Journal of Biology, Agriculture and Healthcare
ISSN 2224-3208 (Paper) ISSN 2225-093X (Online)
Vol.13, No.3, 2023

Impact of Wheat Cluster Farming Practice on Productivity among Smallholder Producers in Arsi Zone of Ethiopia

Getachew Nigussie Tefera (Corresponding author)
School of Agricultural Economics and Agribusiness, Haramaya University
PO box 253, Dire Dawa, Ethiopia
getachewnigussie@gmail.com

Mengistu Ketema Aredo
Ethiopian Economic Associations, Addis Ababa, Ethiopia
mengisuket@yahoo.com

Zekariyas Shumeta Nejo
Ethiopian Agricultura Research Institute, Addis Abeba, Ethiopia
zekishum@gmail.com

Kedir Jemal Ibrahim
School of Agricultural Economics and Agribusiness, Haramaya University
PO box 253, Dire Dawa, Ethiopia
Jemalkedir1@gmail.com

There is no external funding source forthis research.

Abstract
Wheat is one of the major cereal crops produced by smallholder farmers in Ethiopia. Despite efforts made to increase wheat production, the desire for wheat is growing rapidly in Ethiopia. Recently, to curb this problem, the government of Ethiopia has set up a cluster farming system for high-potential agricultural commodities like wheat as a means of improving productivity and ensuring food security of smallholder farmers. In light of the problems and the research gaps identified, this study seeks to address the impact of wheat cluster farming practices on wheat productivity among smallholder producers in the Arsi zone of Ethiopia. Data was collected from 383 sample wheat-producing households. The data was analyzed using descriptive statistics, endogenous switching regression (ESR), and propensity score matching (PSM) models. The results of the ESR indicated that wheat productivity was strongly and positively influenced by endogenous wheat cluster farming participation. The impact of cluster farming participation on cluster participants was about 13 qt of wheat yield increment, representing a 42% increase in participant productivity. The causal effect on non-participants is about 9 qt of wheat yield if they participate in cluster farming practices, representing a 26% increase in wheat productivity. The results of the PSM also reveal that participating in cluster farming significantly increases wheat productivity. As a result, stakeholders should pay close attention to promoting and scaling up wheat cluster farming practices in order to improve the productivity of smallholder farmers who can contribute to achieving food security in Ethiopia.

Keywords: Cluster farming, Productivity, Endogenous switching regression, Propensity score matching
DOI: 10.7176/JBAH/13-3-02
Publication date: February 28th 2023

1. Introduction

1.1. Background of the study

Agriculture contributes significantly to the Ethiopian economy. It is the main focus of the Ethiopian government's plan for the growth and development of the country's economy. It accounts for about 32.7% of the country's gross domestic product (GDP) (NBE, 2020). It also provided employment opportunities for about 65.6% of the total population in 2020 (ILO, 2021).

Cereal crop production is the dominant sub-sector in Ethiopian agriculture (Nicholas *et al.*, 2015). It creates about 60% of the rural job opportunities for the Ethiopian economy. It is also a source of more than 60% of the total calorie intake of the country's population (Workineh *et al.*, 2020). *Teff*, wheat, maize, and sorghum are the most important cereal crops cultivated in Ethiopia (NBE, 2020).

Wheat is the fourth most important cereal crop cultivated after *teff*, maize, and sorghum and the third in production after maize and *teff. About* 4.7 million farm households are directly dependent on wheat production (CSA, 2021). Ethiopia is one of the major wheat-producing countries in Africa, which accounted for about 20% of the nation's total wheat production in 2018 (USDA, 2019). More than 90% of Ethiopia's wheat production is grown mainly by smallholder farmers (Kaleb, 2017; World Bank, 2018). Ethiopia produces about 4.8 million metric tons of wheat, which was cultivated on 1.8 million hectares of land in 2020/21 (NBE, 2020). The central highland areas of Ethiopia, such as Arsi, West Arsi, Bale, and East Shewa zones, cover about 42% of Ethiopia's wheat production, with 1.89 million tonnes in 2018 (USDA, 2019).

Rapid urbanization and population growth greatly increase the demand for wheat products like wheat flour, bread, biscuits, pasta, macaroni, and spaghetti (Minot *et al.*, 2015). Even though Ethiopia is a potential wheat producer, a huge gap between production and consumption due to increasing demand for wheat products makes Ethiopia an importer of wheat. The Agricultural Transformation Agency (ATA) indicates that Ethiopia has more than 600 small and large flour mills with a total production capacity of 4.5 million tons of wheat flour per year (ATA, 2017). Domestic demand for wheat is estimated at 6.3 million tons. In order to close the gap between the demand and supply of wheat, the government has been continuously importing wheat from the Black Sea region for the last several years.

In Ethiopia, agriculture is largely characterized by small-scale subsistence farming and low productivity (Gebresilassie *et al.*, 2017). Wheat is an important staple food crop in Ethiopia, even though it is dominated by smallholder farmers. In Ethiopia, wheat production and productivity are relatively low by global standards (USDA, 2019).

Recently, to curb this problem, the government of Ethiopia has set up a cluster farming system for high-potential agricultural commodities, including wheat, as a means of poverty reduction and smallholders' income maximization. To deal with this, the Growth and Transformation Plan (GTP II) of Ethiopia intends to increase production and productivity of high-potential crops through the implementation of cluster farming practices (ATA, 2017).

Wheat cluster farming is a new farming practice applied by smallholder farmers in Ethiopia. Thus, no empirical research has been done to evaluate the impact of wheat cluster farming on the productivity of smallholder producers in the study area. Therefore, the aim of this study is to analyze the determinant factors that affect smallholder producers' wheat productivity and the impact of wheat cluster farming on smallholder producers' wheat productivity.

1.2. Overview of agricultural cluster farming practice

Cluster farming is a farming practice that is implemented as part of a complete farming package. It creates real profit by merging several smallholder farms into a solid entrepreneurial group of clusters that is capable of sharing both the benefits and the challenges (ATA, 2017).

Cluster farming is an agricultural commercial farming practice that is growing crops on adjacent farmland with the aim of increasing productivity. Increasing productivity means producing more output with the same amount of inputs or using fewer inputs to produce the same level of output. The farming practice improve productivity by using improved seeds at the same time, using fertilizers that are suitable for the same agro-ecology, benefiting from the same technical advisory support, and harvesting their crops with the same machinery (Louhichi *et al.*, 2019).

2. Materials and methods

2.1. Description of the study area

Arsi Zone, is found in the south-eastern highlands of the Oromia National Regional State of Ethiopia. It is located between 6 °45'N and 8° 58'N and 38° 32'E to 40° 50'E. According to the 2021 population projection, the total population of the Arsi zone is 3.71 million (CSA, 2021) (Fig 1).

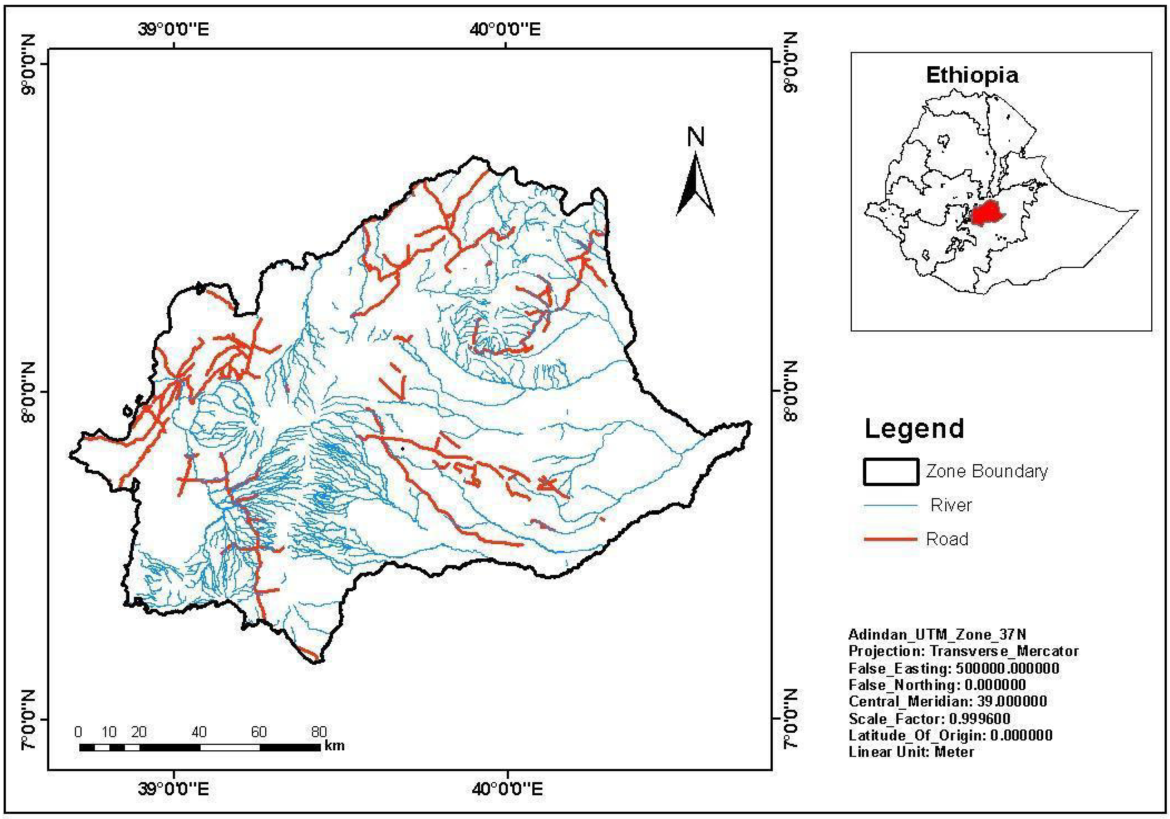

Figure 1. Study area map. Source: Authors 2021

According to USDA (2018), the Arsi zone is one of the major wheat producing areas in the south-eastern Ethiopian highlands, mainly known for its widespread wheat production and called the "wheat belt of Ethiopia." Wheat production comprises about 7.2 million quintals (41%) of the total annual cereal production of the Arsi zone through the engagement of 360,697 wheat producers. Arsi zone wheat production in 2020/21 constitutes about 22% and 12.5% of the Oromia region and Ethiopia's wheat production, respectively, which makes it the leading wheat-producing zone in Ethiopia. Wheat production accounts for about 39% of the total cereal cultivated area in the Arsi zone (CSA, 2021).

The Arsi zone was chosen by the Minister of Agriculture to undertake wheat cluster farming in Ethiopia due to its potential for wheat production. Wheat cluster farming has been practiced in Ethiopia in general and in the Arsi zone in particular since 2014. It is expected to play an important role in increasing wheat productivity in the country (ATA, 2017). Farms that are located near each other as a geographic cluster benefit from being able to utilize the same logistics, such as transportation, improved input supplies, machines, extension services, and a shared pool of skilled labor (Timothy and Randall, 2016).

2.2. Sources and methods of data collection

The multistage sampling method was employed to select samples. First, three districts from the Arsi zone were selected purposively. These districts were in areas where wheat cluster farming technology was better implemented. In the second stage, a total of nine *kebeles* were randomly selected from the three districts. The allocation of *kebeles* to districts was based on their number of wheat producers. In the third stage, respondents were selected. The total sample size is determined using a formula given by Kothari (2004) that affords the maximum sample size to ensure the anticipated precision. The formula given by Kothari (2004) as follows:

$$n = \frac{Z^2pqN}{e^2(N-1)+Z^2pq} = \frac{(1.96)^2(0.5)(0.5)(401{,}404)}{(0.05)^2(401{,}404)+(1.96)^2(0.5)(0.5)}$$

$= 385{,}508.4 / 1004.47 = 383.43$

Where: n is the anticipated sample size; Z is the standard cumulative distribution that corresponds to the level of confidence with the value of 1.96; e is the desired level of precision; p is the estimated proportion of an attribute present in the population with the value of 0.5 as recommended by Israel (1992) to get the desired minimum sample size of households at 95% confidence level and ±5% precision; q=1-p; and N is the size of the total population from which the sample is drawn. Thus, 383 sample households, out of which 134 are from cluster farming and 249 from non-cluster farming, were selected from nine *kebeles* using random sampling with probability proportional to size.

2.3. Data

The study included both primary and secondary data. Using a structured questionnaire and in-person interviews, the primary data was gathered. The information covered includes information on the production of wheat, wheat cluster farming techniques, the types of farming inputs used, the prices of inputs and outputs, sources of income, information on the ownership of assets such as land and livestock, and other institutional, socioeconomic, and demographic traits of households from both groups of wheat producers (cluster participants and non-participants). Secondary data were gathered through reports, office papers, journal articles, and other literature.

2.4. Method of data analysis

2.4.1. Endogenous switching regression estimation

This study uses an endogenous switching regression model to assess how wheat cluster farming affects smallholder wheat productivity. The switching regression was modeled in two stages (Kirtti and Phandindria, 2018; Gorst *et al*., 2018). The first is the selection model for wheat cluster farming participation, denoted by the binary variable. The latent variable is specified as:

$$Cluster^*_i = X_i\Omega + \mu_i \qquad (1)$$

$$Cluster_i = \begin{cases} 1, & \text{if } Cluster_i^* = X_i\,\Omega + \mu_i > 0 \\ 0, & \text{if } Cluster_i^* = X_i\,\Omega + \mu_i \leq 0 \end{cases}$$

Where: the wheat producer's household i chooses to participate in wheat cluster farming (A_i=1) in response to the benefit realized from the practice, if cluster* > 0, and 0 otherwise. X_i is nxm matrix of explanatory variables, W is m x1 vector of model parameters to be estimated, and m is n × 1 vector of normally distributed mean zero random error terms.

According to Ndeye *et al*. (2018), the second stage of the ESR is the outcome equation (wheat yield measured as quintal per ha) that splits the endogenous model into two regimes (participants and non-participants). The selection equation is specified as follows:

$$Wheat_{i1} = X_{i1}\,\beta_1 + \varepsilon_{i1} \quad \text{if Cluster} = 1, \qquad (2)$$

$$Wheat_{i0} = X_{i1}\,\beta_0 + \varepsilon_{i0} \quad \text{if Cluster} = 0, \qquad (3)$$

Where: X_i is a vector of institutional, technological, and socio-economic characteristics related to wheat producers. B_1 and β_0 are vectors of parameters. e_{i1} and e_{i0} are the error terms for participants and non-participants, respectively. In the switching regression model, the selection bias would manifest itself in the error terms ε and μ. As long as the unobserved variables are not captured by the explanatory variables, the error terms of the production and selection equation are correlated with corr (ε, μ) ≠ 0. The error terms μ_i, ε_{1i} and ε_{2i} follow a trivariate normal distribution with zero mean, and the covariance matrix is specified as:

$$Cov\,(\mu,\varepsilon_{1i}\,\varepsilon_{2i}) = \begin{bmatrix} \sigma^2_\mu & \sigma_{1\mu} & \sigma_{2\mu} \\ \sigma_{1\mu} & \sigma^2_1 & . \\ \sigma_{2\mu} & . & \sigma^2_2 \end{bmatrix} \qquad (4)$$

Where: σ^2_μ , σ^2_1 and σ^2_2 is representing the variance of the error terms in the selection equation and the two-production wheat 1 and 2 is respectively. $\sigma_{1\mu}$ and $\sigma_{2\mu}$ are the covariance of the selection equation error term (μ_i) and the wheat production regimes 1 (ε_{1i}) and 2 (ε_{2i}) is respectively. The dot (.) shows that the wheat 1 and 2 outcomes cannot be simultaneously observed for a farmer and hence the covariance is not present (Madalla, 1999). Equations (2-4) should be estimated in a way that accounts for the correlation between the error terms.

Using the ordinary least squares or the maximum likelihood estimation approach may yield inconsistent and biased estimates. A two-stage procedure can be used to estimate the parameters in Equations (2) – (4). The procedure involves estimating the probabilities of participating in wheat cluster farming using the maximum likelihood estimation approach. The estimated probabilities (cluster) are used to estimate selectivity terms that account for the endogeneity of wheat cluster farming participation. The selectivity terms denoted by 0 and 1 are expressed as:

$$\gamma_1 = \phi(Cluster_i)/\ \Phi(Cluster_i) \qquad (5)$$

$$\gamma_0 = -\ \phi(Cluster_i)/\ (-\ \Phi(Cluster_i)) \qquad (6)$$

Where: ϕ is the standard normal probability distribution and Φ is the standard normal cumulative distribution. The selectivity terms are included in Equations (2) and (3) as explanatory variables in addition to Xi. The modified equations are specified as:

$$Wheat_{i1} = X_{i1}\,\beta_1 + \gamma_1\omega_1 + \varepsilon_{i1} \quad \text{if Cluster} = 1, \qquad (7)$$

$Wheat_{i0} = X_{i0}\beta_0 + \gamma_0\omega_0 + \varepsilon_{i0}$ if Cluster = 0, (8)

The parameters in Equations (7) and (8) are computed using the OLS estimator. The two-stage procedure generates consistent coefficients, but the standard errors are inconsistent because they produce heteroskedastic residuals of the error terms. To obtain homoscedastic residual terms, the full information maximum likelihood estimation (FIMLE) method is more suitable (Emanuel and Victor, 2019; Ndye *et al.*, 2018; Kiritti and Phanindra, 2018). The participation and outcome equations (wheat productivity) are estimated simultaneously with the FIMLE. The log likelihood function of the FIML estimator is specified as:

$$\ln L = \sum(Cluster_i\theta_i\{\ln(\ln\beta(\gamma_{1i})\} + \ln\{\Omega(\tau_{1i}/\sigma_1)/\sigma_1)\}$$
$$+ (1-Cluster_i)\,\theta_i(\ln\{1-\Psi(\gamma_{0i})\} + \ln\{\Omega(\tau_{0i}/\sigma_0)/\sigma_0)\}\} \quad (9)$$

Where: $\Psi(.)$ is a cumulative normal distribution function, $\Omega(.)$ indicates a normal density distribution function, θ_i is an optional weight for observation i and λij are defined as:

$$\gamma_{ij} = \frac{x_i\Psi + \rho_j\tau_{ij}/\sigma_j}{\sqrt{1-\rho_j^2}} \quad j=0,1 \quad (10)$$

Where: $\rho_1 = \sigma^2_{\tau 1e}/\sigma_e\sigma_1$ is the correlation term between e_i and τ_1 and $\rho_0 = \sigma^2_{\tau 0e}/\sigma_e\sigma_0$ is the correlation term between e_i and τ_0. To ensure that ρ_1 and ρ_0 is bounded between -1 and 1 and that the estimated σ_1 and σ_0 is always positive, the FIMLE directly estimates $\ln\sigma_1$ $\ln\sigma_1$ and and atanhρ_j (Emanuel and Victor, 2019). atanhρ_j is further computed as:

$$\text{atanh}\rho_j = \frac{1}{2}\ln\left(\frac{1+\rho_j}{1-\rho_j}\right) \quad j = 0,1 \quad (11)$$

The estimation procedure yields an estimate of the following parameters of interest for each period t under consideration: the average treatment effect on the treated (ATE) is the average 'treatment' effect in the population (treatment = cluster participant). The main emphasis of this study is the ATE, as it estimates the expected impact of wheat cluster participation technology among the entire underlying population of wheat producers' households rather than the sub-population of wheat cluster farming participants only. Once the parameters of the models are estimated, the impact of wheat cluster participation, as denoted by the ATE, can be computed as:

$$ATE_{ESR} = E(Wheat_{i1}| Cluster_i = 1, X_{i1}) - E(Wheat_{i1}| Cluster_i=0, Xi_1)$$
$$= \sigma_1\rho_1\Phi(X_i\Psi)/\Omega(X_i\Psi) - \sigma_1\rho_1\Phi(X_i\Psi)/\{1-\Omega(X_i\Psi)\} \quad (12)$$

The validity of the ESR requires exclusionary restrictions that are correlated with participation but do not play a role in the wheat productivity of smallholders (Emanuel and Victor, 2019; Ndye *et al.*, 2018; Kiritti and Phanindra, 2018; Gorst *et al.*, 2015). Thus, the study used the information set of variables as selection instruments (selection variables).

Distance to the extension center and topography of wheat farm land were considered instrumental variables for this study. The researcher argues that distance to the extension center and topography of wheat farm land are critically important in determining participation in wheat cluster farming. Empirically, the validity of ESR instruments is tested.

The first test is running a probit model for participation with instruments and other variables. The second is the falsification test, which checks whether the instruments played an important role in production. As indicated by Kiritti and Phanindra (2018), this test indirectly checks whether instruments are correlated with the unobservable. The standard t-test is employed to determine if the difference is statistically different from zero.

2.4.2. Propensity score matching

The study also employed the PSM technique in order to check the robustness of the ESR findings. The PSM technique used in this work followed a process that includes propensity score estimation, common support region identification, best matching algorithm selection, testing of matching quality, the average treatment effect on the treated (ATT) calculation, and sensitivity analysis.

The PSM assesses the impact of cluster farming on wheat productivity by comparing the difference between cluster participants and non-participants using the ATT. The ATT measures the impact of wheat cluster farming participation on the wheat productivity of smallholders who actually participated in cluster farming rather than across all wheat producers who potentially could have participated in wheat cluster farming. ATT is calculated as:

$$ATT = E[Y(1)- Y(0)||G = 1] = E[Y(1)||G = 1]- E[Y(0)||G = 1] \quad (13)$$

Where Y is wheat productivity (potential outcome), G represents participation in cluster farming with a value equal to 1 if household participates and 0 otherwise and X represents the explanatory variables.

A sensitivity analysis was performed to determine whether or not the treatment effects findings were influenced by unobserved factors (hidden bias). The estimation of treatment effects with matching estimators is based on the selection of observable characteristics. However, a hidden bias might arise if there are unobservable variables that affect assignment to treatment and the outcome variable at the same time (Becker and Caliendo, 2007; Rosenbaum, 2002). Because the sensitivity analysis supports it, the robustness of the estimated intervention results will be included in this study, primarily to ensure whether the inference made about the impact of participating in wheat cluster farming, which has higher wheat productivity increment than non-participant households, is reliable or not.

3. Result and Discussion

3.1. Results of descriptive statistics

Smallholder wheat producers in the study area owned an average of 1.44 ha of wheat farm land per household, which is higher than the national average of 0.41 ha per household (CSA, 2021). Wheat cluster participants produce wheat over a larger area, about 1.5 ha on average, than non-participating households, which produced 1.41 ha. On wheat farms, there is no statistically significant difference in holding size between participants and non-participants.

The study area's wheat producers get income mainly from sales of wheat, sales of other crops, sales of livestock, and off/non-farm activities. However, sales of wheat constitute the major source of income for all groups of sampled wheat producer households in the study area. The mean gross income (the value of the total wheat yield at the current market price) of wheat for sampled households in the *meher* production season in the study area was ETB 176768.70.

3.2. Econometric results

3.2.1. Factors determining participation in wheat cluster farming

The probit model was used to determine the predicted probability values of cluster participants (the treatment group) and non-participants (control group) in wheat cluster farming. The presence of multicollinearity problems among the selected explanatory variables was tested with the Variance Inflation Factor (VIF) for continuous variables and the contingency coefficient (CC) for discrete variables. As indicated in Appendix Table 6, none of the explanatory variables were dropped from the estimated model since no serious problem was detected.

The probit regression result, given on Table 1, reveals that fourteen explanatory variables were hypothesized to determine wheat cluster participation. According to the computed coefficients, seven explanatory variables have a substantial effect on participation in wheat cluster farming at various probability levels. The outcome of the probit regression shows that the education level of the household head, access to credit services, extension contacts, off/non-farm income, and access to mechanization services do positively influence the decision to participate in wheat cluster farming. On the other hand, the age of the household head and wheat production experience have a negative role in households' decisions on wheat cluster farming participation.

Table 4. Estimation result of probit model on factors determining wheat cluster participation

Explanatory variables	Coefficients	Robust Std.err	Z
Age of household head	-0.015**	0.01	-1.49
Sex of household head	-0.43	0.54	-0.81
Education level	0.21**	0.1	2.11
Credit access	0.004***	0.0001	3.22
Extension contacts	0.079***	0.017	4.6
Livestock holding	0.039	0.045	0.87
Off/non farm income	0.03**	0.004	0.17
Farm size	0.08	0.008	1.01
Labor	0.007	0.023	0.32
Wheat experience	-0.076**	0.034	-2.2
Dependence ratio	-0.06	0.25	-0.24
Farmer to farmer extension	0.22	0.15	1.52
Oxen	0.147	0.146	1
Mechanization	0.44***	0.11	4.02
Constant	-3.68**	1.22	-3
Pseudo R2	0.27		
LR Chi-Squared	73.7		
Prob.>Chi-squared	0.000		
Sample size	383		

Source: Survey results.
Note: ***, ** and * represent significance at 1%, 5% and 10% probability levels, respectively.

3.2.2. Impact analysis of endogenous switching regression model

In order to determine whether wheat producer households benefited from joining wheat cluster farming, the study estimated the status of the two groups' wheat productivity. The study used distance to the extension center and topography of the wheat farm as selection instruments for identification. The identified instrumental variables are expected to fulfill the main conditions for being considered valid instruments. The instruments should not be directly related to the outcome variable of the study.

Instrumental variables of the study should directly affect participation in wheat cluster farming. For instance, if we take the distance to the extension center, it directly affects the initiation of participating in wheat cluster farming practice; it doesn't have a direct effect on the smallholder producer's wheat yield. Information gotten by wheat producers from extension experts is considered a critical component of participating in new farming practices. Smallholder producers with difficulty accessing information about new technology practices due to distance from the extension center will fail to participate in new farming practices, and vice versa. Nevertheless, it is tough to verify the validity of the instruments without undertaking proper statistical tests.

The validity of instrumental variables hypothesized by the study was tested by using different robustness checks. Probit regression was used to regress wheat cluster participation (the dependent variable in the selection equation) on all covariates, including instrumental variables. Accordingly, the effect of instruments is jointly significant at a 1% probability level (Appendix Table 7). As was expected, out of the sixteen variables entered into the probit model, the instrument variables, i.e., distance to the FTC and topography of the wheat farm, have a negative significant effect on wheat cluster farming participation at 1% and 5% significance levels, respectively. Other covariates, with the exception of a few variables (age, education, wheat experience, livestock, oxen, sex, farmer-to-farmer extension), significantly influenced the wheat cluster participation of smallholder wheat producers positively at different probability levels.

The second test is the falsification test, conducted by using ordinary least square (OLS) regression on the wheat yield (outcome equation) of non-participants with selection instruments and other covariates. The result of the test indicates that the instrument's joint effect on the non-participants is insignificant (Appendix Table 8).

Once instrumental variable testing is complete, it is possible to turn to the productive implications of the selection instrument (wheat cluster participation). The simplest approach to studying the impact of wheat cluster participation on wheat productivity is by using an OLS model. Wheat yield for the pooled sample was estimated using the OLS estimation technique by considering wheat cluster participation as an independent variable. As indicated in Table 2, wheat cluster farming participation has a positive and significant effect on wheat productivity.

The result of OLS estimates would lead us to conclude that there is a difference in the amount of wheat yield per hectare produced by wheat producers that participated in cluster farming with respect to the amount of wheat yield produced by wheat producers that did not participate in wheat cluster farming.

As a result of what OLS indicated, we conclude that if all other factors remain constant, wheat producer smallholder households that participated in wheat cluster farming can expect a 6% increase in wheat yield over their non-participant counterparts. Nevertheless, taking this result as an indicator of the impact of cluster farming practices on smallholders' wheat productivity is not appropriate since OLS considers participation in wheat cluster farming to be exogenously determined, while it is a potentially endogenous variable. The OLS model's estimated results are biased and inconsistent because it fails to account for the problems of selection bias and unobservable heterogeneity. Besides, OLS estimates do not explicitly account for potential structural differences in wheat productivity between wheat cluster participants and non-participants. As a result, the ESR model for smallholders' wheat productivity among cluster participants and non-participants was estimated.

The results of the ESR model for the estimated selection and wheat productivity are presented in Table 2. The correlation coefficient (rho1=0.17) between participation in wheat cluster farming practice and cluster participant outcome was statistically significant for the outcome variable of wheat productivity at the 1% significance level. Since ρ is positive and significantly different from zero, individuals who participate in the wheat cluster would have a higher wheat yield than a random individual from the sample. The significance of the coefficient of correlation between the participation equation and the wheat productivity also indicates that self-selection occurred in the wheat cluster farming participation decision.

On the other hand, the non-significance of the covariance estimates for the non-participants indicates that without participation in the wheat cluster farming, there would be no evident difference in the average wheat income between non-participants and a random household caused by unobserved factors. This confirms that the endogenous switching model is an appropriate one for controlling for self-selection and inherent differences between cluster participants and non-participants. The result is consistent with the studies of Tezera *et al*. (2020) and Musa *et al*. (2017) but inconsistent with the studies of Temitope *et al*. (2021), Tsega *et al*. (2019), and Emmanuel and Victor (2019).

The Wald test of independence (χ^2 (2) = 22.31) is statistically significant at the 1% significance level. The significant value for the Wald tests confirms the joint significance of the error correlation coefficients in the selection and outcome equations, providing further evidence of endogeneity. The result indicates that the hypothesized independent variables studied in the ESR model jointly affect the participation in wheat cluster farming and its impact on wheat productivity of cluster participants and non-participants. It also indicates the existence of selection bias and slope heterogeneity between cluster participants and non-participants. There are also some independent variables that affect cluster participants and non-participants differently. Hence, it requires the estimation of two separate wheat yield functions.

Table 2. Determinants of wheat productivity (wheat yield in qt)

Variables	OLS	FIML Endogenous Switching Regression		
		Participants	Non-participants	Selection
Cluster participation	5.91***(1.2)			
Age	-0.15(0.08)	-0.018(0.11)	-0.22(0.12)	0.43(.69)
Education	-0.004(0.004)	-0.01(0.01)	0.004(0.01)	-0.11**(0.06)
Experience	0.03(0.04)	-0.004(0.05)	0.05(0.06)	-0.49(0.36)
Livestock	-0.003(0.004)	-0.0005(0.007)	-0.007(0.005)	0.022(0.03)
Oxen	0.01(0.01)	0.012(0.02)	0.007(.013)	0.09(0.08)
Mechanization	0.01**(0.008)	0.055**(0.029)	-0.0001(.01)	0.26***(.06)
Labour	0.09***(0.01)	0.085***(0.024)	0.093***(0.022)	0.13(0.11)
Extension contacts	0.03**(0.005)	0.08**(0.034)	-0.001(0.03)	0.98***(0.12)
Credit	0.05***(0.02)	0.114***(0.035)	0.004(0.024)	0.43***(0.15)
Off/non-farm income	0.07***(0.05)	0.092***(0.024)	0.06***(0.005)	0.16*(0.09)
Improved seed	0.06***(0.02)	0.114***(0.043)	0.018(0.024)	
Urea	0.02**(0.01)	0.05**(0.027)	0.005(0.014)	
NPS	0.002***(0.07)	0.002***(0.002)	0.001***(0.0001)	
Row planting	0.06***(0.02)	0.086**(0.042)	0.076**(0.03)	
Weeding method	0.013**(0.006)	0.002(.0001)	0.017(0.007)	
Rust treatment	0.006(0.01)	0.01(.001)	0.017(0.015)	
Sex				0.004(2.3)
Farm size				0.04(0.08)
Dependence				0.11(0.08)
Farmer-farmer extn				0.05(0.07)
Constant	2.61***(0.23)	1.74***(0.48)	3.1***(0.32)	6.49***(2)
Distance to FTC				-0.26**(0.12)
Topography				-0.2***(0.18)
Sigma		1.74***(0.07)	1.68***(0.09)	
Rho		0.17***(0.19)	-0.44(0.01)	
Observations	383	134	249	383
R-square	0.873			
F (17, 365)	147.56			
Log likelihood		-60.94		
Wald χ^2 (16)		1231.2***		
Wald test of independent equations χ^2 (2)		22.31***		

Source: Survey results.

Note: ***, ** and * represent significance at 1%, 5% and 10% probability levels, respectively. The number in parenthesis shows robust standard errors.

The constant terms for the participation in cluster farming show a significant positive effect, which shows that wheat producers in the study area are cooperating to participate in new farming practices. The result justifies the relevance of using the FIMLE from the second stage of ESR for wheat productivity. The presence of heteroskedasticity in the model for both outcome variables was controlled using robust standard errors based on White (1982). Robust specifies that the Huber/White/sandwich estimator of variance be used in place of the conventional Maximum Likelihood Estimates variance estimator.

The Full Information Maximum Likelihood of the endogenous switching regression model estimate identifies the determinants of wheat productivity for cluster participants and non-participants. As was expected, out of the fourteen variables entered into the FIML model, extension contact, mechanization use, labor, credit use, off/non-farm income, improved seed, urea application rate, NPS quantity, row planting, and weeding method were identified as the significant determinants of the wheat productivity of cluster participants and non-participants.

The estimates for the ATT, the average treatment effects on the untreated (ATU), and the heterogeneity effect (HE), which show the impact of wheat cluster farming participation on wheat productivity and also the effect due to their integral features on wheat yield, are shown in Table 3. The FIML estimates of ATT and ATU account for selection bias arising from the fact that cluster participants and non-participants may be systematically different.

Table 3 presents the expected amount of wheat yield produced per hectare under actual and counterfactual conditions. Cells (a) and (b) denote the expected amount of wheat yield per hectare observed in the sample. The expected amount of wheat yield per hectare by wheat producers that participated in wheat cluster farming is about 44 qt, while it is about 33 qt for the group of wheat producers that did not participate. Based on this simple comparison, on average, the wheat producers that participated in cluster farming produced about 11 kg (33%) more than the wheat producers that did not participate. A flawed conclusion could result from using a direct difference between the anticipated mean outcomes of wheat cluster participants and non-participants.

Table 3. Impact of wheat cluster farming practice on wheat productivity (Yield/ha)

Participation status	Participants	Non-participants	ATE
Producers that participate (ATT)	(a) 43.85	(c) 30.9	12.95***
Producers that did not participate (ATU)	(d) 41.92	(b) 33.2	8.72***
Heterogenous effect (HE)	$BH_1 = 1.93$	$BH_2 = -2.3$	TH = 4.23

Source: Survey results

Note: *** is statistical significance at 1% significance level.

Predicted wheat productivity for participant households compared to what it would have been if they had not been participants (i.e., non-participants) could provide a more accurate comparison of the difference in wheat productivity between participants and non-participants. Similarly, non-participant households' wheat productivity was compared to an outcome variable if they had participated in wheat cluster farming. The model's findings showed that wheat cluster farming significantly increases participants' wheat yields while also having the potential to raise productivity among non-participants.

The causal effect of cluster farming for participants is about 13 qt of wheat yield, representing a 42% increase in the wheat yield of cluster farming participants. In other words, the treatment effects of cluster participation on wheat yield revealed that households of wheat producers who participated in the wheat cluster farming practice would have had a yield loss of about 13 qt had they not participated. The causal effect of cluster farming practice on non-participants is about 9 qt of wheat yield if they participate in cluster farming practice, representing a 26% increase in wheat yield.

The base heterogeneity (BH_1, the difference between cells a and d) for wheat productivity indicates that if the current non-cluster participants had participated, they would have gained less wheat yield (1.93 qt) than wheat producers that participated. On the other hand, the base heterogeneity (BH_2, the difference between cells b and c) shows that participants would have a lower wheat yield of 2.3 qt than actual non-participants even if they had not participated. Overall, the transitional heterogeneity (TH, the difference between BH_1 and BH_2) effect is positive, implying that the impact of cluster farming participation had a positive effect of 4.23 qt on the household's wheat productivity. The result is consistent with the studies of Wogayehu and Shery (2021), Tezera *et al*. (2020), Adane *et al*. (2019), Moti *et al.* (2018), and Tesfaye *et al*. (2018), which found that participation in new farming practices had a significant and positive effect on wheat productivity.

3.2.3. Impact analysis of propensity score matching

Estimating the propensity scores is the first step in the PSM approach. Propensity scores are estimated for both cluster participants and non-participants in the study. As indicated in Table 4, the estimated propensity score for wheat cluster participants had a minimum of 0.0979355 and a maximum of 0.9133318, with a mean propensity score of 0.4700. Whereas the estimated propensity score for non-participants was a minimum of 0.0226726 and a maximum of 0.9029357, with a mean propensity score of 0.22852. Accordingly, based on comparing the minima and maxima of the propensity score of participants and non-participants, the minimum propensity score of non-participants was 0.0226726 and the maximum propensity score of participants was 0.9133318. Thus, the common support region of the data lies between 0.0226726 and 0.9133318. Consequently, households with a propensity score of less than the minimum (0.0226726) and larger than the maximum (0.9133318) are off-support and not considered for matching and estimation of the average treatment effect.

Table 4. Summary of estimated propensity score of households

Propensity score	Observation	Mean	Std. Dev.	Minimum	Maximum
Cluster participants	131	0.47	0.1928	0.0979355	0.913332
Cluster non-participants	249	0.2852	0.1853	0.0226726	0.902936
Total households	383	0.3498	0.2074	0.0226726	0.913332

Source: Survey result

Figure 2 also depicts the distribution of households with respect to the estimated propensity scores and the common support region. The upper halves of the histogram show the propensity score distribution of wheat cluster farming participant households, while the bottom halves show the propensity score distribution of non-participants. The yellow color in the bars indicates the propensity score of cluster participants (treated as wheat producers) that are outside the common support region. Similar to this, the blue color of the bars represents the propensity score of non-participants (untreated wheat producers) who are outside the common support zone.

Figure 2. Propensity scores distribution and common support region
Source: Survey data, plotted by psgraph

The appropriate matching algorithm which balances the most explanatory variables, results in a low pseudo-R^2 value, reduces more standard bias, and also results in large matched sample size should be selected by using such performance criteria. Accordingly, the kernel matching algorithm (bandwidth of 0.25), was found appropriate matching algorithm for the data presented for this study. The kernel (bandwidth 0.25) algorithm estimate identified a total of 346 observations. Of these, 214 are from cluster participants and 132 are from the non-participants.

After ensuring the common support region and verifying the matching quality of the PSM model, the study then analyzed the impact of the wheat cluster farming on wheat productivity, which is the mean production of wheat in quintal per hectare, using the selected algorithm technique. Wheat cluster farming was evaluated as one of the improved farming practices to see if there is an increase in wheat productivity as a result of cluster farming.

As indicated in Table 5, before matching, the mean wheat productivity was significantly higher for cluster participants than for non-participants. The mean wheat productivity reported by participants was 42.85 qt, while that for non-participants was 36.04 qt, which was 18.5% lower than that of participants. However, after controlling for other factors using the PSM method, the mean wheat productivity of participants is 43.01 qt, while that of non-participants is 32.8 qt. As a result of participating in wheat cluster farming, participants' productivity increased by an average of 10.21 qt, or 31%, more than that of non-participants.

It is anticipated that cluster participants' wheat productivity will rise significantly as a result of the assistance and access to agricultural packages provided to them, such as better wheat seed, mechanization services, and extension services. The findings of the studies by Adane *et al*. (2019), Tesfaye *et al*. (2018), and Tesfaye *et al*. (2016) are in line with the results of this study. According to the studies, farmers' productivity with regard to producing wheat and other crops increased significantly and favorably as a result of the implementation of new farming practices.

Table 5. PSM based wheat productivity impact of participation in cluster farming

Outcome Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Wheat yield	Unmatched	42.85	36.04	6.80	0.70	9.59 ***
(qt/ha)	ATT	43.01	32.80	10.21	0.84	7.3***

Source: Survey data
Note: *** represent significance at 1% probability level.

The study employed rbounds to test sensitivity. As shown in Appendix Table 9, if the matched pairs of participants and non-participants in wheat cluster farming were allowed to differ in odds of participating in wheat cluster farming by a factor of 2.5 (150%) in unobserved characteristics, the impact of participating in wheat cluster farming on wheat productivity is still significant at the 10% level of significance. This indicates the estimated impacts are insensitive to hidden bias. According to Hui *et al*. (2020), if the results of sensitivity analysis are significant until the value of gamma (Γ) is close to 2, it can be considered that the empirical result of propensity score matching is insensitive to unobserved bias. Based on this, the results have a good justification to ensure the average treatment effects on subjects treated using PSM were insensitive to unobserved bias. This revealed that the study included important covariates that affected participation in wheat cluster farming and outcome variables, and the estimated ATT wheat productivity from wheat production was insensitive to unobserved covariates.

4. Conclusions and recommendations

The study's average treatment effect result indicates that wheat cluster farming had a positive impact on wheat productivity. The wheat productivity of cluster participants' producers was higher than that of non-participants. Even though the wheat yield of cluster participants' wheat producers in the study area was higher than the national average, the current yield was not up to the area's potential. The reason was that wheat producers were not employed and used full agricultural packages as per the recommendations. Thus, stakeholders should have to develop strategies to help the farmers get full agricultural packages that improve wheat productivity at reasonable prices. As a result, smallholder farmers will use the extra wheat to improve their food security and standard of living.

Table 6. Multicollinearity test on factors affecting cluster participation

Variable	VIF	1/VIF
Wheat Experience	5.23	0.19
Age of household head	5.16	0.19
Wheat farm size	1.55	0.64
Labor	1.31	0.76
Extension contacts	1.17	0.85
Dependence ratio	1.07	0.93
Livestock holding	1.07	0.93
Education status	1.07	0.93
Oxen holding	1.06	0.94
Mean VIF	1.73	

Source: Survey result

Table 7. Probit model estimation result articulating the effect of instruments on wheat cluster participation

Explanatory variables	Coefficient (Robust St. Err)
Age of household head	0.68(0.65)
Education level of household head	-0.09(0.06)
Wheat experience	-0.045(0.33)
Livestock (TLU)	0.039(0.035)
Oxen	0.10(0.09)
Mechanization access	0.22(0.08) ***
Labour in man days	1.94(.37) ***
extension contact	0.93(.14) ***
Credit access	0.32(0.16) *
Off/non-farm income	0.15(0.06) **
Sex of household head	-0.31(0.32)
Wheat farm size	0.74(0.65) ***
Dependence	0.36(0.21) *
Farmer to farmer extension	0.03 (0.02)
Distance to FTC*	-0.38(0.11) ***
Topography of wheat land*	-0.18(0.10) **
Constant	6.04(2.71) ***
Wald test χ^2 (16)	162.13***
Log likelihood	-166.87
Observation	383

Source: Survey results.

Note: "*" indicates instruments; statistical significance at ***, ** and * significance at 1%, 5% and 10% significance levels, respectively.

Table 8. OLS regression of cluster non-participants in wheat productivity

Variables	Coefficient (Robust St. Err)
Age of household	-5.62(3.33)
Education level	-0.029(0.23)
Wheat experience	1.27(1.70)
Livestock	-0.19(0.13)
Oxen	-0.21(0.35)
Mechanization	-0.053(0.24)
Labour	0.12(0.51)
Extension contacts	0.49(0.53)
Credit use	0.48(0.65)
Off/non-farm income	-0.21(0.17)
Improved seed	1.34**(0.65)
Urea	0.24(0.29)
NPS	0.02***(0.003)
Row planting	0.01***(0.02)
Weeding method	0.08**(0.2)
Rust treatment	-0.17(0.41)
Topography	-0.44(0.401)
Distance to FTC	-2.88(0.41)
Constant	83.6(9.5)
Observations	249
R-square	0.38
F-stat	1.78

Source: Survey results.
Note: ***, ** and * represent significance at 1%, 5% and 10% probability levels, respectively.

Table 9. Result of sensitivity analysis using Rosenbaum bounding approach

Gamma e^g =	Productivity
1	P<0.002
1.25	P<0.000
1.5	0
1.75	1.70E-06
2	1.40E-07
2.25	1.20E-08
2.5	9.90E-10
2.75	8.20E-11
3	6.80E-12

Source: Survey result
Note: e^g (Gamma) = log odds of differential due to unobserved factors where Wilcoxon significance level for each significant outcome variable is calculated

Appendix
See Tables 6, 7, 8 and 9.

References

Adane Tufa, Arega Alene, Manda, J., Akinwale, M.G., Chikoye,D., Shiferaw Feleke, Tesfamicheal Wossen, and Manyong, V. 2019. The productivity and income effects of adoption of improved soybean varieties and agronomic practices in Malawi. *World Development*, 124:104631.

ATA. 2017. Agricultural Transformation Agency. *Transforming Agriculture in Ethiopia 2016/17 Annual Report*, Addis Ababa, Ethiopia.

CSA. 2021. Central Statistical Agency the Federal Democratic Republic of Ethiopia. Agricultural Sample Survey 2020/21 (2013 E.C.), Report on Area and Production of Major Crops (Private Peasant Holdings, *Meher* Season). Statistical Bulletin 590, Addis Ababa, Ethiopia.

Emmanuel, D. and Victor, O. 2019. Mineral fertilizer participation and land productivity: implications for securing stable rice production in Northern Ghana. *Land Journal*, 8, 59; doi:10.3390/land8040059.

Gorst, A., Dehlavi, A. and Groom, B. 2018. Crop productivity and adaptation to climate change in Pakistan. *Environment and Development Economics*, 23 (6): 679–701.

Gorst, A., Groom, B. and Dehlavi, A. 2015. Crop productivity and adaptation to climate change in Pakistan. London: Center for Climate Change, Economics and Policy, ESRC Research Center.

ILO. 2021. International labour organization, ILOSTAT data base.

Kaleb Kelemu. 2017. Determinants of farmers' access to information about improved wheat varieties: case of farmers in major wheat growing regions of Ethiopia. *International Journal of Research in Agricultural Sciences*, Vol 4 (1).

Kirtti, R. P. and Phanindra, G. 2018. Impact of farmer education on farm productivity under varying technologies: case of paddy growers in India. Paltasingh and Goyari. *Agricultural and Food Economics*, vol 6(7).

Kothari, C. R. 2004. *Research Methodology*: *Methods and Techniques*, 2nd Edition. New age International, New Delhi, India.

Louhichi, K., Temursho, U., Colen, L. Gomez, Y. and Paloma, S. 2019. Up scaling the productivity performance of the agricultural commercialization cluster initiative in Ethiopia, EUR 29950 EN, publications office of the European Union, Luxembourg, ISBN 978-92-76-12941-7, doi:10.2760/57450, JRC117562.

Maddala, G. S. 1999. *Limited Dependent Variable in Econometrics*. Cambridge University Press, New York.

Moti Jaleta, Menale Kassie, Paswel, M., Chilot Yirga and Olaf, E. 2018. Impact of Improved Maize Adoption on Household Food Security of Maize Producing Smallholder Farmers in Ethiopia. *Food Security*.

NBE. 2020. National Bank of Ethiopia Annual report 2020/21. Addis Ababa, Ethiopia.

Ndèye, F., Abdoulaye, D., Kimseyinga, S., and Djiby, D. 2018. Impact of participation of improved pearl millet varieties on productivity in central Senegal. *Journal of Agriculture and Environmental Sciences*, Vol. 7, No. 2, pp. 90-100.

Nicholas, M., James, W., Solomon Lemma, Leuleseged Kasa, Gashaw Abate, Shahidur, R. 2015. The wheat supply chain in Ethiopia: patterns, trends, and policy options. Research for Ethiopia's Agriculture Policy (REAP) Project working paper, Washington, DC.

Tesfaye Haregewoin, Baye Belay, Eyob Bezabeh, Kaleb Kelemu, Daniel Hailu and Fitsum Daniel. 2018. Impact of improved wheat variety on productivity in Oromia regional state, Ethiopia. *Greener Journal of Agricultural Sciences,* ISSN: 2276-7770 Vol. 8 (4), pp. 074-081.

Tesfaye Solomon, Bedada Begna and Mesay Yami. 2016. Impact of improved wheat technology participation on productivity and income in Ethiopia. *African Crop Science Journal*, Vol. 24(1).

Timothy, P. N. and Randall, F. 2016. *Economic contributions of the wheat cluster to the Washington economy*. Washington State University, America.

Tezera Woldemeskel, Mengistu Ketema, Jema Haji and Lemma Zemedu. 2020. Welfare Impact of Moringa Market Participation in Southern Ethiopia. *Sustainable Agriculture Research*; Vol. 9, No. 3; 2020 ISSN 1927-050X E-ISSN 1927-0518

USDA. 2019. United States Department of Agriculture. Ethiopia - *Grain and Feed Annual.* Global Agricultural Information Network (GAIN), Report Number ET1903.

Wogayehu Fitawek, and Shery, L. H. 2021. Evaluating the Impact of Large-Scale Agricultural Investments on Household Food Security Using an Endogenous Switching Regression Model. *Land*, 10, 323.

Wogderes Ejigu. 2020. Contributions of cluster farming to the livelihood of farmers, a comparative study at Siyadebrina Wayu district, North Shewa, Ethiopia. MSc Thesis, Center for Environment and Sustainable Development, Addis Ababa, Ethiopia.

Workineh Ayenew, Tayech Lakew and Ehite Hailekristos. 2020. Agricultural technology participation and its impact on smallholder farmer's welfare in Ethiopia. *African Journal of Agricultural Research*, ISSN: 1991-637. Vol 15(3), pp.431- 445.