EAJOS Vol. 2,No. 1, June 2026 ISSN (Online): 3088-5426 | ISSN (Print): 2953-2728

Eastern Africa Journal of Official Statistics

Website: https://journal.eastc.ac.tz/index.php/eajos/index

Determinants of Agricultural Productivity Among Smallholder Farmers in Tanzania: Evidence from the National Panel Survey, 2021

[1]Zainabu Hassan, [2]Magreth Kimaro and [3]Evod Rimisho

[1,2,3]Eastern Africa Statistical Training Centre P. O. Box 35103, Dar es Salaam, Tanzania.

Corresponding Author Email: evod.rimisho@eastc.ac.tz

KEYWORDS

Agricultural productivity, Smallholder farming, Generalized Linear Model

ABSTRACT

Agricultural productivity in Tanzania remains low, limiting food security and rural livelihoods. This study analyzed determinants of smallholder staple crop harvests using Wave 5 of the National Panel Survey, based on 237 households with complete socio-economic, farm input, and environmental data. A generalized linear model (GLM) with a Gamma distribution and log link was employed to account for the right-skewed, strictly positive nature of harvest data. Sequential block-wise models were estimated for socio-economic, farm input, and environmental/institutional factors, followed by a full combined model. The socio-economic model showed positive effects of asset ownership ($\beta = 0.1072$, $p < 0.001$) and negative effects of older household heads ($\beta = -0.0159$, $p < 0.05$), with log pseudolikelihood = −1804.43, AIC = 15.28, BIC = −907.20, and deviance = 355.92. The farm input model identified plot size as strongly positive ($\beta = 0.216$, $p < 0.001$) and irrigation as negative ($\beta = -1.701$, $p < 0.05$), yielding log pseudolikelihood = −1802.23, AIC = 15.28, BIC = −895.19, and deviance = 351.53. The environmental/institutional model performed poorly (log pseudolikelihood = −1840.04, AIC = 15.59, BIC = −830.50, deviance = 427.15). The full model provided the best fit (log pseudolikelihood = −1775.08, AIC = 15.15, BIC = −889.35, deviance = 297.22), showing positive effects of plot size, asset ownership, household size, and organic fertilizer, and negative effects of irrigation, hired labor, and older household heads. Findings emphasize the importance of strengthening household resources and improving input efficiency to enhance productivity.

1.0 Introduction

Agriculture remains a key sector of Tanzania's economy, contributing significantly to employment, food security, and economic growth. In 2022, crop cultivation accounted for the largest share of agricultural GDP, contributing 15.0% (The United Republic of Tanzania [URT], 2023). Despite the agricultural importance in the economy, smallholder productivity continues to face persistent challenges (Zerssa *et al.,* 2021) . Smallholder farmers, who typically cultivate less than two hectares, rely heavily on rain-fed agriculture and traditional practices, leaving them vulnerable to climate shocks, soil degradation, and low input use (Assefa et al., 2022; Zeleke et al., 2024). Limited access to credit, poor infrastructure, weak market linkages, and insufficient extension services further constrain productivity.

Several initiatives, including the Agricultural Sector Development Programme Phase II, have been introduced to improve agricultural productivity through technology adoption, input support, and climate-smart practices. However, maize, bean, and paddy yields remain low and below regional and international benchmarks (Kadigi, 2025).

Existing studies on productivity determinants often focus on a single crop (e.g., maize, beans, or paddy) and analyze a limited set of factors, such as technology adoption, credit access, or land size (Gcaba *et al*., 2025; Wambede *et al.*, 2025). Consequently, these studies provide only a fragmented understanding of the broader determinants of smallholder productivity, limiting the usefulness of findings for designing holistic interventions (Shokati *et al.*, 2025; Zeleke *et al.*, 2021).

To address these gaps, this study adopts a comprehensive approach examining multiple staple crops and integrating socio-economic, farm input, environmental, and institutional factors. The goal is to generate actionable evidence to inform policies and interventions that enhance smallholder productivity, resilience, and sustainable rural development, in alignment with the Sustainable Development Goals (SDGs): SDG 1 – No Poverty, SDG 2 – Zero Hunger, and SDG 8 – Decent Work and Economic Growth.

2.0 Materials and Methods

2.1 Study Area

The study was conducted in the United Republic of Tanzania (URT). The country is located in East Africa and lies between Longitudes 29^0 and 41^o East and Latitudes 1^o and 12^o South (URT, 2013). In total, the country has 948,710 square kilometers with a total population of 61,741,120. The country is divided into thirty-one (31) regions, of which five (5) are located in the Zanzibar islands and twenty-six (26) in the Tanzania mainland (URT, 2022). Each region is further divided into several districts. The districts are subdivided into wards, and the wards are subdivided into villages/streets.

Agriculture is the backbone of Tanzania's economy, encompassing crop production and livestock rearing for food and commercial use. The crop sub-sector is dominated by smallholder farmers who mainly grow staple crops such as maize, beans, and paddy, which are essential for household food security and national food supply. Tanzania was selected as the study area due to the importance of agriculture and the central role of these staple crops in rural livelihoods and national food security.

2.2 Population of the Study

The study population comprised all households in Tanzania that participated in Wave 5 of the National Panel Survey and were actively engaged in the production of maize, paddy, or beans during the reference period. These staple crops were selected due to their significance for household food security and rural livelihoods. Although the survey collected data for both short and long rainy seasons, this study used only long rainy season data to reduce seasonal variation. This period was chosen because it is the main agricultural season in Tanzania, when most households cultivate staple crops, ensuring better comparability and reflecting peak production activity.

2.3 Sampling Method

This study used the full available subsample, which included all households in the Wave 5 National Panel Survey (NPS) that met the following inclusion criteria: (1) actively cultivating at least one of the selected staple crops (maize, paddy, or beans) during the reference period, and (2) having complete data for key production and socio-economic variables. No additional sampling was conducted, ensuring comprehensive coverage of eligible smallholder farming households and minimizing sampling bias within the survey frame. It is important to note that the National Panel Survey itself used a multi-stage stratified sampling design to select households, ensuring national and regional representativeness (URT, 2022). While this design was implemented by the NPS and not by the current study, it provides confidence that the dataset accurately reflects the diversity of smallholder farming households across Tanzania.

2.4 Data Collection Methods

This study used secondary data from Wave 5 of the Tanzania National Panel Survey (NPS), conducted by the National Bureau of Statistics (NBS) in collaboration with the World Bank under the Living Standards Measurement Study-Integrated Surveys on Agriculture initiative. The NPS is a nationally representative longitudinal household survey that collects detailed information on agriculture and socio-economic conditions. Data were collected through face-to-face interviews using structured Household, Agriculture, and Community Questionnaires. The Agriculture Questionnaire captured information on crop production, farm inputs, irrigation, mechanization, and hired labor, while the Household Questionnaire provided socio-economic data such as household size, age, education, sex of the household head, and asset ownership. Secondary data were obtained from the Tanzania National Data Archive managed by the NBS, providing reliable information for analyzing smallholder agricultural productivity in Tanzania.

2.5 Study Variables

The study included 20 variables: one dependent variable, five socio-economic variables, eight farm input variables, and six environmental/institutional variables. Agricultural productivity, the dependent variable, was measured as the total annual quantity harvested per household. Although productivity is commonly measured as yield, area-planted data were not consistently available; therefore, total crop output was used as a proxy. Table 1 summarizes the variables, their descriptions, measurement scales, NPS Wave 5 data sources, and corresponding dataset variable names.

Table 1: Description of Study Variables and Their Measurement

Category	Variable Name	Description	Scale	NPS Section (variable name)
Dependent	Agricultural productivity	Total annual harvest of all three staple crops (maize + beans + paddy) measured in kilogram	Ratio	AG_SEC_4B.dta (ag4b_28)
Socio-economic Factors (Independent Variables)	Household size	Number of people in the household	Ratio	HH_SEC_B.dta Computed from the household roster using y5_hhid
	Education level of household head	The highest grade completed	Nominal Re-categorized into four categories	HH_SEC_C.dta (hh_c07)
	Sex	Whether the household head is male or female	Binary	HH_SEC_B.dta (hh_b02)
	Age	Age of household head in years	Ratio	HH_SEC_B.dta (hh_b04)
	Assets	Ownership of assets (Count of assets owned)	Ratio	HH_SEC_M.dta (count owned asset using hh_m00 == 1)
Farm-Input Factors (Independent Variables)	Organic fertilizer	Whether organic fertilizer was applied to a plot	Binary	AG_SEC_3A (ag3a_41)
	Inorganic fertilizer	Whether inorganic fertilizer was used	Binary	AG_SEC_3A (ag3a_47)
	Pesticides	Whether plot was treated with pesticides	Binary	AG_SEC_3A (ag3a_65a)
	Herbicides	Whether plot was treated with herbicides	Binary	AG_SEC_3A (ag3a_60)
	Plot size	Size of the plot (measured by GPS or reported measured in acres)	Ratio	AG_SEC_02 (ag2a_09) or (ag2a_04)
	Labour	Whether the household hired labour	Binary	AG_SEC_3A (ag3a_73)
	Mechanization	Whether the household used mechanization (eg tractors)	binary	AG_SEC_3A (ag3a_71a)
	Irrigation	Whether plot was irrigated	Binary	AG_SEC_3A (ag3a_18)
Environmental / Institutional Factors	Extension Services	Whether household accessed agricultural extension	Binary	AG_SEC_12A.dta (ag12a_02)

Category	Variable Name	Description	Scale	NPS Section (variable name)
(Independent Variables)	Distance	Distance from the plot to the road (km)	Ratio	AG_SEC_3A (ag3a_02_2)
	Region	Region of the household categorized into agro-ecological zones	Nominal	HH_SEC_A.dta (hh_a01_1)
	Crop diseases	Whether crop was affected by diseases	Binary	HH_SEC_R.dta (shockid - 102)
	Drought	Whether there was any problem with drought	Binary	HH_SEC_R.dta (Shockid -101)
	Erosion	Whether there was any problem with erosion	Binary	AG_SEC_3A (ag3a_13)

Source: Research Data (2026)

2.6 Data Analysis Techniques

Prior to data analysis, the extracted datasets from the National Panel Survey (NPS) Wave 5 were carefully processed to ensure accuracy and consistency. This involved merging the household and agriculture datasets using unique household identifiers and selecting only the relevant variables for maize, beans, and paddy. The original crop-level data contained 752 observations of households that cultivated at least one of the selected staple crops: maize (520 observations, 69.15%), beans (204 observations, 27.13%), and paddy (28 observations, 3.72%). However, since some households cultivated more than one crop during the same reference period, the data were combined so that each household appeared only once. This produced 432 unique households, and agricultural output was measured as the total harvest of the selected crops, which serves as the dependent variable in this study.

The aggregation of the dependent variable is used to capture overall household food productivity rather than crop-specific performance. The approach is consistent with studies that examine household-level agricultural output across multiple crops when the focus is on total production capacity rather than crop-specific productivity. To ensure transparency, the descriptive statistics in Table 2 report the share of each crop in the total harvest reported by the 752 households that engaged in cultivating maize, beans and paddy during the reference period. The results show that household agricultural output is dominated by maize, which contributes 47.89% of total annual production. Paddy follows closely with 38.80%, while beans account for the smallest share at 13.31%. This indicates that maize is the most important staple crop in terms of quantity produced by households in the sample, with paddy also playing a major role in overall production.

Table 2: The Share of each Crop in Total Harvest

Crop	Mean (kg/year)	Total harvest (kg/year)	Share of Total Harvest (%)
Maize	429.3423	8640	47.89
Paddy	1146.821	7000	38.80
Beans	104.4466	2400	13.31
Total	-	18,040	100

Source: Research Data (2026)

The agricultural data were then merged with socio-economic, farm input, and environmental variables from other survey modules. Although 1,941 households had complete information on these explanatory variables, only 237 households had complete data for both the total harvest (dependent variable) and all required independent variables. This was because 195 households reported harvest outcomes but lacked information in one or more explanatory variables.

The final sample of 237 households was therefore obtained using a complete-case approach, meaning that only households with no missing values were included in the analysis. This reduction was caused by missing information across different survey sections rather than any intentional selection of households. Given that the missingness arises from survey non-response across modules, the analysis proceeds under the assumption that the missing data are conditionally random with respect to key observable characteristics. While this approach ensures consistency in the analysis, it also implies that the results are based on a smaller sample than the full survey.

After completing the data processing steps described above, a clean dataset was obtained for analysis. The dataset includes full information on household agricultural output and all selected explanatory variables for the final sample of 237 households. The following sections present the descriptive and inferential analysis conducted in this study.

2.6.1 Descriptive Statistics

Descriptive Statistics were first computed to summarize the characteristics of the study variables. The study used percentages, means, standard deviations, minimum values, and maximum. In addition, skewness, and kurtosis were also determined to assess the shape and distribution of the data. The distribution of agricultural productivity was further examined using box plot to visually assess normality, spread, and the presence of extreme values. To formally test for normality, the Shapiro–Francia test was applied.

2.6.2 Correlation Analysis and Multicollinearity Tests

A Spearman correlation coefficient was first employed to examine pairwise associations among the study variables. This step provided preliminary evidence of how independent variables were related to agricultural productivity. Given the possibility of overlap among predictors, a Variance Inflation Factor (VIF) test was subsequently conducted following an ordinary least squares (OLS) regression. The VIF was calculated to detect multicollinearity within the regression framework. According to Shrestha (2020) multicollinearity arises when two or more independent variables are highly correlated with each other, which can distort the estimation of their individual effects on the dependent variable. In this study, a VIF value greater than 10 was considered indicative of potential multicollinearity.

2.6.3 Generalized Linear Models

Preliminary analysis of the dependent variable (agricultural productivity) indicated a highly right-skewed distribution. Although logarithmic transformations were considered to reduce skewness, the data remained non-normal. Given the positive, continuous, and skewed nature of agricultural productivity, a Generalized Linear Model (GLM) with a Gamma distribution and a log link function was employed. The Gamma distribution is appropriate for modeling strictly positive continuous outcomes with skewed distributions, while the log link ensures predicted values remain positive and allows for multiplicative effects of explanatory variables. The general specification of the model is given by: -

$$\ln\left(E(Y_i)\right) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ki} + \varepsilon_i$$

Where:

Y_i =Total staple harvest (agricultural productivity) for household i

$\mathrm{E}(Y_i)$ =Expected value of agricultural productivity

X_{ki} =Vector of explanatory variables

β_k =Parameters to be estimated

ε_i =Error term

To examine the contribution of different categories of determinants, four models were estimated sequentially. The first model included socio-economic variables: age, sex, education level, household size, and asset ownership. The second model included farm input variables: plot size, labour, irrigation, herbicide, inorganic fertilizer, mechanization, organic fertilizer, and pesticide use. The third model included environmental and institutional variables: extension services, crop diseases, agro-ecological zone, erosion, drought/flood exposure, and distance to services. The last model represented the full model, incorporating all socio-economic, farm input, and environmental/institutional variables simultaneously.

Robust standard errors were used in all regression models to address potential heteroscedasticity and ensure consistent variance estimates, particularly given the skewed distribution and extreme values in agricultural productivity data. To compare the model fitting performance of the four Gamma GLM models, log pseudolikelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and deviance were used. A higher (less negative) log pseudolikelihood, lower AIC, lower BIC, and lower deviance all indicate better model performance.

3.0 Results and Discussion

3.1 Results of Descriptive Analysis

The descriptive statistics for the study's categorical variables are presented in Table 3. The data reveals that the sample (n=237) is primarily concentrated in the Lake Zone (63.3%) and the Northern Zone (24.9%). Other regions, including the Eastern, Southern, Western, and Zanzibar zones, each accounted for less than 10% of the total participants. In terms of gender, the sample was heavily skewed toward male respondents, who accounted for 79.3% of the participants. The educational background of the respondents was relatively low; a majority (67.1%) had completed only primary education, while 23.2% reported having no formal education. Only 3.0% of the sample had attained post-secondary or university-level qualifications.

Agricultural practices during the long rainy season were characterized by low levels of modern input adoption. Only 8.4% of farmers utilized inorganic fertilizers, and a similarly low proportion applied herbicides (9.3%) and pesticides (14.8%). Organic fertilizer use was slightly more prevalent at 17.7%. Mechanization was limited to 8.0% of the plots, and irrigation was utilized by only 2.5% of the respondents, indicating a strong dependence on rain-fed agriculture. Despite the low use of chemical and mechanical inputs, hiring labor was a relatively common practice, reported by 37.6% of the households.

Farmers in the study area faced several environmental constraints. The most common challenge was crop diseases, which affected approximately 21% of households. In addition, 13.5% of respondents reported soil erosion during the long rainy season, while 11.8% indicated that their plots had been affected by droughts or floods within the past two years. Institutional support to mitigate these challenges appeared minimal, as only 9.3% of respondents reported receiving agricultural advice or extension services from the government or other institutions.

Table 3: Descriptive Statistics for the Categorical Variables

Categorical Variables	**Number (Percentage)**
Zone	
Eastern Zone	16 (6.75%)
Northern Zone	59 (24.89%)
Lake Zone	150 (63.29%)
Southern Zone	3 (1.27%)
Southern highland Zone	3 (1.27%)
Western Zone	4 (1.69%)
Zanzibar	2 (0.84%)
Sex	
Male	188 (79.3%)
Female	49 (20.7%)
Education level	
No education	55 (23.2%)
Primary	159 (67.1%)

Secondary	16 (6.8%)
Post-Secondary/University	7 (3.0%)
Erosion	
Yes	32 (13.5%)
No	205 (86.5%)
Irrigation	
Yes	6 (2.5%)
No	231 (97.5%)
Organic Fertilizer	
Yes	42 (17.7%)
No	195 (82.3%)
In organic fertilizer	
Yes	20 (8.4%)
No	217 (91.6%)
Herbicide	
Yes	22 (9.3%)
No	215 (90.7%)
Pesticide	
Yes	35 (14.8%)
No	202 (85.2%)
Mechanization	
Yes	19 (8.0%)
No	218 (92.0%)
Labour	
Yes	89 (37.6%)
No	148 (62.4%)
Extension services	
Yes	22 (9.3%)
No	215 (90.7%)
Drought/flood	
Yes	28 (11.8%)
No	209 (88.2%)
Crop diseases	
Yes	50 (21.1%)
No	187 (78.9%)

Source: Research Data (2026)

Table 4 shows the descriptive statistics for the continuous variables (age, household size, asset count, distance to the plot, and plot size). The table shows on average, farmers reported a total staple harvest (agricultural productivity) of about 898 kilograms, though the distribution was highly unequal. While most households harvested relatively small amounts, a few reported extremely large harvests, with yields reaching up to 29,000 kilograms. The distribution was extremely positively skewed (skewness = 8.99) and highly leptokurtic (kurtosis = 94.48), indicating that the majority of households clustered around low harvest values, while a small number of extreme outliers with very high yields heavily influenced the mean.

The demographic profile of respondents shows that farming is largely undertaken by middle-aged individuals, with an average age of 50 years. The age distribution was fairly balanced, ranging from 22 to 89 years, with only mild positive skewness (0.40) and a kurtosis value of 2.55, which is close to the normal distribution benchmark of 3. This suggests that the age variable is approximately normally distributed, with no extreme outliers.

Household size averaged approximately 6 members, with some households as small as one person and others as large as twenty. The distribution was moderately positively skewed (1.15) and moderately leptokurtic (kurtosis = 4.89), indicating that most households are medium-sized, but a few very large households exist, creating a heavier tail in the distribution.

Asset ownership provides another dimension of socioeconomic status. On average, households reported 12 assets, with a range from 3 to 27. The distribution showed moderate positive skewness (0.83) and kurtosis of 4.32, suggesting that while most households own a modest number of assets, a few relatively wealthier households own significantly more, resulting in a slightly peaked distribution with heavier tails.

Spatial characteristics also reveal significant inequalities. The average reported distance from the plot to the road was 1.7 kilometers, but the distribution was extremely skewed (skewness = 13.02) and highly leptokurtic (kurtosis = 188.41). This indicates that while most households are located close to roads, a small proportion are situated at very long distances, creating strong outliers and a distribution with extreme tails.

The plot size variable shows that households cultivate relatively small parcels of land, with an average size of 2.48 acres. The minimum reported plot is extremely small at 0.02 acres, while the maximum is 6.42 acres, which falls just below the national benchmark of 6.67 acres. The standard deviation of 1.52 acres indicates moderate variation across households, but the skewness value of 0.72 suggests that the distribution is positively skewed, meaning most households operate smaller plots, with a few larger ones pulling the distribution to the right. The kurtosis of 3.06, slightly above the normal distribution benchmark, implies a sharper peak and heavier tails, showing that while most households cluster around the mean, there are still some notable extremes. Overall, the plot size data reflects the typical smallholder farming structure in Tanzania, where most households cultivate plots in the 2–3 acre range, with limited cases of larger holdings.

Table 4: Descriptive Statistics for the Continuous Variables

Variables	Mean	Std. Dev.	Min	Max	Skew.	Kurt.
Agricultural productivity	897.81	2445.096	1	29000	8.99	94.48
Age	50.17	14.61	22	89	0.4	2.55
Household size	6.46	3.431	1	20	1.15	4.89
Asset	12.03	3.99	3	27	0.83	4.32
Distance	1.66	5.82	0	86	13.02	188.41
Plot size	2.48	1.52	0.02	6.42	0.72	3.06

Source: Research Data (2026)

Fig. 1 presents the box plot of total annual harvest of staple crops showing the distribution of agricultural productivity data. It can be observed that the distribution is not normally distributed as the whisker is longer on the positive side. The plot indicates the data is positively skewed and cannot take a zero value. It can also be observed that agricultural productivity data have positive outliers.

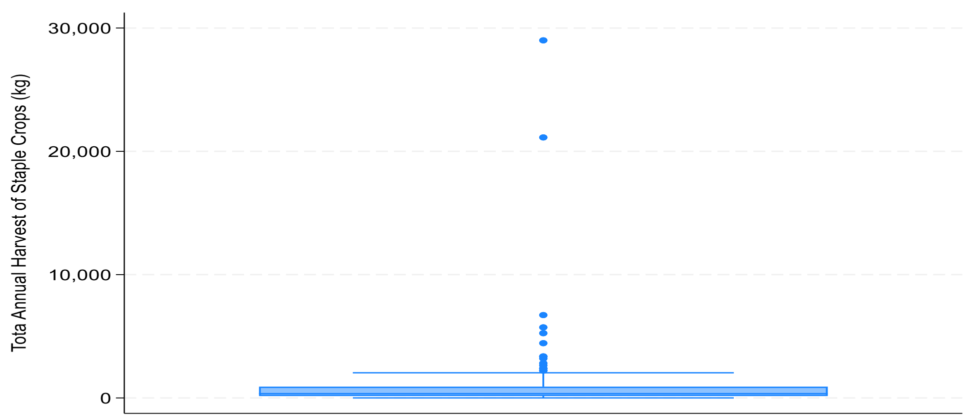

Fig. 1: Box Plot of Total Annual Harvest of Staple Crops

Source: Research Data (2026)

The Shapiro–Francia W' test was used to formerly test for normality. The Shapiro–Francia test results (W' = 0.2655, z = 10.305, $p < 0.001$) indicate that the distribution of total annual harvest significantly deviates from normality. A logarithmic transformation was applied to reduce skewness; however, the Shapiro–Francia test still rejected the null hypothesis of normality (W' = 0.981, $p = 0.004$), suggesting that the variable remains slightly non-normal.

3.2 Results of Inferential Analysis

3.2.1 Results of Correlation Analysis

A Spearman correlation analysis was conducted to examine the relationships between independent variables and total staple crops harvest (agricultural productivity). Table 5 presents the results of Spearman correlation. The table shows that plot size ($\rho = 0.201$, $p < 0.01$), household size ($\rho = 0.210$, $p < 0.01$), and asset count ($\rho = 0.172$, $p < 0.05$) are all positively and significantly correlated with total harvest, suggesting that larger landholdings, more family members, and greater resource endowments contribute to higher agricultural productivity. The positive relationship between household size and total harvest observed in this study is consistent with previous empirical evidence suggesting that larger households provide greater labour availability for farm activities, thereby improving agricultural productivity (Degefu, 2020). Moreover, the positive association between plot size and harvest outcomes contrasts empirical evidence which suggest that an apparent inverse size-productivity relationship as mentioned by Ayalew et al. (2024). Distance to services also shows a weak but significant positive correlation ($\rho = 0.156$, $p < 0.05$), which may reflect better access to markets or agricultural support.

In contrast, several variables exhibit statistically significant negative associations with harvest. Female household headship ($\rho = –0.115$, $p < 0.05$), age of household head ($\rho = –0.080$, $p < 0.05$), irrigation use ($\rho = –0.133$, $p < 0.05$), labour ($\rho = –0.102$, $p < 0.05$), and exposure to drought or floods ($\rho = –0.119$, $p < 0.01$) are all negatively correlated with production. However, the result for irrigation use should be interpreted with caution, as only a very small proportion of households (2.5%, n = 6) reported using irrigation, making this estimate highly sensitive to a few observations and potentially less stable than other correlations reported.

These findings imply that demographic characteristics, reliance on irrigation or labour, and climate shocks may constrain productivity. The adverse effect of climate shocks such as droughts and floods is also consistent with demonstrating that climate variability significantly reduces crop productivity through crop damage, soil degradation, and unstable rainfall patterns as explained by Lobell and Gourdji (2012). Again study by Ortiz-bobea et al. (2021) indicated that climate change has reduced global agricultural productivity growth, with particularly severe effects in warmer regions such as Africa. Other variables such as erosion ($\rho = –0.016$), crop diseases ($\rho = –0.025$), extension services ($\rho = –0.036$), and mechanization ($\rho = –0.031$) display very weak or non-significant correlations, indicating limited direct association with harvest outcomes in this dataset.

The correlation matrix also highlights clustering among input-use variables. Inorganic fertilizer is strongly correlated with irrigation ($\rho = 0.531$, $p < 0.01$), herbicide ($\rho = 0.478$, $p < 0.01$), and pesticide ($\rho = 0.558$, $p < 0.01$). Likewise research by Abate (2024) concur with the findings that fertilizer, irrigation, and crop protection inputs tend to be jointly adopted to maximize yield gains. Similarly, mechanization correlates with zone ($\rho = 0.315$, $p < 0.01$), and household size correlates with plot size ($\rho = 0.398$, $p < 0.01$). These patterns suggest that input-use variables tend to move together, which could raise concerns about multicollinearity in regression models.

Table 5: Spearman's rank correlation coefficients

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(1) total staple harvest	**1.00**																			
(2) plot size	0.20	**1.00**																		
(3) zone	0.06	0.03	**1.00**																	
(4) education	0.09	-0.05	0.06	**1.00**																
(5) Sex	-0.12	-0.15	0.05	-0.30	**1.00**															
(6) Age	-0.08	-0.01	-0.10	-0.25	0.32	**1.00**														
(7) household size	0.21	0.40	0.22	-0.01	-0.17	-0.07	**1.00**													
(8) asset count	0.17	0.19	-0.11	0.17	-0.15	0.04	0.22	**1.00**												
(9) distance	0.16	-0.03	0.05	-0.02	0.03	-0.14	0.02	-0.15	**1.00**											
(10) erosion	-0.02	-0.06	0.01	-0.08	0.02	0.03	-0.02	0.06	-0.06	**1.00**										
(11) irrigation	-0.13	0.13	0.07	0.07	-0.12	-0.06	0.10	-0.04	0.03	-0.06	**1.00**									
(12) Organic fertilizer	0.09	0.03	0.05	-0.03	-0.06	-0.12	-0.07	-0.26	0.08	0.11	-0.07	**1.00**								
(13) inorganic fertilizer	-0.08	0.08	0.17	-0.06	-0.11	-0.08	0.13	-0.18	0.13	-0.03	0.53	-0.06	**1.00**							
(14) herbicide	-0.06	-0.01	0.24	0.06	0.02	0.04	0.11	-0.25	0.08	-0.04	0.32	0.00	0.48	**1.00**						
(15) pesticide	-0.11	0.01	0.23	0.01	-0.08	0.05	0.01	-0.18	-0.05	-0.03	0.39	0.09	0.56	0.48	**1.00**					
(16) mechanization	-0.03	-0.05	0.32	-0.02	0.04	-0.04	0.06	-0.20	0.06	0.02	0.25	-0.06	0.25	0.28	0.18	**1.00**				
(17) labour	-0.10	-0.07	0.14	-0.12	-0.06	-0.03	0.16	-0.15	0.03	0.03	0.15	0.05	0.30	0.20	0.27	0.25	**1.00**			
(18) extension services	-0.04	-0.03	0.01	-0.05	-0.02	-0.06	-0.02	-0.05	0.06	0.04	0.13	0.08	0.11	0.00	0.15	0.01	0.11	**1.00**		
(19) drought/flood	-0.12	-0.09	-0.05	0.10	0.06	-0.02	-0.13	0.01	-0.09	-0.03	-0.06	0.00	-0.11	-0.12	-0.08	-0.11	-0.10	0.02	**1.00**	
(20) crop diseases	-0.03	-0.14	-0.06	-0.07	0.01	0.06	-0.07	-0.03	-0.12	0.04	-0.08	-0.02	0.03	0.05	0.10	-0.15	0.01	0.15	-0.03	**1.00**

Note: Spearman rho = -0.029. Orange = |r| ≥ 0.40. Diagonal = 1.00.

Source: Research Data, (2026)

3.2.2 Results of Multicollinearity Test

The results of VIF test for the independent variables of the study show that all VIF values were well below the conventional threshold of 10, with most predictors ranging between 1 and 2. The highest VIFs were observed for the zone categories (4.02–4.58) and inorganic fertilizer (2.26), but these values remain within acceptable limits. The mean VIF across all predictors was 1.65, confirming that the overall level of multicollinearity in the model is low. These findings suggest that, although some input-use variables and zone categories are moderately correlated, their combined effect does not inflate variance to problematic levels. Consequently, the independent variables can be retained in the GLM models without concern for severe multicollinearity. This diagnostic strengthens the reliability of the subsequent regression estimates and supports the robustness of the analytical framework.

3.2.3 Results of Models Fitting and Estimation

Table 6 shows the results of the first model that considered socio-economic characteristics. The results indicate that age of the household head has a statistically significant negative effect on total staple harvest ($\beta = -0.0159$, $p < 0.05$). This suggests that an increase in the age of the household head is associated with a decrease in agricultural productivity. Research by Barrett at el (2013) concur to findings that age can negatively affect farm productivity when it reduces the capacity to adopt modern inputs and innovative farming practices. Again study by Suri and Udry (2022), argued that agricultural productivity growth in Africa depends heavily on the adoption of improved technologies and modern inputs. Older farmers may be less likely to adopt new technologies and innovative farming practices, thereby reducing their productivity relative to younger farmers. Specifically, holding other factors constant, a one-year increase in age is associated with approximately a 1.6 percent reduction in staple harvest. This finding may reflect the declining physical capacity of older farmers or a lower likelihood of adopting improved agricultural practices.

Household asset ownership shows a positive relationship with staple harvest ($\beta = 0.0125$) and is marginally significant at the 10 percent level ($p < 0.10$). This implies that households with a greater number of productive assets tend to achieve higher staple harvests, possibly due to improved access to farming equipment, financial resources, or better production capacity. Similarly, this finding concurs with Hlatshwayo et al. (2023) who found that access to productive resources and socio-economic assets enhances crop productivity and food security by enabling farmers to invest in improved technologies and agricultural inputs, resulting in better harvests.

The remaining socio-economic variables, including household size, education level of the household head, and sex of the household head, were not found to have statistically significant effects on staple harvest. Although household size and education display positive coefficients, suggesting that larger households and more educated farmers may potentially produce more, these effects are not statistically distinguishable from zero in this model. Similarly, the sex of the household head does not appear to significantly influence staple production outcomes.

Overall, the results suggest that among the socio-economic characteristics considered in this model, age and household asset ownership are the most relevant factors associated with staple harvest.

Table 6: Results of Socio- Economic Model (Model 1)

Deviance = 355.9214085						
			AIC		=	15.27788
Log pseudolikelihood = -1804.428511			BIC		=	-907.2005
total_staple_harvest	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
hhsize	0.0298023	0.024754	1.20	0.229	-0.0187147	0.0783193
educ_cat	0.1351631	0.179627	0.75	0.452	-0.2168995	0.4872256
Age	-0.0158836	0.0065493	-2.43	0.015	-0.02872	-0.0030472
Sex	-0.0015158	0.2410544	-0.01	0.995	-0.4739738	0.4709421
asset_count	0.1072345	0.029819	3.60	0.000	0.0487904	0.1656786

Source: Research Data (2026)

Table 7 presents the findings of the second model. The table shows that plot size has a positive and statistically significant effect on total staple harvest ($\beta = 0.216$, $p < 0.01$). This suggests that an increase in the size of cultivated land is associated with higher staple production. Specifically, holding other factors constant, a one-unit increase in plot size leads to an approximate 21.6 percent increase in staple harvest, highlighting the importance of land availability in determining agricultural output.

Among the farm input variables, irrigation has a statistically significant negative effect on staple harvest ($\beta = -1.701$, $p < 0.05$). This result suggests that households using irrigation reported lower staple harvest compared to those not using irrigation. While this result may appear counterintuitive, it may reflect underlying conditions such as poor irrigation infrastructure, inefficient water management, or the possibility that irrigation is more commonly adopted in areas with unfavorable rainfall conditions or poorer soils.

The remaining input variables labour, herbicide, inorganic fertilizer, mechanization, organic fertilizer, and pesticide were not found to have statistically significant effects on staple harvest in this model.

Although labour, herbicide, inorganic fertilizer, mechanization, and pesticide display positive coefficients, indicating a potential positive association with agricultural output, these effects are not statistically significant at conventional levels. Organic fertilizer shows a negative coefficient but is also not statistically significant. These findings suggest that, within this sample, the contribution of these inputs to variations in staple harvest cannot be clearly distinguished from zero when controlling for other factors.

Overall, the results indicate that plot size is a key determinant of staple harvest among the farm input variables considered, while irrigation shows a significant negative association with production.

Table 7: Results of Farm Inputs Model (Model 2)

Deviance =	351.5252295					
		AIC		=		15.28464
Log pseudolikelihood =	-1802.230421	BIC		=		-895.1925
total_staple_harvest	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
plot_size	0.2163484	0.0544678	3.97	0.000	0.1095935	0.3231034
labour	-0.3345843	0.1906244	-1.76	0.079	-0.7082012	0.0390327
herbicide	0.2918028	0.3961487	0.74	0.461	-0.4846345	1.06824
inorganic_fertilizer	0.353295	0.4344024	0.81	0.416	-0.498118	1.204708
mechanization	0.4482665	0.297654	1.51	0.132	-0.1351245	1.031658
Organic_fertilizer	0.2801466	0.2523531	1.11	0.267	-0.2144565	0.7747496
pesticide	-0.8944545	0.5525577	-1.62	0.106	-1.977448	0.1885387
irrigation	-1.701019	0.7316618	-2.32	0.020	-3.13505	-0.2669885
_cons	9.022161	1.277677	7.06	0.000	6.517961	11.52636

Source: Research Data (2026)

Table 8 presents the findings of the third model. The table indicate that none of the explanatory variables related to environmental and institutional conditions namely extension services, crop diseases, agro-ecological zone, drought or flood shocks, and distance to market have statistically significant effects on staple harvest at conventional significance levels (their P-Values were greater than 0.05). However, this does not necessarily imply that these factors are unimportant for agricultural productivity, but rather suggests limitations in measurement and in the channels through which they operate.

In the case of extension services, the insignificant effect is consistent with evidence that the impact of agricultural extension depends not merely on access or contact, but on the quality, intensity, and relevance of services as well as farmer adoption of recommended practices (Abed et al., 2020). In many low-income settings, including Sub-Saharan Africa, limited linkages between agricultural research, extension services, and farmers together with uneven service delivery further reduce effectiveness even when services are available (Mapiye & Dzama, 2024).

Similarly, the lack of significant effects for environmental shocks may reflect measurement error in self-reported data and heterogeneity in exposure and adaptive capacity across households. Studies have shown that farmer-reported indicators of shocks such as droughts and crop diseases are often subject to recall bias and may not accurately capture severity or timing, leading to attenuation bias in regression estimates (Nguyen & Nguyen, 2020).

Table 8: Results of Environment/Institutional Model (Model 3)

Deviance =	427.1497854					
			AIC		=	15.58686
Log pseudolikelihood =	-1840.042699		BIC		=	-830.504
total_staple_harvest	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
extension_services	-0.0800569	0.3751185	-0.21	0.831	-0.8152756	0.6551618
crop_diseases	-0.5024309	0.4028638	-1.25	0.212	-1.292029	0.2871676
zone	-0.077378	0.1501691	-0.52	0.606	-0.371704	0.216948
erosion	0.1131045	0.2869726	0.39	0.693	-0.4493515	0.6755605
drought_flood	-0.021487	0.2340673	-0.09	0.927	-0.4802504	0.4372765
distance	0.0260845	0.0376274	0.69	0.488	-0.0476638	0.0998327
_cons	7.86646	1.057428	7.44	0.000	5.793939	9.93898

Source: Research (2026)

Table 9 presents the findings of the fourth model. The fourth model highlights several key determinants of agricultural productivity. Plot size is positively and significantly associated with harvests (0.115, p = 0.022), confirming that larger plots provide greater land availability and support higher yields. Similarly, asset ownership shows a positive and significant effect (0.054, p = 0.038), suggesting that households with more assets have better access to resources and are more resilient. Labour, in contrast, has a negative and significant impact (–0.403, p = 0.014), indicating that reliance on labour alone may reduce productivity, potentially due to inefficiencies or limited access to complementary technologies.
The use of organic fertilizer improves productivity significantly (0.560, p = 0.008), emphasizing its role in enhancing soil fertility, while household size also contributes positively (0.064, p = 0.015), likely by increasing available labour.

Some socio-demographic factors show marginal effects. Age of the household head is negatively associated with productivity (–0.010, p = 0.080), suggesting that older heads may face reduced physical capacity or slower adoption of innovations. Female-headed households tend to have lower productivity (–0.330, p = 0.072), though this effect is only marginally significant, possibly reflecting limited access to resources.

Table 9: Results of the full Model (Model 4)

Deviance	= 297.2181605					
		AIC		=		15.14833
Log pseudolikelihood	= -1775.076887	BIC		=		-889.3509
total_staple_harvest	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
plot_size	0.1153506	0.0502887	2.29	0.022	0.0167865	0.2139146
asset_count	0.0543159	0.0261149	2.08	0.038	0.0031316	0.1055003
labour	-0.4025992	0.1638153	-2.46	0.014	-0.7236712	-0.0815272
extension_services	0.0773033	0.2852159	0.27	0.786	-0.4817097	0.6363163
herbicide	0.2085809	0.3314088	0.63	0.529	-0.4409684	0.8581301
inorganic_fertilizer	0.2554753	0.3607183	0.71	0.479	-0.4515197	0.9624703
mechanization	0.3764318	0.2820696	1.33	0.182	-0.1764145	0.9292781
Organic_fertilizer	0.5596343	0.2122399	2.64	0.008	0.1436517	0.9756169
pesticide	-0.5025997	0.3314182	-1.52	0.129	-1.152167	0.146968
crop_diseases	-0.0701257	0.2105297	-0.33	0.739	-0.4827562	0.3425049
hhsize	0.0639785	0.0264308	2.42	0.015	0.0121751	0.115782
educ_cat	0.1227921	0.1357497	0.90	0.366	-0.1432725	0.3888566
Age	-0.0095986	0.0054859	-1.75	0.080	-0.0203508	0.0011535
zone	0.0879485	0.0814877	1.08	0.280	-0.0717644	0.2476614
Sex	-0.3300833	0.1833388	-1.80	0.072	-0.6894207	0.0292542
irrigation	-2.062058	0.6655846	-3.10	0.002	-3.36658	-0.7575366
erosion	-0.1515562	0.2325211	-0.65	0.515	-0.6072891	0.3041767
drought_flood	-0.1387225	0.2331176	-0.60	0.552	-0.5956245	0.3181796
distance	0.0322404	0.017628	1.83	0.067	-0.0023098	0.0667905
_cons	9.101108	1.559101	5.84	0.000	6.045326	12.15689

Source: Data Source (2026)

Table 9 further shows that irrigation consistently exhibits a statistically significant negative association with staple harvest, with coefficients remaining negative and significant across specifications (β = –1.701, p < 0.05 in the farm inputs model and β = –2.062, p = 0.002 in the full model). This finding is unexpected, given the theoretical expectation that irrigation enhances agricultural productivity. Therefore, the result should be interpreted with caution due to the very small number of irrigating households in the sample, which limits the statistical reliability of the estimated effect and makes it sensitive to a few observations. The negative coefficient should therefore not be interpreted as evidence that irrigation reduces productivity. Instead, it is more likely to reflect the specific characteristics of this sample and the context in which irrigation is practiced, where effects may vary depending on local conditions and implementation.

Distance to services exhibits a weak positive association (0.032, $p = 0.067$), potentially reflecting that households farther from services are situated in more fertile rural zones. Other variables, including extension services, herbicide, inorganic fertilizer, mechanization, pesticide use, crop diseases, agro-ecological zone, erosion, and exposure to drought or floods, did not show statistically significant effects in this model.

3.2.4 Models Performance Comparison

A comparison of the four Gamma GLM models based on fit statistics (Log pseudolikelihood, AIC, BIC, and Deviance) indicates varying explanatory performance. The socio-economic model (Model 1) shows the lowest AIC (15.2779) and BIC (-907.20), suggesting the best balance between model fit and complexity. The farm input model (Model 2) has the highest log pseudolikelihood (-1802.23) and lowest deviance (351.53), indicating slightly better ability to explain variation in staple harvest, though its AIC is marginally higher than Model 1. The environmental and institutional model (Model 3) performs worst, with the lowest log pseudolikelihood (-1840.04), highest AIC (15.5869), and highest deviance (427.15), showing limited explanatory power. The full model (Model 4), which combines socio-economic, farm input, and environmental/institutional variables, outperforms all others with the highest log pseudolikelihood (-1775.08), lowest AIC (15.15), lower BIC (-889.3509) compared to model 2 and 3, and lowest deviance (340.87), indicating it provides the most comprehensive fit and captures the greatest variation in staple harvest among smallholder households. Overall, model performance ranks as Model 4 > Model 2 ≈ Model 1 > Model 3.

4.0 Conclusion and Recommendations

This study examined determinants of agricultural productivity among smallholder farmers growing maize, paddy, and beans in Tanzania using Wave 5 of the Tanzania National Panel Survey (NPS). Results show that productivity is mainly driven by household resources and farm characteristics. Asset ownership, household size, plot size, and organic fertilizer use positively influenced harvests, while older household heads, hired labour, and irrigation were negatively associated with productivity. These findings align with evidence from Tanzania and other Sub-Saharan African countries, emphasizing the role of resource access and efficient farm management in improving smallholder performance. Overall, productivity gains depend not only on input access but also on effective resource use.

From a policy perspective, interventions should strengthen farmers' access to productive assets, promote organic fertilizer use, and improve farm management practices. Attention is also needed to address lower productivity associated with hired labour and irrigation use. Such measures can enhance food security, rural livelihoods, and contribute to SDGs 1, 2, and 8.

Several limitations should be noted. The study used a complete-case approach, reducing sample size and potentially introducing selection bias, which may limit generalizability. Crops (maize, paddy, beans) were aggregated into a single output measure, masking crop-specific differences. Future studies should conduct crop-specific analyses using multivariate methods. Productivity was measured using total harvest as a proxy due to missing area-planted data, preventing yield estimation; future research should use complete land data to compute yields per hectare. The small number of irrigation users also limited statistical power and the reliability of related estimates. Additionally, the cross-sectional design limits control for unobserved household heterogeneity. Future research should use larger and more balanced samples, distinguish irrigation technologies and intensity, and apply panel data with multilevel or mixed-effects models to better capture heterogeneity and dynamic effects.

Acknowledgement: The authors gratefully acknowledge the Eastern Africa Statistical Training Centre (EASTC) for providing academic and institutional support that facilitated the successful completion of this research.

Conflict of Interest: Authors declare no conflict of interest

REFERENCES

Abate, T. W. (2024). Analysis of agricultural technology adoption : the use of improved maize seeds and its determinants in Ethiopia , evidence from Eastern Amhara. *Journal of Innovation and Entrepreneurship*. https://doi.org/10.1186/s13731-024-00421-4

Abed, R., Sseguya, H., Flock, J., Mruma, S., & Mwango, H. (2020). *An Evolving Agricultural Extension Model for Lasting Impact : How Willing Are Tanzanian Farmers to Pay for Extension Services ?*

Assefa, D., Woldeamanuel, T., & Haile, F. (2022). Smallholder farmers ' vulnerability and adaptation to climate change induced shocks : The case of Northern Ethiopia highlands. *Journal of Agriculture and Food Research*, *8*(April), 100312. https://doi.org/10.1016/j.jafr.2022.100312

Ayalew, H., Chamberlin, J., Abay, K. A., Newman, C., Kosmowski, F., & Sida, T. (2024). *Revisiting the size – productivity relationship with imperfect measures of production and plot size*. *May 2021*, 595–619. https://doi.org/10.1111/ajae.12417

Barrett AT, McIntosh RA, Pritchard M, Hannan M, Alam Z, and M. M. (2013). *Asset Values : Why Are Some Households Doing Better Than Others ?* (Issue November).

Degefu, S. (2020). *Review on the Determinants of Agricultural Productivity and Rural Household Income in Ethiopia*. *11*(18), 1–11. https://doi.org/10.7176/JESD/11-18-01

Gcaba, K., Christian, M., Letsoalo, S., & Jagadeesh, D. (2025). *Factors Influencing the Willingness and Ability of Farmers to Adopt TELA Maize Seed in Alfred Nzo District , Eastern Cape , South Africa*. 1–14. https://doi.org/https:// doi.org/10.3390/su17030973

Hlatshwayo, S. I., Simon, M., Ngidi, C., Ojo, T. O., Modi, A. T., Mabhaudhi, T., & Slotow, R. (2023). The determinants of crop productivity and its e ect on food and nutrition security in rural communities of South Africa. *Frontiers in Sustainable Food Systems*.

Kadigi, I. L. (2025). *Improved Seeds and Rice Productivity in Tanzania : Policy Insights from National Sample Census Data*. https://doi.org/https://doi.org/10.21203/rs.3.rs-7280225/v1

Lobell, D. B., & Gourdji, S. M. (2012). The Influence of Climate Change on Global Crop Productivity. *American Society of Plant Biologists*, *160*(December), 1686–1697. https://doi.org/10.1104/pp.112.208298

Mapiye, O., & Dzama, K. (2024). Strengthening research-extension-farmer-input linkage system for sustainable smallholder livestock farming in Africa : progress and prospects. *Tropical Animal Health and Production*, *56*(363). https://doi.org/https://doi.org/10.1007/s11250-

024-04210-9

Nguyen, G., & Nguyen, T. T. (2020). Exposure to weather shocks: A comparison between self-reported record and extreme weather data. *Economic Analysis and Policy, 65*, 117–138. https://doi.org/DOI: 10.1016/j.eap.2019.11.009

Ortiz-bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G., & Lobell, D. B. (2021). Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, *11*(April). https://doi.org/10.1038/s41558-021-01000-1

Shokati, M., Sadegh, M., & Baghernejad, J. (2025). Environmental and Sustainability Indicators Factors affecting the livelihood sustainability of smallholder farmers in Iran. *Environmental and Sustainability Indicators*, *26*(October 2024), 100601. https://doi.org/10.1016/j.indic.2025.100601

Shrestha, N. (2020). Detecting Multicollinearity in Regression Analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42. https://doi.org/10.12691/ajams-8-2-1

Suri, T., & Udry, C. (2022). *Agricultural Technology in Africa*. *36*(1), 33–56.

The United Republic of Tanzania. (2013). *Tanzania in Figures 2012. National Bureau of Statistics*.

The United Republic of Tanzania. (2022). *Tanzania Panel Survey wave 5, 2020-2021. National Bureau of Statistics*. National Bureau of Statistics.

United Republic of Tanzania. (2023). *THE ECONOMIC SURVEY 2022 THE UNITED REPUBLIC OF TANZANIA Ministry of Finance*.

Wambede, N. M., Milliam, K., Denis, E., Andrew, M., Robert, T., & Geoffrey, M. (2025). *Determinants of Fruit Tree Adoption as a Climate Change Adaptation Strategy Amongst Smallholder Farmers in Lake Kyoga Basin : A Case Study of Budaka District , Eastern Uganda*. *2025*. https://doi.org/https://doi.org/10.1155/tswj/9642641

Zeleke, G., Teshome, M., & Ayele, L. (2024). *Determinants of Smallholder Farmers ' Decisions to Use Multiple Climate-Smart Agricultural Technologies in North Wello Zone ,*.

Zeleke, T., Beyene, F., Deressa, T., Yousuf, J., & Kebede, T. (2021). *Vulnerability of Smallholder Farmers to Climate Change-Induced Shocks in East Hararghe Zone , Ethiopia*. 1–19. https://doi.org/https:// doi.org/10.3390/su13042162 Academic

Zerssa, G., Feyssa, D., Kim, D., & Eichler-löbermann, B. (2021). *Challenges of Smallholder Farming in Ethiopia and Opportunities by Adopting Climate-Smart Agriculture*. 1–25.