(0.02)	(0.03)	(0.02)
Acres of owned land	-0.00	-0.00^{*}	0.00	-0.00^{**}
	(0.01)	(0.00)	(0.01)	(0.00)
Distance to nearest agricultural market (km) (log)	0.10	0.24	-0.12	0.01
	(0.19)	(0.16)	(0.18)	(0.15)
Distance to district headquarter (km) (log)	0.06	0.02	0.00	0.01
	(0.09)	(0.03)	(0.09)	(0.03)
Belongs to cooperative (dummy)	0.37^{*}	0.12	0.19	0.13^{*}
	(0.19)	(0.07)	(0.20)	(0.07)
Belongs to SACCO (dummy)	1.46^{**}	0.04	1.06	0.00
	(0.66)	(0.12)	(0.65)	(0.10)
Avg. *masika*-rainfall over past five years (mm) (log)	1.07^{**}	1.08^{**}	0.67	0.70
	(0.48)	(0.48)	(0.55)	(0.55)
In Wave 3 (2012/13) (dummy)	0.29	0.05	0.29	0.03
	(0.22)	(0.08)	(0.21)	(0.08)
Cultivated in short rains (dummy)	-0.16	0.26^{***}	-0.14	0.22^{***}
	(0.23)	(0.08)	(0.22)	(0.08)
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.55^{***}	-0.10	0.42^{**}	-0.10
	(0.20)	(0.06)	(0.18)	(0.06)
Experienced crop disease/pests (dummy)	-0.15	0.15^{*}	-0.36^{*}	0.04
	(0.24)	(0.07)	(0.20)	(0.07)
Experienced large fall in sale prices for crops (dummy)	0.10	0.12^{*}	0.08	0.06
	(0.20)	(0.07)	(0.22)	(0.07)
Experienced large rise in price of food (dummy)	0.34^{*}	0.10^{*}	0.23	0.05
	(0.19)	(0.06)	(0.21)	(0.06)
Experienced large rise in agricultural input prices (dummy)	-0.45^{**}	0.11	-0.58^{***}	0.05
	(0.22)	(0.07)	(0.20)	(0.07)
Experienced livestock death or theft (dummy)	-0.66^{**}	0.16^{*}	-0.65^{**}	0.00
	(0.30)	(0.09)	(0.28)	(0.08)
Estimated threshold	180,108		180,108	
Linearity test				
F statistic (p-value in parenthesis)	93.2^{***}		66.11	
	(0.00)		(0.2780)	

Note: Sample size: $N = 2,664$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level.

Appendix 3-IX. Dataset Expansion and Different Thresholds

Expanding the Range of the Dataset

My main analysis has two statistical limitations. First, the sample size is limited. I had to restrict the sample size to construct a balanced panel. Second, my estimation strategy may be susceptible to the endogeneity of threshold variables (food consumption levels). For example, if threshold variables are serially correlated, using a lagged variable may still be insufficient even in the fixed-effects approach. My results may depend on the Hansen's (1999) approach.

To check the robustness to these limitations, this appendix first extends the range of my dataset from Wave 2 (2010/11) and Wave 3 (2012/13) to all waves including Wave 1 (2008/09), Wave 4 (2014/15), and the recently uploaded wave, "National Panel Survey 2019/2020 - Extended Panel with Sex Disaggregated Data." Despite its name, the last wave comprises information on the 2018/19 agricultural season. In what follows, I call the last wave Wave 5 (2018/19) to avoid confusion.

Table A6 displays the number of households in each wave under the expanded dataset. To expand the dataset, I remove two restrictions imposed in my main analysis. First, the expanded dataset is an unbalanced panel and comprises different numbers of households, depending on waves. As noted in Section 3.3, Wave 4 (2014/15) was not used because the NPS refreshed the sampled households in this wave. This refreshed households comprise the majority in Wave 4; households who were interviewed in the past waves are the minority in Wave 4. In addition, Wave 5 (2018/19) surveyed only this minority households in Wave 4; it did not interview households in the refreshed majority in Wave 4. Consequently, the number of households drops dramatically in Waves 4 and 5 (additionally, Wave 5 has fewer households than Wave 4, because Wave 5 has many missing values in shock variables). Second, the expanded dataset includes households that split off from the original households, as long as they were interviewed at least in two waves after splitting off. My main analysis excluded split-off households because it is uncertain whether they inherited the unobserved characteristics of their original households. However, it is unproblematic to use information after splitting off from original households. By containing households interviewed at least twice after splitting off, the numbers of households in Waves 2 and 3 increase in the expanded dataset relative to the pre-expanded one.[2]

Because the number of households in the expanded dataset declines dramatically after Wave 4, the sample selection bias may arise in this unbalanced panel. To test for the sample

[2]My main analysis using the pre-expanded dataset removed households who split off between Waves 1 and 2 as well as those between Waves 2 and 3. This is because my main analysis relied on the consumption data in the preceding wave; I should not regard past consumption data of original households as that of split-off households.

Table A6: Number of Households in Each Wave under the Expanded Dataset

	Wave 1 (2008/09)	Wave 2 (2010/11)	Wave 3 (2012/13)	Wave 4 (2014/15)	Wave 5 (2018/19)	Total Households
Number of Households	1,356	1,620	1,608	366	166	5,116
(Proportion to Total)	(26.51)	(31.67)	(31.43)	(7.15)	(3.24)	(100.00)

selection bias, I add a lagged selection indicator to covariates, estimate the FE model that will appear later, and see the statistical significance for the indicator (see pp 832-833 in Wooldridge (2010) for details). Because I find no evidence of the sample selection bias, I proceed without correcting for it.

Using a Different Threshold

This appendix specifies a threshold in advance without estimating it, unlike threshold models in my main analysis. I define the threshold to be at the the bottom 40% of the sample distribution of asset holdings per adult equivalent, as in Makate et al. (2022). Food consumption levels are flow values determined by households' decisions in each period, whereas asset levels are stock values, an accumulation of all flow values in the past. Therefore, asset levels may be less serially correlated and less endogenous as wealth indicators than food consumption levels.[3]

The threshold wealth level adopted in Makate et al. (2022) is 40%, which differs from that in my main analysis, the bottom 16% of the sample distribution in food consumption levels (see Table 3.2). However, it is meaningful to examine how my results are changed or unchanged in comparison to Makate's et al. (2022) settings.

Asset Variable

The asset variable is defined as the sum of purchased values over all items in a household (e.g., televisions, motor vehicles, houses, and farm implements such as hoes, spraying machines, and tractors).[4] Waves 3-5 provide information on past prices at which households purchased

[3] I also attempt to estimate a switching regression model using Murtazashvili and Wooldridge (2016); the regime switching, which is observed or specified in advance, also depends on covariates. If the threshold by asset levels is used for defining the regime switching, I find no evidence that the switching is endogenous. Conversely, I find that the switching is endogenous if the threshold by the contemporary (not preceding) energy consumption (the poverty line of 2,300 kcal per adult per day) is used . This suggests that asset levels may be less endogenous than food consumption levels

[4] Land sizes and income are not used as wealth indicators for two reasons. First, these indicators may reflect farmers' engagements in non-farm activities, rather than wealth levels. Even if farmers are wealthy, their investments in land and agricultural assets seem low under low commitments to on-farm activities. In fact, land sizes differ insignificantly between poor and non-poor farmers, as was shown in the summary statistics. Second, data information provided in the NPS is often inconsistent among households. For example, households answered a questionnaire on their non-agricultural income, based on different time intervals (day/week/month/year).

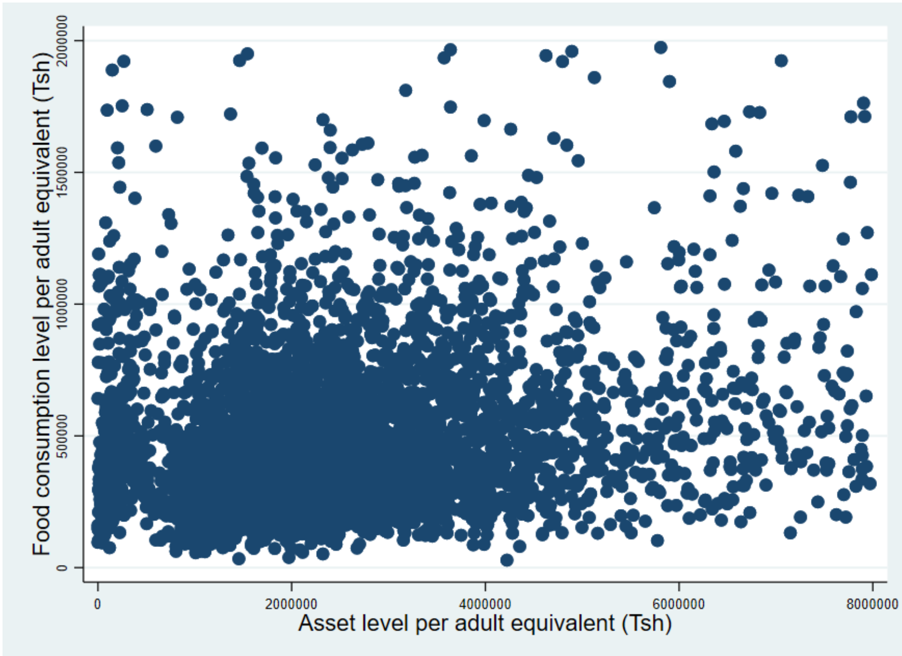


Figure A6: The Scatter plot between food consumption and asset levels per adult equivalent (Tsh)

each asset. Unfortunately, the data on purchasing prices of each asset do not exist in Waves 1 and 2; only the data on the number of asset holdings exist for each household. Thus, I first calculate predicted asset prices by taking averages of purchasing prices over households for each item in Waves 3-5. By multiplying predicted asset prices and the number of asset holdings together, I calculate the sum of predicted purchased values of assets in a household in each wave. Finally, I divide the sum of predicted purchased values of assets by adult equivalents. Because items owned by wealthy farmers may be of high quality and have high prices, purchased values of assets by poor households may be overestimated to some extent by using averaged asset prices.

Figure A6 displays a scatter plot between food consumption and asset levels per adult equivalent. I do not observe a clear positive relationship between them, though the correlation coefficient is 0.31 and positive at the 1% significance level.

Fixed-Effects Estimation Using Interaction Terms

This appendix estimates the following FE model:

$$SCD_{ikt} = \alpha_{ik} + x'_{it}\beta_{Lk} + s'_{it}\gamma_{Lk} + TA_{it} \times (x'_{it}\beta_{Uk} + s'_{it}\gamma_{Uk}) + \epsilon_{ikt} \tag{6.3.3}$$

for household i, year t, and indicator type k. The variables SCD_{ikt}, x_{it} and s_{it} represent standardized indicators of crop diversification, household characteristics, and shock variables, respectively. The dummy variable TA_{it} is equal to one if the asset level of household

i is above the bottom 40% of the sample distribution. By introducing interaction terms between covariates and the dummy variable, I represent a regime switching at the threshold. If household i belongs to the lower regime, then $TA_{it} = 0$ and Equation (6.3.3) becomes

$$SCD_{ikt} = \alpha_{ik} + x'_{it}\beta_{Lk} + s'_{it}\gamma_{Lk} + \epsilon_{ikt}. \tag{6.3.4}$$

In contrast, if household i belongs to the upper regime, then $TA_{it} = 1$ and Equation (6.3.3) becomes

$$SCD_{ikt} = \alpha_{ik} + x'_{it}(\beta_{Lk} + \beta_{Uk}) + s'_{it}(\gamma_{Lk} + \gamma_{Uk}) + \epsilon_{ikt}. \tag{6.3.5}$$

Therefore, the null hypothesis to be tested for each shock q is $\gamma_{Lkq} = 0$ for lower-regime farmers, whereas it is $\gamma_{Lkq} + \gamma_{Ukq} = 0$ for upper-regime farmers.

Covariates and Shock Variables

Some of the covariates used in my main analysis are removed owing to the lack of corresponding questionnaires after Wave 4 (the distances to the nearest agricultural market and the district head quarter and a dummy variable concerning a cooperative). Moreover, the information on the periods of shocks is not provided in the questionnaires after Wave 4, unlike Waves 1-3. For consistency over all waves, shock variables are defined, based on *roughly* five years from the interview (not from the cultivation) (the modification "roughly" is based on the fact that several households in Waves 1 to 3 answered shocks occurring more than five years before the interview).

Results

Table A7 shows estimation results of Equation (6.3.3). For brevity, Table A7 displays estimated coefficients for shock variables and their interaction terms with the asset dummy and omits other covariates. In Columns (1)-(2), indicators of crop diversification are calculated by differentiating individual crops regardless of crop types, whereas they are calculated based on the classification by five crop types in Columns (3)-(4).

Column (2) shows that lower-regime farmers tend to increase the SI if they experienced drought/flood, which coincides with my main results. Conversely, in Columns (3)-(4), lower-regime farmers are less likely to adopt crop diversification if they experienced livestock loss. Though food consumption levels are not clearly positively correlated with asset levels in Figure A6, I still observe similar results among lower-regime farmers. However, there are some different results than my main ones. In Columns (1)-(4), lower-regime farmers promote crop diversification in response to crop diseases/pests. Moreover, they also tend to decrease the SI in response to a large fall in crop prices in Columns (2) and (4).

Table A7: FE Estimates of Indicators of Crop Diversification

Explanatory Variable	Individual Crops (1) Crop Number	Individual Crops (2) Simpson Index	Five Crop Types (3) Type Number	Five Crop Types (4) Simpson Index
Agricultural shocks over past five years:				
Experienced drought/flood (dummy)	0.05	0.11^{**}	0.00	0.05
	(0.05)	(0.06)	(0.04)	(0.06)
Experienced crop disease/pests (dummy)	0.18^{***}	0.10^{*}	0.14^{***}	0.13^{**}
	(0.05)	(0.06)	(0.04)	(0.06)
Experienced large fall in sale prices for crops (dummy)	-0.05	-0.12^{**}	-0.04	-0.12^{**}
	(0.06)	(0.06)	(0.04)	(0.06)
Experienced large rise in price of food (dummy)	0.04	0.03	0.05	0.02
	(0.05)	(0.06)	(0.04)	(0.06)
Experienced large rise in agricultural input prices (dummy)	0.04	0.06	0.01	0.04
	(0.06)	(0.06)	(0.04)	(0.06)
Experienced livestock death or theft (dummy)	-0.09	-0.07	-0.10^{**}	-0.11^{*}
	(0.06)	(0.06)	(0.04)	(0.06)
Interaction terms between shock variables and the asset dummy:				
$TA_{ikt}\times$drought/flood (dummy)	-0.02	-0.08	0.02	-0.02
	(0.06)	(0.07)	(0.05)	(0.07)
$TA_{ikt}\times$crop disease/pests (dummy)	-0.16^{**}	-0.11	-0.12^{**}	-0.16^{**}
	(0.07)	(0.08)	(0.05)	(0.08)
$TA_{ikt}\times$large fall in sale prices for crops (dummy)	0.07	0.15^{*}	0.09^{*}	0.16^{**}
	(0.07)	(0.08)	(0.05)	(0.08)
$TA_{ikt}\times$large rise in price of food (dummy)	0.00	0.04	0.01	0.03
	(0.06)	(0.07)	(0.05)	(0.07)
$TA_{ikt}\times$large rise in agricultural input prices (dummy)	0.01	-0.06	-0.00	-0.03
	(0.07)	(0.08)	(0.06)	(0.08)
$TA_{ikt}\times$livestock death or theft (dummy)	0.11	0.10	0.10^{*}	0.16^{**}
	(0.07)	(0.08)	(0.05)	(0.08)

Note: Sample size: $N = 5,116$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01. Standard errors in parenthesis are clustered at the household level.

In contrast, many coefficients have opposite signs between shock variables and corresponding interaction terms. The coefficients of the interaction terms are significantly negative for crop diseases/pests in Columns (1), (3), and (4), positive for a fall in crop prices in Columns (2)-(4), and positive for livestock loss in Columns (3)-(4). Interestingly, the coefficient for the interaction term with a fall in crop prices outnumbers that for the corresponding shock variable in absolute values in all columns. Moreover, Columns (1) and (3) show that a fall in crop prices has no significant impact if not interacted, whereas it has significantly positive impacts if interacted.

Nevertheless, I do not find statistical significances for $\gamma_{Lkq}+\gamma_{Ukq}$ in most shocks. Table A8 displays values of $\gamma_{Lkq}+\gamma_{Ukq}$ as well as γ_{Lkq}, and tests the null hypotheses using t and F tests, with p-values in parenthesis. Because many coefficients in Table A7 have opposite signs between shock variables and corresponding interaction terms, $\gamma_{Lkq}+\gamma_{Ukq}$ tends to be close to zero. Though $\gamma_{Lkq}+\gamma_{Ukq}$ is consistently positive in all cases concerning a fall in crop prices, there is no significance. Exceptionally, I find a statistical significance for a rise in food prices in Column (6).

Some of the results in this appendix may differ from my main results, possibly because different wealth indicators are used for thresholds (food consumption versus assets), rather than because the threshold wealth levels are different (16% in food consumption levels versus 40% in asset levels). As Figure A6 suggests, food consumption levels and asset levels may exhibit different economic aspects and may not necessarily move together. Studies on poverty traps have shown that poor households attempt to maintain their asset levels by cutting food consumption levels (asset smoothing), unlike wealthy households who attempt to smooth consumption levels by selling assets (Zimmerman and Carter 2003; Carter and Lybbert 2012; Janzen and Carter 2019). If asset smoothing works among poor households, asset levels may not capture households' subsistence requirements adequately, compared with food consumption levels.

Table A8: Values of γ_{Lkq} and $\gamma_{Lkq} + \gamma_{Ukq}$ and Results of t test for $\gamma_{Lkq} = 0$ and F test for $\gamma_{Lkq} + \gamma_{Ukq} = 0$

	Individual Crops				Five Crop Types			
	Crop Type		Simpson Index		Type Number		Simpson Index	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explanatory Variable	γ_{Lk}	$\gamma_{Lk}+\gamma_{Uk}$	γ_{Lk}	$\gamma_{Lk}+\gamma_{Uk}$	γ_{Lk}	$\gamma_{Lk}+\gamma_{Uk}$	γ_{Lk}	$\gamma_{Lk}+\gamma_{Uk}$
Shock Variables:								
drought/flood	0.05	0.02	0.11**	0.03	0.00	0.03	0.05	0.03
	(0.36)	(0.55)	(0.05)	(0.54)	(0.92)	(0.41)	(0.39)	(0.49)
crop disease/pests	0.18***	0.01	0.10*	-0.01	0.14***	0.02	0.13**	-0.02
	(0.00)	(0.77)	(0.10)	(0.87)	(0.00)	(0.57)	(0.03)	(0.64)
large fall in sale prices for	-0.05	0.02	-0.12**	0.02	-0.04	0.05	-0.12**	0.05
crops	(0.33)	(0.64)	(0.04)	(0.66)	(0.29)	(0.15)	(0.05)	(0.35)
large rise in price of food	0.04	0.05	0.03	0.07	0.05	0.06*	0.02	0.06
	(0.42)	(0.24)	(0.55)	(0.11)	(0.17)	(0.05)	(0.69)	(0.21)
large rise in agricultural	0.04	0.05	0.06	0.01	0.01	0.01	0.04	0.00
input prices	(0.52)	(0.31)	(0.31)	(0.88)	(0.84)	(0.89)	(0.54)	(0.95)
livestock death or theft	-0.09	0.03	-0.07	0.03	-0.10**	-0.01	-0.11*	0.05
	(0.13)	(0.54)	(0.29)	(0.55)	(0.02)	(0.83)	(0.08)	(0.33)

Note: $N = 5,116$, Significance levels: * p<0.1; ** p<0.05; *** p<0.01 with p-values in parenthesis.

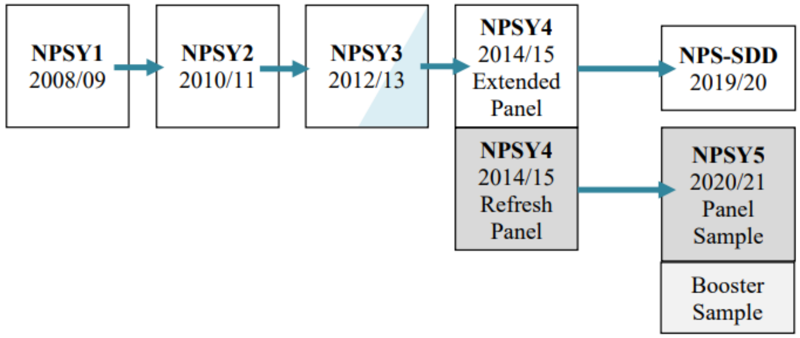


Figure A7: History of Data Collection over Waves in the NPS. This figure is cited from pp 4 in NBS (2022).

Appendices for Chapter 4

Appendix 4-I. Supplementary Information on the Short Rainy Season

Data Collection History over Waves in the NPS

NBS (2022) clearly depicts the history of the data collection over waves in the NPS in pp. 4. Figure A7 provides its picture. As I explain in the chapter, the NPS data basically comprise the same households from Wave 2008/09 to Wave 2012/13. However, only a part of households in Wave 2012/13 were interviewed as “Extended Panel” in Wave 2014/15. They were interviewed again in Wave 2018/19, which corresponds to “NPS-SDD 2019/20” in Figure A7 (my year notation, 2018/19, is adapted to the year of the long rainy season in this wave). In contrast, new households were added as “Refreshed Panel” in Wave 2014/15. They were interviewed again in Wave 2020/21, which corresponds to “NPS5 2020/21” in Figure A7.

Distributions of Harvest Months

Using the household-level dataset, Figure A8 displays the number of months in which farmers start and end harvesting annual crops in their plots after long and short rainy seasons.[5] In Figures A8(a) and A8(b), the distributions of months of starting and ending harvest in the long rainy season are unimodal and concentrate from May to August. In Figures A8(c) and A8(d), the distributions of months of starting and ending harvest in the short rainy season are unimodal and concentrate from January to Match. Figure A8 suggests that farmers tend to harvest crops in accordance with the timeline of Table 4.2.

[5] If months of starting and ending harvesting differ by multiple plots within a household, we count all these months in Figure A8.

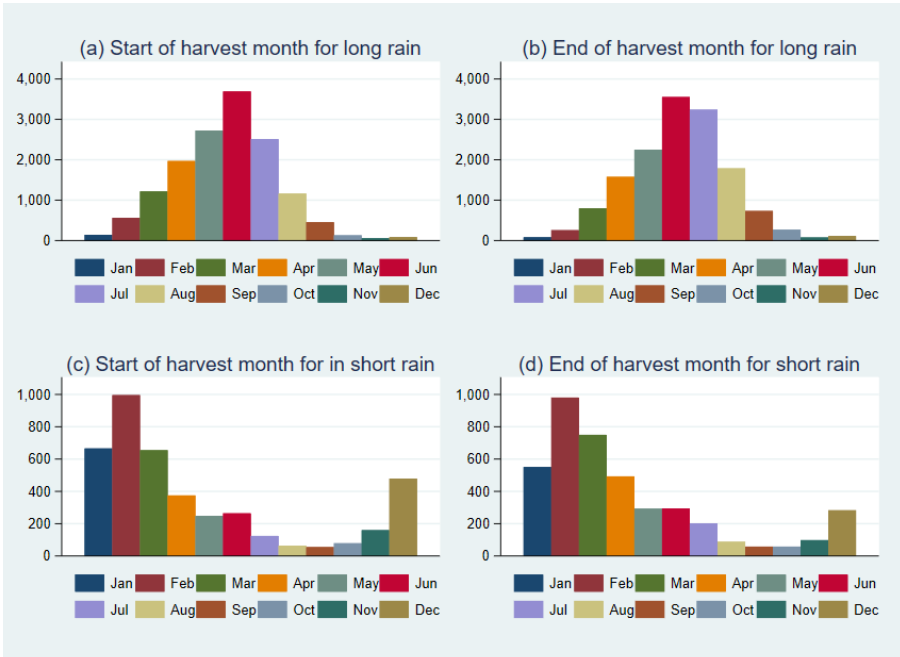


Figure A8: Distribution of First and Last Harvest Months in Long and Short Rainy Seasons

Rainfall Pattern in the Short Rainy Season

In the short rainy season, rainfall tends to be lower and more uncertain than in the long one (Mutai and Ward 2000). Figure A9 plots rainfalls in croplands of Kilimanjaro region from 1982 to 2021, with long rains (March-May) depicted by the blue-solid line and short rains (October-December) by the red-dotted line. In Figure A9, rainfalls in the short rainy season are obviously lower and seem to fluctuate more than those in the long one, overall. The coefficient of variation (CV), defined by the ratio of the standard deviation to the mean, is 45.35% for short rains, higher than 26.48% for long rains. Because of the interannual variability plus relatively low rainfall, farmers' decisions on whether to cultivate in the short rainy season may depend on water management facilities to efficiently use short rains (e.g., irrigation).

Price Seasonality in Tanzania

I examine the seasonality of food prices in Tanzania. Using open data from the Vulnerability Analysis and Mapping of the World Food Programme (WFP VAM), I obtain average wholesale prices of maize per kilogram in each region and month from January 2006 to May 2023.[6] I classify all months into four periods: January-March (the harvest season after short rains), April-June (the long rainy season), July-September (the harvest season after long rains), October-December (the short rainy season). I introduce four dummy variables that

[6]The data are available at `https://dataviz.vam.wfp.org`

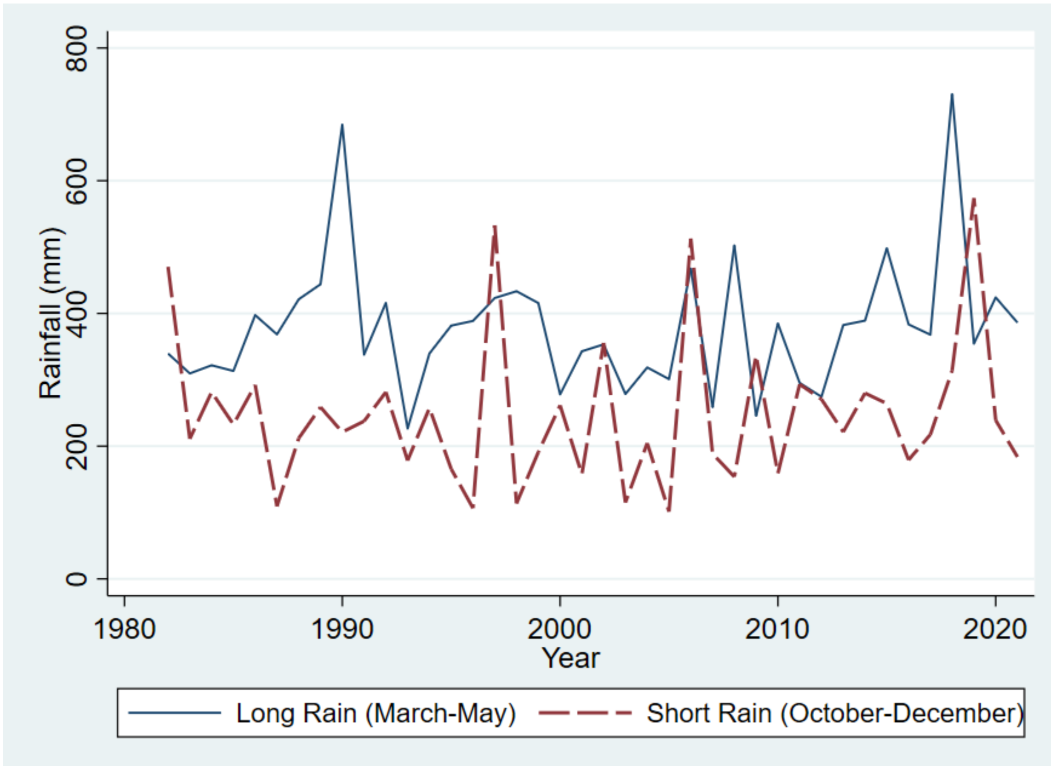


Figure A9: Rainfall during Long and Short Rainy Seasons in Kilimanjaro Region from 1982 to 2021

Table A9: OLS Estimates of Wholesale Maize Prices (Tsh) on Three Period Dummies

Explanatory Variable	(1) Northeast/Lakes	(2) Other Areas
= 1 if a month is in October-December	55.43***	56.11***
	(0.00)	(0.00)
= 1 if a month is in January-March	25.89*	62.44***
	(0.06)	(0.00)
= 1 if a month is in April-June	2.94	5.64
	(0.85)	(0.52)
Are year dummies included?	YES	YES
Are region dummies included?	YES	YES

Note: Significance levels: * p<0.1; ** p<0.05; *** p<0.01. with p-values in parenthesis. Standard errors are clustered at the household level.

Table A10: Number of Individuals Engaging in Each Type of Non-Farm Occupation and Amounts of Monthly Non-Farm Income

	Engage in fishing/ mining/tourism	Employed by government	Employed by parastatal	Employed by private sector	Employed by NGO/religious group	Operate own business
Number of individuals	188	511	99	6,340	176	3,170
(Proportion to total)	(0.02)	(0.05)	(0.01)	(0.60)	(0.02)	(0.30)
Avg. monthly income (Tsh)		513,136	247,536	108,830	217,034	381,661
(US dollar)		($257)	($124)	($54)	($109)	($191)

Note: Exchange rate: $1 = 2,000 Tsh.

are equal to one for each period. In this analysis, July-September is used as a baseline so that I can examine how maize prices change relative to those in the main harvest season in Tanzania. I also introduce year and region dummies.

In Table A9, I conduct the OLS regression of wholesale maize prices (Tanzanian shillings) on the three period dummies and region and year dummies, separately by two areas, Northeast/Lakes and other areas (see Figure 4.1). In Northeast/Lakes, the wholesale maize price tends to be the highest during the short rainy season (October-December), whereas the highest price in other areas comes before the start of the long rainy season (January-March). Importantly, the maize price in the harvest season of the short rainy season (January-March) tends to be even lower in Northeast/Lakes than in other areas by 36.55 Tsh ($= 62.44 - 25.89$), which suggests that the supply of crops harvested in the short rainy season decreases food prices in markets. This indicates the possibility that, if the cultivation in the short rainy season became more prevalent, even non-farmers would enjoy welfare improvements through lower food prices. I find no evidence of higher maize prices in April-June relative to July-September, though the coefficients are positive in the two areas.

Types and Amounts of Off-Farm Income

I construct an individual-level dataset that corresponds to the household-level dataset in Table 4.1, and count the number of individuals engaging in each type of non-farm occupations in Table A10 (proportions to total individuals in parenthesis). Table A10 classifies types of non-farm occupations as engagement in fishing/mining/tourism, employment in a central or local government, employment in a parastatal, political party, or cooperative, employment in the private sector, and operation of own business.

Table A10 shows that, out of individuals who are engaged in non-farm occupations, 60% are employed in the private sector. This is followed by the number of individuals who operate their own business (30%). Other types of employment and engagement in non-farm activities are limited.

The NPS provides the information on monthly income earned by non-farm employment and business.[7] Thus, Table A10 also displays average (Avg.) amounts of monthly non-farm

[7] In the NPS, individuals report non-farm income, based on different time intervals (e.g., daily, weekly, monthly, yearly). For consistency, I focus my attention to only monthly income. To avoid measurement

income per employed individual or business in Tanzanian shillings (US dollars in parenthesis), except for the engagement in fishing/mining/tourism. Table A10 shows that the average monthly income is the lowest in the private sector (108,830 Tsh) and is the highest in the government (513,136 Tsh). The average monthly income from operating business is 381,661 Tsh.

Importantly, all amounts of non-farm income per individual or business are lower than the average monthly food consumption in my dataset. The average value of real food consumption per adult equivalent per 28 days is 577,000 Tsh in our household dataset (this value is not displayed in Table A10). This value outnumbers the highest non-farm income in the government (513,136 Tsh). This suggests that non-farm income may not be adequate for many farmers to fully meet their food demands. Harvested crops during the short rainy season and their cash income from sales may enable farmers to satisfy their desire for free food consumption during the lean season.

Relationship between Basic Caloric Needs and Types of Food Consumed in a Household

I examine whether households tend to consume food more diversely once they meet the basic caloric needs. As in Dercon et al. (2009) and Michler and Josephson (2017), I adopt a poverty line of 2,300 kcal per adult per day as the basic caloric needs and introduce a dummy variable that is equal to one if the caloric intake per adult equivalent exceeds 2,300 kcal. Dependent variables are defined as dummy variables that are equal to one if a household consumes each food item in the HDDS (cereals, root/tubers, vegetables, fruits, meat/poultry/offal, eggs, fish/seafood, pulses/legumes/nuts, milk and milk products, oil/fats, sugar/honey, and miscellaneous).

Using the CRE probit model, Table A11 displays the estimated APE for the dummy variable of the poverty line in each food item (covariates used in my main analysis are also included in the analysis, although they are omitted in Table A11). Households are more likely to consume all food items once the basic caloric needs are met. However, the increase in the probability of consuming cereals is low relative to other food items than vegetables. This suggests that households who meet the basic caloric needs prefer to consume food more diversely, rather than increase cereals with high caloric contents.

errors, I do not convert non-farm income with other intervals into monthly income (e.g., multiplying daily income by 30).

Table A11: CRE Probit Estimates of APE for the Consumption of Each Food Item on Exceeding the Poverty Line

Explanatory Variable	(1) Cereals	(2) Root/tubers	(3) Vegetables	(4) Fruits
= 1 if caloric intake per adult equivalent exceeds 2,300 kcal	0.03*** (0.00)	0.08*** (0.00)	0.01** (0.01)	0.09*** (0.00)
Explanatory Variable	**(5) Meat/poultry/offal**	**(6) Eggs**	**(7) Fish/seafood**	**(8) Pulses/legumes/nuts**
= 1 if caloric intake per adult equivalent exceeds 2,300 kcal	0.12*** (0.00)	0.05*** (0.00)	0.08*** (0.00)	0.11*** (0.00)
Explanatory Variable	**(9) milk & milk products**	**(10) Oil/fats**	**(11) Sugar/honey**	**(12) Miscellaneous**
= 1 if caloric intake per adult equivalent exceeds 2,300 kcal	0.09*** (0.00)	0.06*** (0.00)	0.08*** (0.00)	0.09*** (0.00)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 with p-values in parenthesis. Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for areas, waves, and interview periods, time-averaged values.

Table A12: FE Estimates of Maize Harvests on its Planted Acre and the Squared Term

Explanatory Variable	(1) Maize Harvests in Long Rain	(2) Maize Harvests in Short Rain
Instrumental variables:		
Planted acre for maize in long rain	163.43^{***}	
	(0.00)	
Squared term of planted acre for maize in long rain	-1.89	
	(0.37)	
Planted acre for maize in short rain		184.08^{***}
		(0.00)
Squared term of planted acre for maize in short rain		-4.49^{**}
		(0.02)
F statistic for planted acre and its squared term (p-value in parenthesis)	12.06^{***}	6.60^{***}
	(0.00)	(0.00)

Note: Sample size: $N = 5,984$ in Column (1) and $N = 1,636$ in Column (2). Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for waves and interview periods.

Appendix 4-II. Supplementary Information in Methodology and Results

Inverse Relationship for Maize Harvests in Long and Short Rainy Seasons

In Table A12, I check whether the inverse relationship holds in each long and short rainy season in my dataset. Using the FE model, I regress total maize harvests in the long rainy season in Column (1), whereas I regress maize harvests in the short rainy season in Column (2). The covariates are the planted acre for maize for each rainy season, its squared term, and the same household characteristics used in our main analysis (however, I omit four area dummies owing to multicollinearity).

Column (1) shows that the coefficient for the planted acre for maize in the long rainy season is significantly positive. The coefficient for its squared term is negative, although it is not significant. The F statistic for the two variables is significant. Column (2) shows that the coefficient for the planted acre in the short rainy season is significantly positive, whereas the coefficient for its squared term is significantly negative. These two variables are also jointly significant. These results support the inverse relationship in our dataset.

First-Stage Results in FE-2SLS

In Table A13, I conduct the FE estimation of S_{it} and S_{it}^2 on four IVs as the first stage of the FE-2SLS model. Other covariates are omitted in the table, although they are included in the analysis. Standard errors are clustered at the household level.

In Table A13, three meteorological indices tend to be individually significant (Even if I find no statistical significance, p-values tend to be close to 0.10). The proportion of farmers planting annual crops in the short rainy season within a district is also individually significant. The four IVs are also jointly significant.

First-Stage Results in ESR

Table A14 displays estimation results in the first stage of the ESR model. I run the CRE probit estimation of $\textit{off-farm income}_{it}$ on IVs and exogenous explanatory variables. Other covariates than IVs are omitted in the table, although they are included in the analysis. In Table A14, meteorological indices are individually significant except for NINO3.4. The proportion of farmers having non-agricultural occupations within a district is also significant. Consequently, the five IVs are jointly significant.

Full Results of Table 4.4

Table A15 displays estimated coefficients for all covariates used in this chapter. However, time-averaged values in the ESR model are omitted in the table, although they are included in the analysis.

Table A13: FE Estimates of S_{it} and S_{it}^2 on IVs

Explanatory Variable	(1) S_{it}	(2) S_{it}^2
Instrumental variables:		
Sum of NINO3.4 in September-February of following year	-0.04^{*}	-1.25^{**}
	(0.10)	(0.04)
Sum of DMI in September-February of following year	0.96^{**}	17.65
	(0.05)	(0.10)
Interaction between NINO3.4 and DMI in September-February of following year	-0.00	-0.07
	(0.12)	(0.22)
Proportion of farmers planting annual crops in short rain within a district	3.03^{***}	31.25^{***}
	(0.00)	(0.00)
F statistic for four IVs (p-value in parenthesis)	16.04^{***}	3.53^{***}
	(0.00)	(0.01)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for areas, waves, and interview periods.

Table A14: CRE Probit Estimates of *off-farm income*$_{it}$ on IVs

Explanatory Variable	(1) *off-farm income*$_{it}$
Instrumental variables:	
Sum of NINO3.4 in September-February of following year	-0.02
	(0.13)
Sum of DMI in September-February of following year	-1.38^{***}
	(0.00)
Interaction between NINO3.4 and DMI in September-February of following year	0.01^{***}
	(0.00)
Proportion of farmers planting annual crops in short rain within a district	-0.11
	(0.50)
Proportion of farmers having non-agricultural occupations within a district	2.65^{***}
	(0.00)
F statistic for five IVs (p-value in parenthesis)	324.88^{***}
	(0.00)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for areas, waves, and interview periods, time-averaged values.

Table A15: Estimates of Welfare Impacts of the Cultivation in the Short Rainy Season

	Homogeneous effects without interactions				Heterogeneous effects with interactions			
	FE		FE-2SLS		FE		ESR	
Explanatory Variable	(1) Kcal	(2) HDDS	(3) Kcal	(4) HDDS	(5) Kcal	(6) HDDS	(7) Kcal	(8) HDDS
Variables non-interacted with *off-farm income$_{it}$*:								
Acres of cultivated land in short rain	28.57*	0.03*	1008.41*	0.89**	50.08*	0.02	1058.38*	0.25
	(0.09)	(0.07)	(0.06)	(0.03)	(0.06)	(0.49)	(0.10)	(0.68)
Squared term of acres of cultivated land in short rain	-0.69	-0.00	-97.57*	-0.07*	-1.81*	-0.00	-105.96*	-0.01
	(0.29)	(0.20)	(0.05)	(0.06)	(0.08)	(0.81)	(0.07)	(0.86)
Age of household head	1.53	-0.01**	2.60	-0.01	1.74	-0.01**	0.77	-0.03**
	(0.73)	(0.03)	(0.76)	(0.16)	(0.73)	(0.04)	(0.95)	(0.02)
Male-headed household (dummy)	-92.70	-0.01	-339.35	-0.19	-132.82	0.05	-554.62**	-0.23
	(0.36)	(0.94)	(0.18)	(0.35)	(0.26)	(0.74)	(0.04)	(0.44)
Years of education of household head	25.64**	0.04***	13.03	0.03	35.73***	0.03*	35.16	-0.00
	(0.01)	(0.00)	(0.60)	(0.12)	(0.01)	(0.08)	(0.45)	(0.99)
Years of living in community by household head	-1.41	-0.00	0.92	-0.00	-0.92	-0.00	1.33	0.00
	(0.28)	(0.14)	(0.73)	(0.70)	(0.64)	(0.27)	(0.75)	(0.67)
Adult equivalents	-225.51***	0.07***	-263.52***	0.05	-230.26***	0.08***	-218.25***	0.04
	(0.00)	(0.00)	(0.00)	(0.17)	(0.00)	(0.00)	(0.00)	(0.37)
Acres of owned land	14.62***	0.01***	129.24*	0.08*	14.83***	0.02**	66.59**	0.00
	(0.00)	(0.01)	(0.06)	(0.09)	(0.00)	(0.03)	(0.04)	(0.95)
Belongs to SACCO (dummy)	60.41	0.18*	639.76	0.58*	349.15*	0.27	-275.89	0.02
	(0.48)	(0.07)	(0.12)	(0.06)	(0.10)	(0.15)	(0.57)	(0.97)
Distance to nearest market from plots (km)	-1.08*	-0.00	-1.00	-0.00	-4.60**	0.00	-5.82*	-0.00*
	(0.07)	(0.37)	(0.29)	(0.88)	(0.01)	(0.95)	(0.07)	(0.09)
Rainfall during long rainy season (mm) (log)	-133.04*	-0.23**	103.49	-0.10	-69.42	-0.41**	517.01	-0.37
	(0.09)	(0.02)	(0.64)	(0.59)	(0.60)	(0.01)	(0.11)	(0.21)
Rainfall during short rainy season (mm) (log)	5.14	-0.04	100.73	0.05	114.85	-0.11	466.75	-0.11
	(0.95)	(0.74)	(0.58)	(0.76)	(0.43)	(0.57)	(0.15)	(0.74)
Sum of long rain over past 5 years (mm) (log)	-74.63	-0.59**	-97.60	-0.67**	-175.89	-0.44	655.50	-0.66
	(0.72)	(0.02)	(0.76)	(0.03)	(0.52)	(0.16)	(0.24)	(0.25)
Sum of short rain over past 5 years (mm) (log)	1.79	0.43*	-533.80	0.04	-39.77	0.42	-391.09	-0.18
	(0.99)	(0.08)	(0.23)	(0.91)	(0.87)	(0.16)	(0.43)	(0.73)
Resides in Northeast (dummy)	1016.92	2.13***	1514.28	2.37**	874.34	2.20**	-590.49*	0.43
	(0.27)	(0.01)	(0.14)	(0.01)	(0.34)	(0.01)	(0.10)	(0.17)
Resides in Coastal (dummy)	-841.39	0.89	-11.39	1.23	-697.16	0.93	-128.21	-0.66***
	(0.30)	(0.30)	(0.99)	(0.22)	(0.39)	(0.29)	(0.39)	(0.00)
Resides in Central (dummy)	-116.15	0.06	304.05	0.18	-134.89	-0.00	-13.84	-0.61***
	(0.88)	(0.95)	(0.71)	(0.83)	(0.86)	(1.00)	(0.92)	(0.00)
Resides in Lakes (dummy)	-233.89	0.66	517.07	0.90	-218.36	0.68	410.58**	-0.76***
	(0.79)	(0.33)	(0.60)	(0.26)	(0.80)	(0.34)	(0.02)	(0.00)
Interviewed from January to March (dummy)	-181.72*	-0.49***	-374.85*	-0.66***	-91.96	-0.40***	-783.53**	-0.40
	(0.08)	(0.00)	(0.09)	(0.00)	(0.47)	(0.01)	(0.01)	(0.17)
Interviewed from April to June (dummy)	16.86	-0.17*	87.63	-0.13	-128.85	-0.17	-354.12	-0.21
	(0.84)	(0.08)	(0.62)	(0.38)	(0.23)	(0.18)	(0.13)	(0.40)
Interviewed from October to December (dummy)	-98.78	-0.27**	-434.95	-0.52**	-51.01	-0.30*	-581.69**	-0.39
	(0.37)	(0.05)	(0.14)	(0.03)	(0.71)	(0.06)	(0.02)	(0.17)
In 2010/11 (dummy)	-494.45***	0.45***	-459.69***	0.48***	-647.03***	0.42***	-534.68***	0.46***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
In 2012/13 (dummy)	-572.10***	0.24***	-613.94***	0.23**	-736.05***	0.19*	-667.15***	0.38**
	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.06)	(0.00)	(0.01)
In 2014/15 (dummy)	-656.58***	0.43***	-955.95***	0.30	-812.02***	0.47***	-660.82***	0.86***
	(0.00)	(0.00)	(0.00)	(0.16)	(0.00)	(0.00)	(0.00)	(0.00)
In 2018/19 (dummy)	-653.30***	0.57***	-950.56***	0.49*	-688.88***	0.74***	-1540.63***	1.22***
	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)
In 2020/21 (dummy)	-677.68***	0.49***	-883.32***	0.48*	-823.02***	0.68***	-1076.48***	1.06***
	(0.00)	(0.00)	(0.00)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)
Any member has a non-agri. occupation (dummy)					-497.55	-0.28	-1245.89	-2.01
					(0.74)	(0.88)	(0.91)	(0.89)
Generalized residuals obtained in probit estimation							616.22*	-0.08
							(0.07)	(0.81)

(Continued.) Estimates of Welfare Impacts of the Cultivation in the Short Rainy Season

	Homogeneous effects without interactions				Heterogeneous effects with interactions			
	FE		FE-2SLS		FE		ESR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Explanatory Variable	Kcal	HDDS	Kcal	HDDS	Kcal	HDDS	Kcal	HDDS
Interaction terms with *off-farm income*$_{it}$:								
Acres of cultivated land in short rain					-32.98	0.01	-521.96	0.87
					(0.28)	(0.71)	(0.52)	(0.31)
Squared term of acres of cultivated land in short rain					1.56	-0.00	52.41	-0.08
					(0.19)	(0.63)	(0.47)	(0.25)
Age of household head					-1.37	0.00	1.21	0.02
					(0.64)	(0.54)	(0.93)	(0.21)
Male-headed household (dummy)					67.33	-0.11	490.83	0.13
					(0.44)	(0.32)	(0.14)	(0.76)
Years of education of household head					-14.61	0.01	-23.71	0.04
					(0.20)	(0.63)	(0.62)	(0.36)
Years of living in community by household head					-0.62	0.00	-1.92	-0.01
					(0.76)	(0.83)	(0.70)	(0.30)
Adult equivalents					15.20	-0.01	-29.41	0.02
					(0.43)	(0.49)	(0.68)	(0.81)
Acres of owned land					-0.38	-0.00	44.23	0.16*
					(0.94)	(0.75)	(0.57)	(0.08)
Belongs to SACCO (dummy)					-365.50	-0.12	751.57	1.01*
					(0.10)	(0.56)	(0.20)	(0.10)
Distance to nearest market from plots (km)					3.94**	-0.00	5.65*	0.01*
					(0.04)	(0.82)	(0.09)	(0.09)
Rainfall during long rainy season (mm) (log)					-117.36	0.26	-766.33**	0.21
					(0.45)	(0.17)	(0.03)	(0.55)
Rainfall during short rainy season (mm) (log)					-152.54	0.09	-547.59	0.21
					(0.37)	(0.68)	(0.11)	(0.59)
Sum of long rain over past 5 years (mm) (log)					106.56	-0.15	-1357.68**	-0.23
					(0.63)	(0.57)	(0.04)	(0.76)
Sum of short rain over past 5 years (mm) (log)					152.93	-0.06	17.40	0.24
					(0.51)	(0.85)	(0.98)	(0.79)
Resides in Northeast (dummy)					-190.46	-0.22	241.56	-0.54
					(0.27)	(0.30)	(0.57)	(0.25)
Resides in Coastal (dummy)					-292.29**	-0.16	-30.16	-0.19
					(0.04)	(0.32)	(0.89)	(0.42)
Resides in Central (dummy)					-331.55***	0.14	-203.53	-0.19
					(0.01)	(0.39)	(0.29)	(0.43)
Resides in Lakes (dummy)					-276.76**	-0.09	-62.94	0.15
					(0.03)	(0.57)	(0.76)	(0.57)
Interviewed from January to March (dummy)					-134.06	-0.13	659.46*	-0.28
					(0.20)	(0.31)	(0.07)	(0.48)
Interviewed from April to June (dummy)					245.00**	-0.02	640.85**	0.22
					(0.01)	(0.86)	(0.02)	(0.51)
Interviewed from October to December (dummy)					-62.15	0.06	378.74	-0.15
					(0.56)	(0.66)	(0.33)	(0.75)
In 2010/11 (dummy)					263.74**	-0.01	180.14	-0.01
					(0.01)	(0.95)	(0.30)	(0.97)
In 2012/13 (dummy)					273.69**	0.06	118.54	-0.33
					(0.02)	(0.61)	(0.51)	(0.14)
In 2014/15 (dummy)					262.01**	-0.08	-51.74	-0.67**
					(0.02)	(0.57)	(0.84)	(0.03)
In 2018/19 (dummy)					39.14	-0.24	1014.34**	-0.57
					(0.81)	(0.26)	(0.05)	(0.27)
In 2020/21 (dummy)					220.14*	-0.30*	660.01*	-0.52
					(0.10)	(0.08)	(0.07)	(0.15)
Generalized residuals obtained in probit estimation							-792.14*	0.24
							(0.06)	(0.60)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use standard errors clustered at the household level in FE, robust standard errors in FE-2SLS, and standard errors obtained by bootstrapping with 500 repetitions in ESR. Time-averaged values and their interaction terms are omitted in Columns (7)-(8), although they are included in the analysis.

Table A16: FE-2SLS Estimates of Crop Diversity and Pooled CF-CRE Probit Estimates of Planting Each Crop Types During the Short Rainy Season on Farmers' Holdings of Non-Agricultural Occupations

Explanatory Variable	(1) Number of Crop Types	(2) Staple Crops	(3) Sorghum/ Tuber/Root	(4) Legume/ Oil	(5) Vegetable/ Fruit	(6) Cash Crops
off-farm income$_{it}$	-0.26	0.07	-0.18*	-0.11	0.20***	-0.17**
	(0.67)	(0.31)	(0.07)	(0.30)	(0.00)	(0.01)

Note: Sample size: $N = 10,355$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Robust standard errors are used in Column (1). In Columns (2)-(6), standard errors are obtained by bootstrapping with 500 repetitions. APEs are displayed in Columns (2)-(6).
Other covariates omitted from the table: age, gender, educational years, and years of living in the community by household head, adult equivalents, acres of owned land, membership to SACCO, distance to nearest market, rainfall in cultivation seasons and over 5 years for each long and short rain, dummies for areas, waves, and interview periods (and time-averaged values in Columns (2)-(6)).

Appendix 4-III. Supplementary Information on Discussion

Crop Diversity

To examine whether farmers who have non-agricultural occupations diversify the composition of planted crops during the short rainy season, I classify planted crops into five crop types: staple crops (e.g., maize, paddy), sorghum/roots/tubes, legumes/oil, fruits/vegetables, and cash crops.[8] I use the sum of crop types planted during the short rainy season as an indicator of crop diversification. I restrict my dataset to farmers who plant annual crops during the short rainy season, because I aim to examine crop diversity in this season. Using the FE-2SLS model in Equation (4.4.1), I regress the sum of planted crop types on *off-farm income*$_{it}$, with the district-level proportion of *off-farm income*$_{it}$ used for its IV. I use the same covariates as my main analysis in Equation (4.4.1), except that I exclude S_{it} and S_{it}^2 from the covariates. I use robust standard errors. Column (1) in Table A16 displays its estimation results. I find no evidence that crop diversity during the short rainy season is associated with whether farmers have non-agricultural occupations.

Combining the pooled CRE probit model with the CF approach (Papke and Wooldridge 2008), I also examine the probability that farmers plant each crop type. Standard errors are obtained by bootstrapping with 500 repetitions. Table A16 displays APEs in estimation results. Farmers with non-agricultural occupations are less likely to plant sorghum/roots/tubers in Column (3) and cash crops in Column (6). Sorghum/tubers/roots (e.g., millet and potatoes) are low-valued but drought-resistant crops. Thus, farmers may have less incentives to plant sorghum/tubers/roots to reduce climatic risks if they have stable off-farm income. They may also not have to plant cash crops (e.g., cotton and tobacco) to each cash income if they can obtain cash income from non-farm activities. In contrast, farmers with non-agricultural occupations are more likely to plant vegetables/fruits in Column (5). They may be able to afford to plant these crops as dietary sources of minerals and vitamins.

[8] Permanent fruit trees and cash crops are included in these categories.

Appendices for Chapter 5

Appendix 5-I. Supplementary Information in Background Information and Conceptual Framework

Distribution of Students in Vocational and Teaching Schools

Table A17 displays the number of students attending the MS+ course, Certificate course, Diploma course, and non-formal education. Non-formal education, which is not depicted in the formal educational system in Figure 5.1, helps out-of-school children return back to the formal education. Table A17 shows that the numbers of students attending vocational and teaching schools are considerably low relative to pre-primary, primary, and lower-secondary schools.

Total Sales Revenues over the Two Rainy Seasons

Using the CRE Tobit model, Table A18 shows estimation results of total revenues of crop sales within a year.[9] Because sales revenues are large monetary values in Tanzanian shillings and the Tobit model is susceptible to the violation of the normality of the error term (Cameron and Trivedi 2010), I transform the dependent variable in logarithmic form after I add one to the original values. APEs are displayed in Table A18. I use the same covariates as my main analysis in Equation (5.4.1). Standard errors are clustered at the household level.

Column (1) of Table A18 shows that the double-cropping is associated with 87% more crop sales revenues within a year. In Column (2), I also add the interaction term $D_{it} \times \ln(Asset_{it} + 1)$, as in Equation (5.4.2). Although D_{it} is not individually significant, the estimated coefficient for D_{it} is positive and that for $D_{it} \times \ln(Asset_{it} + 1)$ is negative. Moreover, D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ are jointly significant. Farmers with high levels of asset values may gain income from non-farm activities rather than from on-farm activities.

Relationship between Farmers' Decisions to Sell Crops and the Schooling of Children

Using the pooled CRE probit model, Table A19 shows estimation results for the probability that farmers sell harvested crops in the long rainy season and that they sell crops in the short one. I focus on two explanatory variables: the sum of enrolling children and the sum of non-schooling children from pre-primary schools to lower-secondary schools. APEs are displayed in Table A19. I use the same covariates as my main analysis except for D_{it}, $D_{it} \times \ln(Asset_{it} + 1)$, and $\ln(Asset_{it} + 1)$ in Equation (5.4.2). Standard errors are clustered at the household level.

[9] Although a part of sales of fruits and permanent cash crops are not incorporated in Figure 5.5 owing to a lack of information on the period of the harvest, the total sales revenues in Table A18 include them. This is because we do not need to differentiate them by the period of the harvest.

Table A17: Number of Students in Each Educational Level

Age of Children	MS+ Course	Certificate Course	Diploma Course	Non-Formal Education
Official School Age of Pre-Primary School:				
5 (PP)				
6 (PP)				
Official School Age of Primary School:				
7 (D1)				
8 (D2)				2
9 (D3)				
10 (D4)				
11 (D5)				1
12 (D6)				2
13 (D7)	5			2
Official School Age of Lower-Secondary School:				
14 (F1)	1			
15 (F2)	6			
16 (F3)	5			1
17 (F4)	11	5	1	1
Official School Age of Upper-Secondary School:				
18 (F5)	4	18	3	
19 (F6)	2	22	5	
Official School Age of University:				
20 (U1)	2	44	5	
21 (U2)	1	26	8	
22 (U3)		22	11	
Total	37	137	33	9

Note: The table displays the number of students for each age of children attending school.

Table A18: Pooled CRE Tobit Estimates of APEs for Total Revenues of Crop Sales over the Two Rainy Seasons

Explanatory Variables	(1) Total Sales Revenue (No interaction term)	(2) Total Sales Revenue (Include interaction term)
D_{it} (double-cropping) (dummy)	0.87^{***}	2.26
	(0.00)	(0.19)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)		-0.09
		(0.41)
$\ln(Asset_{it}+1)$ (log of asset values)	0.20^{***}	0.23^{***}
	(0.00)	(0.00)
F statistic (p-values in parenthesis): $\beta_{1j} = \beta_{2j} = 0$		14.25^{***}
		(0.00)

Note: Sample size: $N = 11,031$. Significance levels: * p<0.1, ** p<0.05, *** p<0.01 in parenthesis. Standard errors are clustered at the household level.
I use the same covariates as Table A22, though they are omitted from the table.

Table A19: Pooled CRE Probit Estimates of APEs for the Probability of Selling Harvested Crops in Each Rainy Season

Explanatory Variables	(1) Prob. of Selling in Long Rain	(2) Prob. of Selling in Short Rain
The sum of children enrolling in pre-primary, primary, and lower-secondary schools	-0.00	0.01^{*}
	(0.88)	(0.05)
The sum of non-schooling children with official ages of pre-primary, primary, and lower-secondary schools	0.01	0.01^{**}
	(0.43)	(0.03)

Note: Sample size: $N = 11,031$. Significance levels: * p<0.1, ** p<0.05, *** p<0.01 in parenthesis.
Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, and educational years by household head, adult equivalents, numbers of children aged 5-17 and adults aged 18-60, acres of owned land, membership to SACCO, residence in urban areas, holding of non-agricultural occupations, rainfall in cultivation period and sum of rainfall over five years in each long and short rainy season, area and wave dummies, time-averaged values.

In Column (1), I find no evidence that the sum of enrolling children and that of non-schooling children are associated with the probability of selling harvested crops in the long rainy season. Conversely, these explanatory variables are significantly positive in Column (2), suggesting that farmers rely on additional agricultural income in the short rainy season to cover educational expenses.

Table A20: CRE Probit/Tobit Estimates of APEs for Children's Enrollment in and Continuation of School with the Covariate of $D_{it} \times Asset_{it}^2$

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.02	0.06***	-0.01	0.01	-0.00	-0.04	-0.01
	(0.15)	(0.00)	(0.49)	(0.66)	(0.75)	(0.29)	(0.69)
$D_{it} \times Asset_{it}$ (interaction term)	-0.00	-0.00**	0.00	-0.00	-0.00	0.00*	-0.00
	(0.41)	(0.01)	(0.31)	(0.95)	(0.53)	(0.07)	(0.80)
$D_{it} \times Asset_{it}^2$ (interaction term)	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.00
	(0.39)	(0.15)	(0.16)	(0.97)	(0.83)	(0.39)	(0.22)
$Asset_{it}$ (log of household asset values in Tsh)	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.72)	(0.54)	(0.48)	(0.12)	(0.88)	(0.22)	(0.28)
F statistic (p-values in parenthesis							
$\beta_{1j} = \beta_{2j} = \beta_{3j} = 0$ (coefficients for D_{it}, $D_{it} \times Asset_{it}$, and $D_{it} \times Asset_{it}^2$)	2.27	8.90**	1.12	0.46	0.94	1.78	0.44
	(0.32)	(0.01)	(0.57)	(0.80)	(0.39)	(0.17)	(0.65)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I obtain SE clustered at the household level. I use the same covariates as Table A22, though they are omitted from the table.

Appendix 5-II. Supplementary Information in Results and Discussion

Estimating Interaction Term between D_{it} and $Asset_{it}^2$

To check whether the heterogeneous impacts of the double-cropping through asset values are inverted U-shaped, I estimate the following equation, using the pooled CRE probit/Tobit models:

$$y_{ijt}^* = \beta_{1j}D_{it} + \beta_{2j}D_{it} \times Asset_{it} + \beta_{3j}D_{it} \times Asset_{it}^2 + \beta_{4j}Asset_{it} + \boldsymbol{\delta_j'x_{it}} + \mu_{ij} + \epsilon_{ijt}, \quad (6.3.6)$$

Table A20 displays estimation results. In Columns (1)-(4), I estimate the probability of children's enrollment in each school, whereas I estimate the number of children continuing to attend it in Columns (5)-(7). I find no individual significance for $D_{it} \times Asset_{it}^2$ in all columns. The F test shows that D_{it}, $D_{it} \times Asset_{it}$, and $D_{it} \times Asset_{it}^2$ are jointly significant in Column (2). However, the coefficient for $D_{it} \times Asset_{it}^2$ is not negative. Conversely, Column (2) shows that D_{it} and $D_{it} \times Asset_{it}$ are individually significant, which suggests that the joint significance is caused by D_{it} and $D_{it} \times Asset_{it}$, rather than $D_{it} \times Asset_{it}^2$. Thus, I find no evidence of the inverted U-shaped effects of the double-cropping over asset values.

First-Step Results in the CF Approach

In Table A21, I conduct the pooled CRE Probit estimation of D_{it} and the pooled CRE Tobit estimation of $D_{it} \times \ln(Asset_{it} + 1)$ on four IVs and other exogenous explanatory variables

Table A21: First-Stage Estimates of D_{it} and $D_{it} \times \ln(Asset_{it} + 1)$ on IVs

Explanatory Variable	(1) Probit D_{it}	(2) Tobit $D_{it} \times \ln(Asset_{it}+1)$
Instrumental variables:		
Sum of NINO3.4 in September-February of following year	0.01^{***}	0.11^{**}
	(0.00)	(0.04)
Sum of DMI in September-February of following year	0.18^{**}	1.88
	(0.05)	(0.17)
Interaction between NINO3.4 and DMI in September-February of following year	-0.00^{**}	-0.01
	(0.03)	(0.15)
Proportion of farmers cultivating in the two rainy seasons within a district	0.67^{***}	14.64^{***}
	(0.00)	(0.00)
F statistic for four IVs (p-value in parenthesis)	615.88^{***}	213.20^{***}
	(0.00)	(0.00)

Note: Sample size: $N = 11,031$. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. Standard errors are clustered at the household level.
Other covariates omitted from the table: age, gender, and educational years by household head, adult equivalents, numbers of children aged 5-17 and adults aged 18-60, acres of owned land, membership to SACCO, residence in urban areas, holding of non-agricultural occupations, rainfall in cultivation period and sum of rainfall over five years in each long and short rainy season, area and wave dummies, time-averaged values.

in the first stage of the CF approach. Other explanatory variables are omitted in the table, although they are included in the analysis. APEs are displayed in Column (1). Standard errors are clustered at the household level.

In Column (1), all IVs are individually significant. In Column (2), the NINO3.4 index and proportion of farmers cultivating in both rainy seasons are individually significant. The four IVs are jointly significant in both columns.

Full Estimation Results

Table A22 displays full results for Table 5.5(a). Table A23 displays full results for Table 5.6(a).

Frequency of Farmers' Double-Cropping

Table A24 displays the frequency of farmers' double-cropping in my dataset. I restrict my dataset to farmers who adopted the double-cropping at least one time in the past six waves from 2008/09 to 2020/21, so that I focus on farmers who can environmentally cultivate in the short rainy season. Because my panel dataset comprises households who were interviewed from two to five times, Table A24 separates the number of interviews from two to five in columns.

Table A24 shows that the largest frequency of the double-cropping is one, regardless of the number of interviews in the NPS. The frequency is dispersed relatively equally in the range of more than one.

Table A22: Full Results for Table 5.5(a)

	CRE model uncombined with CF				CRE model combined with CF			
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary	(8) Take PSLE
Double-cropping and assets:								
D_{it} (double-cropping) (dummy)	0.00	0.02	-0.00	0.01	-0.02	0.07**	-0.01	0.01
	(0.58)	(0.17)	(0.86)	(0.50)	(0.43)	(0.01)	(0.61)	(0.77)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00
	(0.34)	(0.87)	(0.52)	(0.88)	(0.46)	(0.96)	(0.53)	(0.89)
Household and climatic characteristics and wave dummies:								
Age of household head	-0.00	-0.00	-0.00*	-0.00***	-0.00	-0.00	-0.00	-0.00**
	(0.47)	(0.96)	(0.09)	(0.01)	(0.56)	(0.94)	(0.13)	(0.02)
Male-headed household (dummy)	0.01	-0.02	-0.02	0.02	0.01	-0.02	-0.02	0.02
	(0.75)	(0.40)	(0.23)	(0.32)	(0.75)	(0.39)	(0.20)	(0.41)
Years of education of household head	-0.00	-0.00	0.00	0.00	-0.00	-0.00	0.00	0.00
	(0.21)	(0.55)	(0.18)	(0.90)	(0.22)	(0.56)	(0.19)	(0.87)
Adult equivalents	-0.03***	-0.03***	0.03***	0.04***	-0.03***	-0.03***	0.03***	0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children aged 7-17	0.06***	0.09***	-0.00	-0.00	0.06***	0.08***	-0.00	-0.00
	(0.00)	(0.00)	(0.29)	(0.70)	(0.00)	(0.00)	(0.25)	(0.70)
Number of working adults	0.03***	0.04***	-0.03***	-0.04***	0.03***	0.04***	-0.03***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Acres of owned land	-0.00	-0.00	0.00	-0.00**	-0.00	-0.00	0.00	-0.00*
	(0.45)	(0.68)	(0.38)	(0.03)	(0.65)	(0.66)	(0.42)	(0.10)
Belongs to SACCO (dummy)	-0.01	0.03	0.00	0.01	-0.01	0.02	-0.00	0.01
	(0.43)	(0.27)	(1.00)	(0.67)	(0.44)	(0.33)	(0.97)	(0.74)
Resides in urban areas (dummy)	-0.04	-0.02	-0.01	0.00	-0.03*	-0.02	-0.01	0.00
	(0.10)	(0.39)	(0.66)	(0.97)	(0.09)	(0.37)	(0.78)	(0.90)
Has non-agricultural occupations (dummy)	0.01	0.00	-0.01*	0.00	0.01	0.00	-0.01	0.00
	(0.28)	(0.67)	(0.09)	(0.97)	(0.31)	(0.71)	(0.10)	(1.00)
Rainfall during long rainy season (mm) (log)	0.03**	-0.00	-0.00	-0.02	0.04**	-0.01	0.00	-0.01
	(0.04)	(0.91)	(0.83)	(0.45)	(0.04)	(0.60)	(0.76)	(0.69)
Rainfall during short rainy season (mm) (log)	0.02	-0.02	0.02	0.01	0.02	-0.02	0.01	0.01
	(0.21)	(0.32)	(0.24)	(0.60)	(0.23)	(0.40)	(0.30)	(0.67)
Sum of long rain over past 5 years (mm) (log)	-0.01	0.05	-0.01	0.01	-0.01	0.05	0.00	0.02
	(0.87)	(0.26)	(0.85)	(0.78)	(0.90)	(0.37)	(0.97)	(0.69)
Sum of short rain over past 5 years (mm) (log)	0.09**	0.03	0.01	0.03	0.08*	0.05	0.01	0.03
	(0.02)	(0.52)	(0.71)	(0.58)	(0.05)	(0.38)	(0.76)	(0.56)
In Wave 2010/11 (dummy)	0.03***	0.01	-0.00		0.03***	0.02	-0.00	
	(0.00)	(0.21)	(0.63)		(0.00)	(0.20)	(0.63)	
In Wave 2012/13 (dummy)	0.03***	0.00	-0.01	-0.02*	0.03***	0.01	-0.01	-0.02**
	(0.00)	(0.72)	(0.17)	(0.07)	(0.01)	(0.60)	(0.14)	(0.05)
In Wave 2014/15 (dummy)	0.04***	-0.01	-0.01*	-0.02*	0.04***	-0.03*	-0.02**	-0.03**
	(0.00)	(0.27)	(0.08)	(0.06)	(0.00)	(0.05)	(0.02)	(0.01)
In Wave 2018/19 (dummy)	0.01	-0.00	-0.00	-0.01	-0.00	0.01	-0.02	-0.03
	(0.52)	(0.86)	(1.00)	(0.51)	(0.98)	(0.60)	(0.22)	(0.21)
In Wave 2020/21 (dummy)	0.01	-0.04**	0.01	-0.01	0.01	-0.04**	0.00	-0.02*
	(0.48)	(0.02)	(0.15)	(0.21)	(0.62)	(0.04)	(0.66)	(0.06)
Resides in Northeast (dummy)	-0.02	0.04**	0.01	-0.01	-0.02	0.06***	0.01	-0.00
	(0.19)	(0.01)	(0.39)	(0.59)	(0.15)	(0.00)	(0.22)	(0.96)
Resides in Coastal (dummy)	-0.04***	-0.01	-0.01	-0.02	-0.04***	-0.01	-0.01	-0.01
	(0.00)	(0.28)	(0.12)	(0.12)	(0.00)	(0.48)	(0.25)	(0.27)
Resides in Central (dummy)	0.00	-0.01	-0.03***	-0.03***	0.00	0.00	-0.03***	-0.03**
	(0.72)	(0.53)	(0.00)	(0.00)	(0.88)	(0.90)	(0.00)	(0.01)
Resides in Lakes (dummy)	0.01*	0.03**	-0.02**	-0.02**	0.01	0.04***	-0.02**	-0.02
	(0.09)	(0.02)	(0.03)	(0.04)	(0.14)	(0.00)	(0.03)	(0.13)
Generalized residuals:								
Generalized residuals for D_{it}					0.02	-0.04**	0.01	0.01
					(0.17)	(0.02)	(0.45)	(0.71)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use SE clustered at the household level in Columns (1)-(4) and SE by bootstrapping with 200 repetitions in Columns (5)-(8). Time-average values are included in the estimation but omitted from the table.

Table A23: Full Results for Table 5.6(a)

	CRE model uncombined with CF			CRE model combined with CF		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Pre-Primary	(5) Primary	(6) Lower-Secondary
Double-cropping and assets:						
D_{it} (double-cropping) (dummy)	-0.01	0.01	-0.01	0.00	-0.00	-0.01
	(0.26)	(0.81)	(0.52)	(0.91)	(0.98)	(0.84)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.63)	(0.71)	(0.59)	(0.62)	(0.72)	(0.60)
Household and climatic characteristics:						
Age of household head	-0.00	-0.00	0.00	-0.00	-0.00	-0.00
	(0.59)	(0.84)	(0.82)	(0.66)	(0.87)	(0.94)
Male-headed household (dummy)	0.01	0.10*	-0.02	0.01	0.10*	-0.02
	(0.46)	(0.08)	(0.38)	(0.49)	(0.06)	(0.45)
Years of education of household head	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	(0.86)	(0.84)	(0.60)	(0.88)	(0.85)	(0.66)
Adult equivalents	-0.02***	0.06***	0.18***	-0.02***	0.06***	0.18***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of children aged 7-17	0.03***	0.37***	-0.11***	0.03***	0.37***	-0.11***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Number of working adults	0.02***	-0.08***	-0.19***	0.02***	-0.08***	-0.19***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Acres of owned land	-0.00	0.00	-0.00	-0.00	0.00	-0.00
	(0.86)	(0.43)	(0.51)	(0.85)	(0.48)	(0.41)
Belongs to SACCO (dummy)	0.03**	0.13**	0.00	0.03**	0.13**	0.00
	(0.03)	(0.01)	(0.93)	(0.05)	(0.01)	(0.91)
Resides in urban areas (dummy)	-0.01	0.09*	0.03	-0.01	0.09*	0.02
	(0.52)	(0.08)	(0.37)	(0.54)	(0.06)	(0.48)
Has non-agricultural occupations (dummy)	0.00	-0.00	0.01	0.00	-0.00	0.01
	(0.78)	(0.87)	(0.18)	(0.77)	(0.86)	(0.23)
Rainfall during long rainy season (mm) (log)	0.00	0.10**	-0.01	0.00	0.11**	-0.02
	(0.99)	(0.02)	(0.79)	(0.97)	(0.02)	(0.52)
Rainfall during short rainy season (mm) (log)	-0.01	-0.01	0.02	-0.01	-0.01	0.02
	(0.23)	(0.89)	(0.48)	(0.26)	(0.86)	(0.47)
Sum of long rain over past 5 years (mm) (log)	0.02	-0.27***	-0.05	0.02	-0.27***	-0.06
	(0.52)	(0.00)	(0.35)	(0.50)	(0.01)	(0.32)
Sum of short rain over past 5 years (mm) (log)	0.01	-0.10	-0.06	0.01	-0.10	-0.06
	(0.78)	(0.33)	(0.26)	(0.68)	(0.29)	(0.31)
In Wave 2010/11 (dummy)	-0.00	-0.10***	0.03***	-0.00	-0.10***	0.03**
	(0.62)	(0.00)	(0.01)	(0.54)	(0.00)	(0.01)
In Wave 2012/13 (dummy)	0.00	-0.16***	0.00	0.00	-0.16***	0.01
	(0.72)	(0.00)	(0.86)	(0.80)	(0.00)	(0.69)
In Wave 2014/15 (dummy)	0.01	-0.22***	-0.01	0.01	-0.23***	0.00
	(0.13)	(0.00)	(0.59)	(0.29)	(0.00)	(0.79)
In Wave 2018/19 (dummy)	0.01	-0.04	0.03	0.01	-0.04	0.06**
	(0.45)	(0.36)	(0.18)	(0.63)	(0.37)	(0.03)
In Wave 2020/21 (dummy)	0.01	-0.05*	0.05***	0.01	-0.07*	0.06***
	(0.22)	(0.10)	(0.00)	(0.20)	(0.05)	(0.00)
Resides in Northeast (dummy)	-0.00	-0.11***	-0.00	-0.00	-0.09**	-0.00
	(0.85)	(0.00)	(0.93)	(0.94)	(0.02)	(0.85)
Resides in Coastal (dummy)	-0.01	-0.06**	-0.04***	-0.01	-0.05*	-0.04***
	(0.45)	(0.01)	(0.01)	(0.42)	(0.06)	(0.01)
Resides in Central (dummy)	-0.00	-0.20***	-0.08***	-0.00	-0.20***	-0.08***
	(0.81)	(0.00)	(0.00)	(0.86)	(0.00)	(0.00)
Resides in Lakes (dummy)	0.00	-0.11***	-0.09***	0.01	-0.10***	-0.09***
	(0.50)	(0.00)	(0.00)	(0.31)	(0.00)	(0.00)
Generalized residuals:						
Generalized residuals for D_{it}				-0.00	0.02	0.01
				(0.70)	(0.49)	(0.64)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use SE clustered at the household level in Columns (1)-(3) and SE by bootstrapping with 200 repetitions in Columns (4)-(6). Time-average values are included in the estimation but omitted from the table.

Table A24: Frequency of Farmers' Double-Cropping

	Number of interviews in the NPS			
Frequency of double-cropping	two	three	four	five
one	1,602	864	100	325
two	712	552	40	115
three		789	44	100
four			16	120
five				100

Table A25: CF-CRE Probit/Tobit Estimates of APEs for Children's Enrollment in and Continuation of School under Restriction to Households with Children Aged 5-17

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.09	0.24	0.20**	0.15	-0.05	-0.04	-0.01
	(0.42)	(0.13)	(0.02)	(0.25)	(0.46)	(0.92)	(0.97)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.01	-0.01	-0.02**	-0.01	0.00	-0.01	0.00
	(0.30)	(0.51)	(0.02)	(0.29)	(0.68)	(0.73)	(0.80)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	-0.00	0.01	0.01**	0.01	-0.00	0.00	-0.00
	(0.86)	(0.41)	(0.05)	(0.53)	(0.67)	(0.89)	(0.71)
F statistic (p-values in parenthesis							
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it})+1$)	1.54	9.45***	5.50*	1.32	1.39	2.84	0.57
	(0.46)	(0.01)	(0.06)	(0.52)	(0.50)	(0.24)	(0.75)

Note: Sample size: $N = 7,813$ in pre-primary, primary, and lower-secondary schools; $N = 6,434$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I obtain SE by bootstrapping with 200 repetitions in all columns. I use the same covariates as Table A22, though they are omitted from the table.

Histogram of Household Asset Values

Figure A10 illustrates the histogram of household asset values per adult equivalent in Tanzanian shillings, with their quantiles depicted by red-dotted lines. Because some farmers have few assets, the asset distribution has two modes, although the distribution is right-skewed as a whole.

Robustness Check: Restriction to Households with Children Aged 5-17

For robustness check, I restrict the sample to households with children at official ages from pre-primary to lower-secondary schools (5-17). Table A25 shows estimation results when I use the CF approach and include the interaction term $D_{it} \times \ln(Asset_{it}+1)$. Results on children's enrollment are displayed in Columns (1)-(4), whereas results on children's school continuation are displayed in Columns (5)-(7). In Column (2), D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are jointly significant. In Column (3), D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are individually and jointly significant. In contrast, I find no statistical significance in the continuation of school in Columns (5)-(7). These results coincide with my main results.

For further robustness check, I create three household groups who have children at offi-

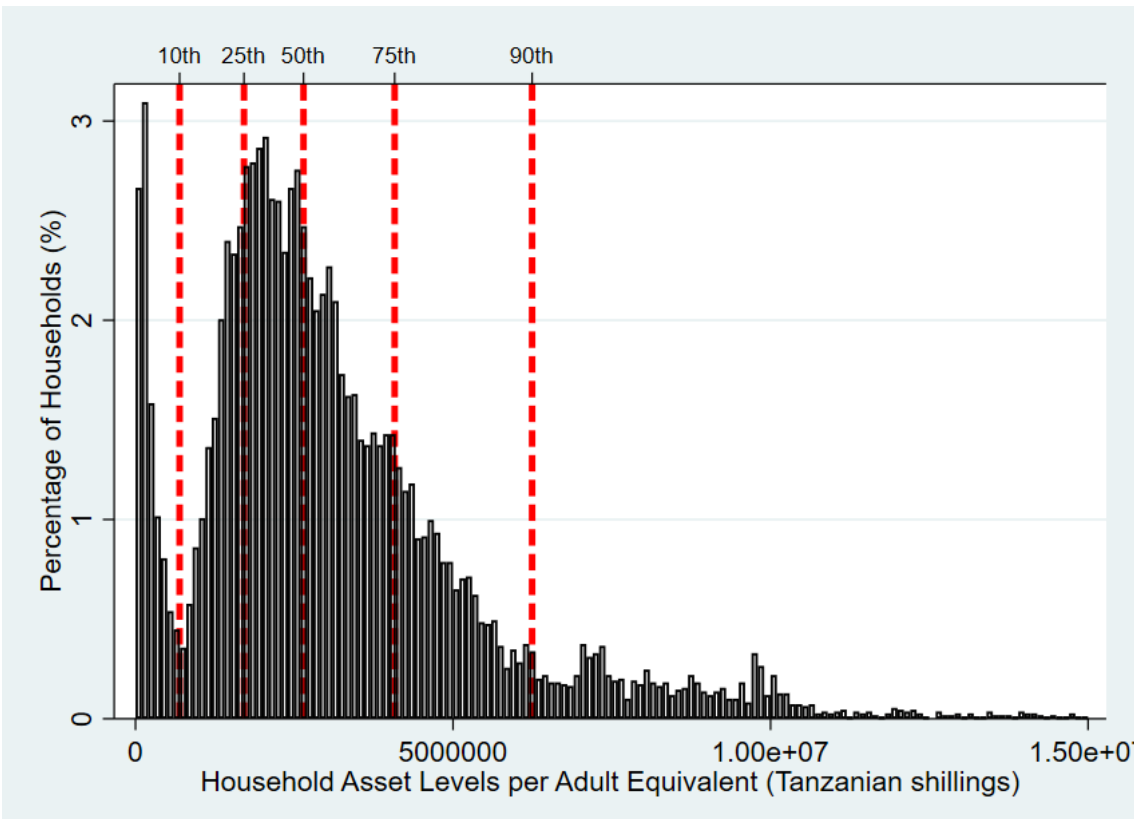


Figure A10: Histogram of Household Asset Values per Adult Equivalent (Red-Dotted Lines: 10th, 25th, 50th, 75th, 90th quantiles)

Table A26: CF-CRE Probit/Tobit Estimates of APEs for Children's Enrollment in and Continuation of School under Restriction to Three Household Groups with Children Aged 5-6, 7-13, and 14-17.

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.09	0.22	0.29	0.00	0.13	-0.06	-0.03
	(0.79)	(0.29)	(0.15)	(0.99)	(0.42)	(0.88)	(0.93)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.01	-0.00	-0.02	-0.00	-0.01	-0.01	0.00
	(0.71)	(0.91)	(0.12)	(0.89)	(0.38)	(0.72)	(0.87)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	0.01	0.01	0.02*	0.01	0.00	0.01	-0.01
	(0.65)	(0.31)	(0.07)	(0.74)	(0.63)	(0.56)	(0.57)
F statistic (p-values in parenthesis							
$\beta_{1j} = \beta_{2j} = 0$ (coefficients for D_{it} and $D_{it} \times \ln(Asset_{it}) + 1$)	0.24	10.88***	2.48	0.31	0.77	3.43	0.12
	(0.89)	(0.00)	(0.29)	(0.86)	(0.68)	(0.18)	(0.94)

Note: Sample size: $N = 1,988$ in pre-primary schools; $N = 5,890$ in primary schools; $N = 3,502$ in lower-secondary schools; $N = 2,628$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. We obtain SE by bootstrapping with 200 repetitions in all columns. We use the same covariates as Table A22, though they are omitted from the table.

cial ages of pre-primary schools (5-6), primary schools (7-13), and lower-secondary schools (14-17), respectively. Then, I estimate impacts of the double-cropping on children's school enrollment and continuation within each corresponding household group, because the school enrollment and continuation tend to concentrate on official age groups, as Table 5.1 shows. For the analysis on the PSLE, I use the same household group as lower-secondary schools, although households in Wave 2008/09 are excluded in the PSLE. Table A26 shows estimation results when I use the CF approach and include the interaction term $D_{it} \times \ln(Asset_{it}+1)$. In Column (2), D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are jointly significant. Although I find no significance in Column (3), the coefficients for D_{it} and $D_{it} \times \ln(Asset_{it}+1)$ are positive and negative, respectively. I find no statistical significance in the continuation of school in Columns (5)-(7). Although the limited sample size may affect the statistical significance here, I obtain similar results.

Robustness Check: Using Acres of Planted Land as Covariates

To incorporate acres of planted land during the short rainy season in covariates, I estimate the following equation:

$$y^*_{ijt} = \gamma_{1j} S_{it} + \gamma_{2j} S^2_{it} + \gamma_{3j} S_{it} \times \ln(Asset_{it}+1) + \gamma_{4j} S^2_{it} \times \ln(Asset_{it}+1) \\ + \gamma_{5j} \ln(Asset_{it}+1) + \boldsymbol{\delta'_j x_{it}} + \mu_{ij} + \epsilon_{ijt},$$

where S_{it} is acres of land planted for annual crops during the short rainy season. I focus on annual crops because the NPS provides information on planted acres for annual crops. I further add to covariates its squared term S^2_{it} to incorporate the inverse relationship between

Table A27: CRE Probit/Tobit Estimates of APEs for Children's Enrollment in and Continuation of School on Acre of Planted Land, its Squared Term, and their Interaction Terms with Log of Asset Values

	Enrollment in school				Continuation of school		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
S_{it} (acres of planted land for annual crops in short rain)	0.04	-0.03	0.05*	0.11**	0.01	-0.07	0.02
	(0.42)	(0.60)	(0.10)	(0.03)	(0.71)	(0.51)	(0.74)
S_{it}^2 (squared term of S_{it})	-0.00	0.00	-0.00***	-0.01***	-0.00	-0.00	-0.00
	(0.52)	(0.94)	(0.00)	(0.00)	(0.77)	(0.90)	(0.41)
$S_{it} \times \log(Asset_{it} + 1)$	-0.00	0.00	-0.00*	-0.01**	-0.00	0.00	-0.00
	(0.44)	(0.52)	(0.09)	(0.04)	(0.70)	(0.53)	(0.71)
$S_{it}^2 \times \log(Asset_{it} + 1)$	0.00	-0.00	0.00***	0.00***	0.00	0.00	0.00
	(0.57)	(0.83)	(0.00)	(0.00)	(0.79)	(0.89)	(0.42)
$\log(Asset_{it} + 1)$	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00
	(0.44)	(0.98)	(0.37)	(0.92)	(0.64)	(0.65)	(0.61)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. We use SE clustered at the household level in Columns (1)-(4) and SE by bootstrapping with 200 repetitions in Columns (5)-(8). We use the same covariates as Table A22, though they are omitted from the table.

farm productivity and farm size, as Chapter 4 did. The marginal increase of harvests per land may decrease as planted acres increase, which may lead to diminishing returns of children's education to planted land.

Using CRE probit/Tobit models, Table A27 displays estimates of APEs for children's enrollment in and continuation of school in the equation above. In Table A27, I do not combine CRE models with the CF approach because I find no individual significance for all IVs when S_{it}^2 and $S_{it}^2 \times \log(Asset_{it} + 1)$ are dependent variables in the first stage of the CF approach. Although estimation methods are not rigorous in Table A27, I still see its results to check the robustness of our main results.

Column (3) shows that one acre of planted land during the short rainy season is positively associated with the probability to enroll in lower-secondary schools by 5 pp. This increase in the enrollment probability is weakened as acres of planted land increase, because S_{it}^2 is significantly negative, which is consistent with the inverse relationship between farm productivity and farm size. Furthermore, the impacts of S_{it} decrease as households' asset levels rise, because the interaction term between S_{it} and $\log(Asset_{it} + 1)$ is significantly negative, which coincides with observations in our main analysis. In contrast, because $S_{it}^2 \times \log(Asset_{it} + 1)$ is significantly positive, the influence of the inverse relationship is weakened as asset levels increase.

The same argument applies in Column (4). Unlike our main results, I find statistical significances in the probability for children to take the PSLE. Thus, estimation results may somewhat depend on a form of estimated models. However, it seems that farmers' cultivation in the short rainy season promotes children's school enrollment, regardless of the difference in estimated models.

Table A28: CRE Probit/Tobit Estimates of APEs for Children's Enrollment/Attendance in Schools on Educational Reform in 2016

	Enrollment in School				Number of Non-Schooling Children		
	(1) Pre-Primary	(2) Primary	(3) Lower-Secondary	(4) Take PSLE	(5) Pre-Primary	(6) Primary	(7) Lower-Secondary
Double-cropping and assets:							
D_{it} (double-cropping) (dummy)	0.19**	0.28**	-0.03	-0.03	0.02	0.09	-0.20
	(0.03)	(0.03)	(0.74)	(0.81)	(0.86)	(0.58)	(0.14)
$D_{it} \times \ln(Asset_{it}+1)$ (interaction term)	-0.01**	-0.02**	0.00	0.00	-0.00	-0.01	0.01
	(0.04)	(0.04)	(0.74)	(0.77)	(0.84)	(0.51)	(0.15)
$D_{it} \times Reform_{it}$ (dummy)	-0.28	0.32	0.03	0.19	-0.87*	-0.31	0.95*
	(0.26)	(0.48)	(0.87)	(0.43)	(0.08)	(0.53)	(0.06)
$D_{it} \times \ln(Asset_{it} + 1) \times Reform_{it}$	0.02	-0.02	-0.00	-0.01	0.06*	0.02	-0.07**
	(0.26)	(0.46)	(0.89)	(0.44)	(0.08)	(0.48)	(0.04)
$\ln(Asset_{it}+1)$ (log of household asset values in Tsh)	0.00	0.01	0.00	-0.00	0.00	0.00	-0.00
	(0.95)	(0.20)	(0.76)	(0.83)	(0.62)	(0.81)	(0.38)

Note: Sample size: $N = 11,031$ in pre-primary, primary, and lower-secondary schools; $N = 9,178$ in PSLE. Significance levels: * p<0.1; ** p<0.05; *** p<0.01 in parenthesis. I use SE clustered at the household level. I use the same covariates as Table A22, though they are omitted from the table.

Impacts of Educational Reform

I examine how the impacts of the double-cropping were changed by the fee-free education. I introduce a dummy variable $Reform_{it}$, which is equal to one for farmers in the wave 2018/19 or 2020/21. I add to covariates in Equation (5.4.2) two interaction terms, $D_{it} \times Reform_{it}$ and $D_{it} \times \ln(Asset_{it}) \times Reform_{it}$, and estimate their APEs. Because it is considerably difficult to satisfy assumptions of IV for as many as four endogenous variables, I do not combine the CRE probit and Tobit models with the CF approach.[10] However, the CRE model may indicate useful information on the impacts of the educational reform.

Instead of using the number of children continuing schools, I construct the number of non-schooling children who are at official ages of pre-primary schools (5-6), primary schools (7-13), and lower-secondary schools (18-19). As mentioned in "Gender-Differential Impacts" in Discussion, my variable on the continuation of pre-primary schools may reflect the enrollment in primary schools.

Table A28 displays estimation results of children' enrollment in Columns (1)-(4) and the number of non-schooling children in Columns (5)-(7). In Column (5), the estimated APE for $D_{it} \times Reform_{it}$ is significantly negative, whereas that for $D_{it} \times \ln(Asset_{it}) \times Reform_{it}$ is significantly positive. Thus, the double-cropping after the educational reform enables farmers to make their children attend pre-primary schools. Moreover, this effect is stronger for asset-poor farmers than for asset-rich farmers. Although the underlying mechanism is unclear, the educational reform seems to contribute to early education of children through enhancing the potential role of the short rainy season in Tanzania.

[10]My IVs are not jointly significant for $D_{it} \times Reform_{it}$ and $D_{it} \times \ln(Asset_{it}) \times Reform_{it}$ in the first step of the CF approach.

References

- Aggarwal, S., Francis, E., and Robinson, J. (2018). Grain today, gain tomorrow: evidence from a storage experiment with savings clubs in Kenya. *Journal of Development Economics*, 134, 1-15.
- Alvi, E., and Dendir, S. (2011). Weathering the storms: credit receipt and child labor in the aftermath of the great floods (1998) in Bangladesh. *World Development*, 39(8), 1398–1409.
- Amin S., Quayes, S., and Rives, J.M. (2006). Market work and household work as deterrents to schooling in Bangladesh. *World Development*, 34(7), 1271-1286.
- Antony, J., and Klarl, T. (2022). Poverty and sustainable development around the world during transition periods. *Energy Economics*, 110, 106016.
- Arslan, A., Cavatassi, R., Alfani, F., Mccarthy, N., Lipper, L., and Kokwe, M. (2018). Diversification under climate variability as part of a CSA strategy in rural Zambia. *The Journal of Development Studies*, 54(3), 457-480.
- Asfaw, S., Pallante, G., and Palma A. (2018). Diversification strategies and adaptation deficit: evidence from rural communities in Niger. *World Development*, 101, 219-234.
- Asfaw, S., Scognamillo, A., Caprera, G.D., Sitko, N., and Ignaciuk, A. (2019). Heterogeneous impact of livelihood diversification on household welfare: cross-country evidence from Sub-Saharan Africa. *World Development*, 117, 278-295.
- Baird, S., McIntosh, C., and Özler, B. (2019). When the money runs out: do cash transfers have sustained effects on human capital accumulation? *Journal of Development Economics*, 140, 169-185.
- Baland, J., and Robinson, J. (2000). Is child labor inefficient? *Journal of Political Economy*, 108(4), 663–679.
- Bandara, A., Dehejia, R., and Lavie-Rouse, S. (2015). The impact of income and non-income shocks on child labor: evidence from a panel survey of Tanzania. *World Development*, 67, 218-237.

- Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E. (2019). Medium- and long-term educational consequences of alternative conditional cash transfer designs: experimental evidence from Colombia. *American Economic Journal: Applied Economics*, 11(3), 54–91.
- Basu, K., and Van, P. (1998). The economics of child labor. *American Economic Review*, 88(3), 412–427.
- Beegle, K., Dehejia, R.H., and Gatti, R. (2006). Child labor and agricultural shocks. *Journal of Development Economics*, 81(1), 80-96.
- Beegle, K., Galasso, E., and Goldberg, J. (2017). Direct and indirect effects of Malawi's public works program on food security. *Journal of Development Economics*, 128, 1–23.
- Bellon, M. R., Kotu, B.H., Azzarri, C., and 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.
- Benedetti, F., Ibarrarán, P., McEwan, P.J. (2016). Do education and health conditions matter in a large cash transfer? Evidence from a Honduran experiment. *Economic Development and Cultural Change*, 64(4), 759-793.
- Benhassine, N., Florencia, D., Esther, D., Pascaline, D., and Victor, P. (2015). Turning a shove into a nudge? a "labeled cash transfer" for education. *American Economic Journal: Economic Policy*, 7(3), 86-125.
- Berhane G., Gilligan D.O., Hoddinott J., Kumar N., and Taesse 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.
- Bevis, L.E. M., Naschold, F., and Rao, T. (2019). An unequal burden: intra-household dimensions of seasonal health in Tanzania. *Food Policy*, 89, 101766.
- Bevis, L.E., and Barrett, C.B. (2020). Close to the edge: high productivity at plot peripheries and the inverse size-productivity relationship. *Journal of Development Economics*, 143, 102377.
- Bhargava, A.K., Vagen, T., and Gassner, A. (2018). Breaking ground: unearthing the potential of high-resolution, remote-sensing soil data in understanding agricultural profits and technology use in Sub-Saharan Africa. *World Development*, 105, 352-366.
- Birthal, P.S., Joshi, P.K., Roy, D., and Thorat, A. (2013). Diversification in Indian agriculture toward high-value crops: the role of small farmers. *Canadian Journal of Agricultural Economics*, 61, 61-91.

- Birthal, P.S., Roy, D., and Negi, D.S. (2015). Assessing the impact of crop diversification on farm poverty in India. *World Development*, 72, 70-92.
- Björkman-Nyqvist, M. (2013). Income shocks and gender gaps in education: evidence from Uganda. *Journal of Development Economics*, 105, 237–253.
- Bozzola, M. and Smale, M. (2020). The welfare effects of crop biodiversity as an adaptation to climate shocks in Kenya. *World Development*, 135, 105065.
- Brander, M., Bernauer, T., and Huss, M. (2021). Improved on-farm storage reduces seasonal food insecurity of smallholder farmer households: evidence from a randomized control trial in Tanzania. *Food Policy*, 98, 101891.
- Burke, M., Bergquist, L.F., and Miguel, E. (2019). Sell low and buy high: arbitrage and local price effects in Kenyan markets. *Quarterly Journal of Economics*, 134(2), 785-842.
- Cameron, A.C. & Trivedi, P.K. (2010). *Microeconometrics using Stata*, revised edition, Stata Press.
- Carter, M. R., and Lybbert, T.J. (2012). Consumption versus asset smoothing: testing the implications of poverty trap theory in Burkina Faso. *Journal of Development Economics*, 99, 255-264.
- Chamberlain, G. (1984) Panel data., Chapter 22 in *Handbook of Econometrics*, Volume 2, edited by Griliches, Z. and Intriligator, M.D. Amsterdam: North Holland Press.
- Channa, H., Ricker-Gilbert, J., Feleke, S., and Abdoulaye, T. (2022). Overcoming smallholder farmers' post-harvest constraints through harvest loans and storage technology: insights from a randomized controlled trial in Tanzania. *Journal of Development Economics*, 157, 102851.
- Chegere, M.J., Eggert, H., and Soderbom, M. (2022). The effects of storage technology and training on postharvest losses, practices, and sales: evidence from small-scale farms in Tanzania. *Economic Development and Cultural Change*, 70(2), 729-61.
- Cogneau, D., and Jedwab, R. (2012). Commodity price shocks and child outcomes: the 1990 cocoa crisis in Côte d'Ivoire. *Economic Development and Cultural Change*, 60(3), 507-534.
- Coles, J.L., and Hammond, P.J. (1995). Walrasian Equilibrium without Survival: Existence, Efficiency, and Remedial Policy, Chapter 3 in *Development, Welfare and Ethics: A Festschrift in Honour of Amartya Sen*, edited by Basu, K., Pattanaik, P.K., and Suzumura K. Oxford University Press.

- Coromaldi, M., Pallante G., and Savastano, S. (2015). Adoption of modern varieties, farmers' welfare and crop biodiversity: evidence from Uganda. *Ecological Economics*, 119, 346-358.

- Cunha, J.M. (2014). Testing paternalism: cash vs. in-kind transfers in rural Mexico. *American Economic Journal: Applied Economics*, 6(2): 195-230.

- Dammert, A.C. (2008). Child labor and schooling response to changes in coca production in rural Peru. *Journal of Development Economics*, 86, 164–180.

- de Janvry A., Finan F., Sadoulet E., and Vakis R. (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics*, 79, 349–373.

- de Walque, D., and Valente, C. (2023). Incentivizing school attendance in the presence of parent-child information frictions. *American Economic Journal: Economic Policy*, 15(3), 256–285.

- Delavallade, D., and Godlonton, S. (2023). Locking crops to unlock investment: Experimental evidence on warrantage in Burkina Faso. *Journal of Development Economics*, 160, 102959.

- Dercon, S., amd Krishnan, P. (2000). Vulnerability, seasonality and poverty in Ethiopia. *Journal of Development Studies*, 36(6), 25-53.

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

- Dercon, S., Gilligan, D.O., Hoddinott, J., and Woldehanna, T. (2009). The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages. *American Journal of Agricultural Economics*, 91(4), 1007-1021.

- Di Falco, S. and Perrings, C. (2005). Crop biodiversity, risk management and the implications of agricultural assistance. *Ecological Economics*, 55(4), 459-466.

- Di Falco, S., and Chavas, J.P. (2008). Rainfall shocks, resilience, and the effects of crop biodiversity on agroecosystem productivity. *Land Economics*, 84(1), 83-96.

- Di Falco, S. and Chavas, J.P. (2009). On crop biodiversity, risk exposure, and food security in the highlands of Ethiopia. *American Journal of Agricultural Economics*, 91(3), 599-611.

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

- Dillon, A. (2013). Child labor and schooling responses to production and health shocks in Northern Mali. *Journal of African Economies*, 22(2), 276–299.
- Dillon, A., McGee, K., and Gbemisola, O. (2015). Agricultural production, dietary diversity and climate variability. *Journal of Development Studies*, 51(8), 976-995.
- Dillon B. (2021). Selling crops early to pay for school: a large-scale natural experiment in Malawi. *Journal of Human Resources*, 56(4), 1296-325.
- Dimitrova, A. (2021). Seasonal droughts and the risk of childhood undernutrition in Ethiopia. *World Development*, 141, 105417.
- Dolislager, M., Liverpool-Tasie, L.S.O., Mason, N.M., Reardon, T., and Tschirley, D. (2022). Consumption of healthy and unhealthy foods by the African poor: Evidence from Nigeria, Tanzania, and Uganda. *Agricultural Economics*, 53(6), 847-984.
- Duelli, P. and Obrist, M.K. (2003). Biodiversity indicators: the choice of values and measures. *Agriculture, Ecosystems and Environment*, 98, 87-98.
- Dunne, M., and Humphreys, S. (2022). The edu-workscape: re-conceptualizing the relationship between work and education in rural children's lives in Sub-Saharan Africa. *World Development Perspectives*, 27, 100443.
- Duryea, S., Lam, D., and Levison, D. (2007). Effects of economic shocks on children's employment and schooling in Brazil. *Journal of Development Economics*, 84(1), 188–214.
- Ecker, O., and Qaim, M. (2011). Analyzing nutritional impacts of policies: an empirical study for Malawi. *World Development*, 39(3), 412-428.
- Ecker, O. (2018). Agricultural transformation and food and nutrition security in Ghana: does farm production diversity (still) matter for household dietary diversity? *Food Policy*, 79, 271-282.
- Edmonds, E.V. (2006). Child labor and schooling responses to anticipated income in South Africa. *Journal of Development Economics*, 81, 386–414.
- Emerson, P.M., Ponczek, V., and Souza. A.P. (2017). Child Labor and learning. *Economic Development and Cultural Change*, 65(2), 265-96.
- Ersado, L. (2005). Child labor and schooling decisions in urban and rural areas: comparative evidence from Nepal, Peru and Zimbabwe. *World Development*, 33(3), 455–480.
- Fink, G., Jack, B.K., and Masiye, F. (2020). Seasonal liquidity, rural labor markets and agricultural production. *American Economic Review*, 110(11), 3351–3392.

- Gazeaud, J., and Ricard, C. (2024). Learning effects of conditional cash transfers: The role of class size and composition. *Journal of Development Economics*, 166, 103194.
- Gelli, A., Aberman, N., Margolies, A., Santacroce, M., Baulch, B., and Chirwa, E. (2017). Lean-season food transfers affect children's diets and household food security: evidence from a quasi-experiment in Malawi. *The Journal of Nutrition*, 147(5), 869-878.
- Gilbert, C.L., Christiaensen, L., and Kaminski, J. (2017). Food price seasonality in Africa: measurement and extent. *Food Policy*, 67(2), 119-32.
- Gitter, S.R., and Barham, B.L. (2007). Credit, natural disasters, coffee, and educational attainment in rural Honduras. *World Development*, 35(3), 498-511.
- Goldstein, M. and Udry, C. (2008). The profits of power: land rights and agricultural investment in Ghana. *Journal of Political Economy*, 116(6), 981-1022.
- Gollin, D., and Rogerson, R. (2014). Productivity, transport costs and subsistence agriculture. *Journal of Development Economics*, 107, 38-48.
- Gollin, D., Lagakos, D., and Waugh, M.E. (2014). Agricultural productivity differences across countries. *American Economic Review: Papers & Proceedings*, 104(5), 165-70.
- Gørgens, T. and Würtz, A.H. (2019). Threshold regression with endogeneity for short panels. *Econometrics*, 7(23).
- Gubert, F., and Robilliard, A. (2008). Risk and schooling decisions in rural Madagascar: a panel data-analysis. *Journal of African Economies*, 117(2), 207–238.
- Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64, 413-430.
- Hansen, B.E. (1999). Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of Econometrics*, 93, 345-368.
- Hazarika, G., and Sarangi, S. (2008). Household access to microcredit and child work in rural Malawi. *World Development*, 36(5), 843–859.
- Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., and Moreira, V. (2014). Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. *Journal of Development Economics*, 107: 144-56.
- Hirvonen, K., and Hoddinott, J. (2017). Agricultural production and children's diets: evidence from rural Ethiopia. *Agricultural Economics*, 48(4), 469-480.

- Hoddinott, J., Sandström, S., and Upton, J. (2018). The impact of cash and food transfers: evidence from a randomized intervention in Niger. *American Journal of Agricultural Economics*, 100(4): 1032-1049.
- Holden, S. and 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.
- Ikeno, J. (2011). Dry-season irrigation farming at the western foot of the North Pare Mountains, Tanzania. *African Study Monographs* Supplementary issue 2011, 42: 59-77,
- Jacoby, H.G., (1994). Borrowing constraints and progress through school: evidence from Peru. *Review of Economics and Statistics*, 76(1), 151– 160.
- Janzen, S., and Carter, M.R. (2019) After the drought: the impact of microinsurance on consumption smoothing and asset protection. *American Journal of Agricultural Economics*, 101(3), 651-671.
- Jensen, R. (2000). Agricultural volatility and investments in children. *American Economic Review*, 90(2), 399-404.
- Jones, A.D., Shrivinas, A., and Bezner-Kerr, R. (2014). Farm production diversity is associated with greater household dietary diversity in Malawi: findings from nationally representative data. *Food Policy*, 46, 1-12.
- Kassie, B.T., Hengsdijk, H., Rötter, R., Kahiluoto, H., Asseng, S., and Ittersum, M.V. (2013). Adapting to climate variability and change: Experiences from cereal-based farming in the Central Rift and Kobo Valleys, Ethiopia. *Environmental Management*, 52:1115-1131.
- Kourtellos, A., Thanasis, S., and Tan, C.M. (2016). Structural threshold regression. *Econometric Theory*, 32, 827-860.
- Kreps, D.M. (2012). *Microeconomic Foundations I: Choice and Competitive Markets*, Princeton University Press.
- Lagakos, D., and Waugh, M.E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2), 948-80.
- Lay, J., Mahmoud, T.O., and M' Mukaria, G.M. (2008). Few opportunities, much desperation: the dichotomy of non-agricultural activities and inequality in Western Kenya. *World Development*, 36(12), 2713-2732.

- Lechicki, A., and Spakowski, A. (1985). A note on intersection of lower semicontinuous multifunctions. *Proceedings of the American Mathematical Society*, 95(1), 119-122.
- Loison, S.A. (2015). Rural livelihood diversification in Sub-Saharan Africa: a literature review. *The Journal of Development Studies*, 51:9, 1125-1138.
- Lu, B., Ren, H.L., Scaife, A.A., Wu, J., Dunstone, N., Smith, D., Wan, J., Eade, R., MacLachlan, C., and Gordon, M. (2018). An extreme negative Indian Ocean Dipole event in 2016: dynamics and predictability. *Climate Dynamics*, 51, 89-100.
- Lukmanji, Z., Hertzmark, E., Mlingi, N., Assey, V., Ndossi, G., and Fawzi, W. (2008). Tanzania food composition tables. Muhimbili University of Health and Allied Sciences (MUHAS), Dar es Salaam in Tanzania, Tanzania Food and Nutrition Centre (TFNC), Dar es Salaam in Tanzania, and Harvard School of Public Health (HSPH), Boston in USA.
- Maggio, G. and Sitko, N.J. (2021). Diversification is in the detail: accounting for crop system heterogeneity to inform diversification policies in Malawi and Zambia, *The Journal of Development Studies*, 57:2, 264-288.
- Makate, C., Angelsen, A., Holden, S.T., Westengen, O.T. (2022). Crops in crises: shocks shape smallholders' diversification in rural Ethiopia. *World Development*, 159, 106054.
- Makate, C., Angelsen, A., Holden, S.T., Westengen, O.T. (2023). Evolution of farm-level crop diversification and response to rainfall shocks in smallholder farming: evidence from Malawi and Tanzania. *Ecological Economics*, 205, 107736.
- Mani, S., Hoddinott, J., and Strauss, J. (2013). Determinants of schooling: empirical evidence from rural Ethiopia. *Journal of African Economies*, 22(5), 693-731.
- Martey, E., Etwire, P. M., and Atinga, D. (2021). To attend or not to attend: Examining the relationship between food hardship, school attendance and education expenditure. *International Journal of Educational Development*, 80, 102304.
- Martey, E., Etwire, P.M., Mockshell, J., Armah, R., and Akorsikumah, E. (2023). Ecological shocks and children's school attendance and farm work in Ghana. *World Development Perspectives*, 31, 100529.
- Mbigi, D. and Xiao, Z. (2021). Tanzanian rainfall responses to El Niño and positive Indian Ocean dipole events during 1951–2015. *Atmospheric and Oceanic Science Letters*, 14(6), 100093.

- McCord, P.F., Cox, M., Schmitt-Harsh, M., and Evans, T. (2015). Crop diversification as a smallholder livelihood strategy within semi-arid agricultural systems near Mount Kenya. *Land Use Policy*, 42, 738-750.

- Mendola, M., and Negasi, M.Y. (2023). Nutritional and schooling impact of a social protection program in Ethiopia: a retrospective analysis of childhood exposure, *Journal of African Economies*, , 00, 00, 1–21

- Michler, J.D. and Josephson, A.L. (2017). To specialize or diversify: agricultural diversity and poverty dynamics in Ethiopia. *World Development*, 89, 214-226.

- Millán, T.M., Macours, K., Maluccio, J.A., and Tejerina, L. (2020). Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *Journal of Development Economics*, 143, 102385.

- Ministry of Education, Science and Technology (MoEST). (2017). Education sector and development plan (2016/17-2020/21). Dar es Salaam, Tanzania (Updated July 2018).

- Morgan, S., Mason, N.M., Levine, N.K., and Mbata-Zulu, O. (2019). Dis-incentivizing sustainable intensification? The case of Zambia's maize-fertilizer subsidy program. *World Development*, 122, 54-69.

- Mulwa, C.K. and 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.

- Muralidharan, K., and Prakash, N. (2017). Cycling to school: increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*, 9(3), 321-50.

- Musumba, M., Palm, C.A., Komarek, A.M., Mutuo, P.K., and Kaya, B. (2022). Household livelihood diversification in rural Africa. *Agricultural Economics*, 53(2), 246-256.

- Mutai, C.C. and Ward, M.N. (2000). East African rainfall and the tropical circulation/convection on intraseasonal to interannual timescales. *Journal of Climate*, 13, 3915-3939.

- Murtazashvili and Wooldridge (2016). A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. *Journal of Econometrics*, 190(2), 252-266.

- National Bureau of Statistics. (2022). Basic information document: National Panel Survey (NPS 2020-2021). Dar es Salaam, Tanzania.
- Ndegwa, M.K., De Groote, H., Gitonga, Z.M., and Bruce, A.Y. (2016). Effectiveness and economics of hermetic bags for maize storage: results of a randomized controlled trial in Kenya. *Crop Protection*, 90, 17-26.
- Nindi, T., Ricker-Gilbert, J., and Bauchet, J. (2023). Incentive mechanisms to exploit intraseasonal price arbitrage opportunities for smallholder farmers: Experimental evidence from Malawi. *American Journal of Agricultural Economics*, 1-24.
- Oestigaard, T. (2016). Rainfed agriculture, drought and hunger in Tanzania. In Tvedt, T. and Oestigaard, T. (eds.). A History of Water, Series 3, Volume 3. Water and Food: From Hunter-Gatherers to Global Production in Africa: 332-354. I.B. Tauris. London.
- Omotilewa, O., Ricker-Gilbert, J., Ainembambazi, J.H., and Shively, G. (2018). Does improved storage technology promote modern input use and food security? Evidence from a randomized trial in Uganda. *Journal of Development Economics*, 135, 176-98.
- 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.
- Premand, P., and Barry, O., 2022. Behavioral change promotion, cash transfers and early childhood development: experimental evidence from a government program in a low-income setting. Journal of Development Economics, 158, 102921.
- Ricker-Gilbert, J., Omotilewa, O., and Kadjo, D. (2022). The economics of postharvest loss and loss-preventing technologies in developing countries. *Annual Review of Resource Economics*, 14, 243–65.
- Rivera-Padilla, A. (2020). Crop choice, trade costs, and agricultural productivity. *Journal of Development Economics*, 146, 102517.
- Seo, M.H. and Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195, 169-186
- Shimamura, Y., and Lastarria-Cornhiel, S. (2010). Credit program participation and child schooling in rural Malawi. *World Development*, 38(4), 567–580.
- Schultz, T.P. (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74, 199-250.

- Smale, M., Hartell, J., Heisey, P.W., and Senauer, B. (1998). The contribution of genetic resources and diversity to wheat production in the Punjab of Pakistan. *American Journal of Agricultural Economics*, 80(3), 482-493.
- Steger, T.M. (2000). Economic growth with subsistence consumption. *Journal of Development Economics*, 62(2), 343-361.
- Swindale, A., and Bilinsky, P. (2006). Household Dietary Diversity Score (HDDS) for measurement of household food access: indicator guide. Food and Nutrition Technical Assistance Project (FANTA), Academy for Educational Development, Washington, DC.
- Takayama, A. (1985). *Mathematical Economics (Second edition)*. New York: Cambridge University Press.
- Tesfaye, W. and Tirivayi, N. (2020). Crop diversity, household welfare and consumption smoothing under risk: evidence from rural Uganda. *World Development*, 125, 104686.
- Thomas, D., Beegle, K., Frankenberg, E., Sikoki, B., Strauss, J., and Teruel, G. (2004). Education in a crisis. *Journal of Development Economics*, 74, 53–85.
- Waha, K., Dietrich, J.P., Portmann, F.T., Siebert, S., Thornton, P.K., Bondeau, A., and Herrero, M. (2020). Multiple cropping systems of the world and the potential for increasing cropping intensity. *Global Environmental Change*, 64, 102131.
- Walls, H., Johnston, D., Matita, M. Chirwa, E., Mazalale, J., Quaife, M., Kamwanja, T., & Smith, R. (2023). How effectively might agricultural input subsidies improve nutrition? A case study of Malawi's Farm Input Subsidy Programme (FISP). *Food Security*, 15, 21-39.
- Weinberger, K., and Lumpkin, T. (2007). Diversification into horticulture and poverty reduction: a research agenda. *World Development*, 35(8), 1464-1480.
- Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data*, second edition. MIT Press, Cambridge, MA, America.
- Wooldridge, J.M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics*, 182, 226-234.
- Wooldridge, J.M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2), 420-445.
- Yu, P., and Phillips, P.C.B. (2018). Threshold regression with endogeneity. *Journal of Econometrics*, 203, 50-68.

- Zimmerman, F.J., and Carter, M.R. (2003). Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints. *Journal of Development Economics*, 71, 233-260.