Determinants of Production of Maize in Yeki Woreda, Sheka Zone, Southwest Ethiopia

Lema Abate (*Department of statistics, Mizan-Tepi University, Tepi, Ethiopia*)

Email address: *lemamtu2007@gmail.com*

Abstract: Agriculture is the sole source of livelihood for the majority of Ethiopian. Ethiopia endowed with huge potential for agricultural development and cereal crops like maize are widely cultivated across a range of environmental conditions. The general objective of this study was to examine the determinants of production of maize in Yeki Woreda, Sheka zone Southwest Ethiopia. The data for this study obtained from Yeki Woreda agricultural office. Different statistical methodologies used to fit collected data like descriptive statistics, Multiple Linear regression model with dummy variables and Analysis of variance (ANOVA). The goodness of fit test like coefficient of determination and adjusted coefficient of determination used to check the effectiveness of the model. In addition, the assumptions of linear regression model and assumption of ANOVA checked. The results of this study revealed that DAP, Urea, farm size in hectare and number of participant had statistical significant relationship with the production of maize. Giving attention to those factors may increase the production of maize. Therefore, all concerned body should give due emphasis to the identified factors to increase maize production.

Key terms: Maize production, Multiple Linear Regression, Dummy variables

1. INTRODUCTION

Maize is annual plant with high yield and enjoys exceptional geographic adaptability, with an important property, which has helped its cultivation to spread throughout the world. The gradual expansion in the America by the natives propagated rapidly in the 16th century following the return of Columbus to Europe. Colonial conquests and trade played a central role in the spread of maize cultivation well beyond the European continent, to Africa and Far East Asia (Abdolreza A., 2006). Maize is the largest and most productive crop in Ethiopia. It was arrived in Ethiopia slightly later, around the late 17th century (Huffnagel H. P., 1961). The maize yield was 4.2 million tons, 40% greater than teff, 56% greater than sorghum, and 75% greater than wheat yield in 2007/08. With an average yield of 1.74 tons per hectare or equal to 3.2 million tons grown over 1.8 million hectares from 1995 to 2008, maize has been the leading cereal crop in Ethiopia from 1990 in terms of crop production. Wheat and sorghum yields have averaged 1.39 and 1.36 tons per hectare, respectively (CSA, 2009).

As reported by Dawit A. et al (2014) Maize is the most important cereal crop both in terms of level of production and area coverage in Ethiopia. About 9 million farmers, i.e., 70% of the total farmers produced about 6 million tons of maize over two million hectares of land. The farmers grow maize mostly for subsistence, with 75 % of all maize produced are consume by the farming households. Maize is an important crop for overall food security and the maize growing environments are classified into four broad agro-ecologies; specifically, mid altitude sub-humid,

mid altitude moisture stress area, high altitude sub-humid and low altitude sub-humid agro ecologies. The main fertilizer types that are used by farmers in Ethiopia are Urea as a source of nitrogen (N) and DAP as a source of Phosphorus (P) (Dawit A. C. Y., 2014).

An average of 69% of all maize grown in 2013 in Ethiopia received some amount of mineral fertilizer application, compared to 56% in 2004. There were appreciable differences in the maize area receiving fertilizer application among the regions. According to study conducted in Ethiopia 92% of the area planted to maize in Tigray and 85% in Amhara received fertilizer in 2013 whereas Oromia, SNNP and Benishangul Gumuz showed lesser area coverage of 67%, 61% and 41%, respectively. As stated earlier, the overall fertilizer application on maize in Ethiopia has shown significant growth over the last decade. The consumption rate grew at more than 12% per annual between 2004 and 2013, in comparison to the sub-Saharan Africa (SSA) average of 3.8% (between 2004 and 2012). Ethiopia has one of the fastest growth rates of fertilizer usage in Sub-Saharan Africa SSA (Tsedeke A. et al., 2015).

Several researches conducted in these areas by applying different approaches to identify factors that determine the production of maize in different countries. Urea and DAP appeared to be the major underlying determinants of maize output. It was establish from a stochastic frontier model that maize yield was positively influence by labor, DAP and Urea fertilizers and farm size (Bealu T., 2013). This study was intended to examine determinants of maize production in Yeki Woreda, Sheka Zone, Southwest Ethiopia. Multiple linear regression model is a model in which we can fit independent variables with dependent variable in the linear form. A multiple linear regression model with fertilizer use, farm size in hectare and a dummy variable designed to capture the effects of the changes induced by fertilizer liberalization measure, as explanatory variables. Maize output was the dependent variable (Ammani A., 2010).

2. Methods and Materials

2.1. Study Area and study design

This study conducted in Sheka Zone, Yeki Woreda, around 611km away from Addis-Ababa Capital city of Ethiopia. Sheka Zone is one of Zones in South Nation Nationalities and Peoples Regional State (SNNPR) in Ethiopia. There are three-woreda and two-city administration in Sheka Zone namely Anderacha, Masha, Yeki, Tepi city administration and Mash city administration. The major town of Yeki Woreda is Tepi and Yeki is bordered on the south by the Bench Sheko Zone, on the west by the Gambela Region, on the north by Anderacha and on the east by the Keffa Zone (SNNPR Livelihood Woreda Report, 2009). Based on the 2007 Census conducted by the Central Statistical Agency, Tepi town has an estimated total population of 19,231 of whom 10,113 were males and 9,118 were females. The retrospective study conducted by taking secondary data from Yeki Woreda Agricultural office.
The dependent variable in this study was the amount of maize production (in quintal) each year and Independent variables included in this study were DAP, Urea, farm size (hectare) and number of participant

2.2. Statistical Analysis Methods

In any applied setting, a statistical analysis should begin with a thoughtful and thorough description of the data. In this study, descriptive statistics like mean and standard deviation used to describe the observed data and, in addition, from inferential statistics multiple linear regression model and analysis of variance (ANOVA) used to identify the determinants of maize production in Yeki Woreda. Data analyzed by SPSS and R 3.6.1 statistical software's.

2.2.1. Regression Analysis

Regression analysis gives information on the relationship between a response (dependent) variable and one or more (predictor) independent variables to the extent that information is contained in the data. The value of each predictor variable can assessed through statistical tests on the estimated coefficients of the predictor variables.

2.2.2. Multiple linear regression model

Multiple linear regression model is the model, which consists of two or more than two independent variables. The model written as follows:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon_{ki} \qquad (1)$$

Where: $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ are regression coefficients (unknown model parameters) and ε is the error due to variability in the observed responses.

$$X_1 = DAP, X_2 = UREA, X_3 = farm\ size,\ X_4 = number\ of\ participant$$

2.2.3. Dummy variable

A dummy variable or Indicator Variable is an artificial variable created to represent an attribute with two or more distinct categories/levels in the linear regression model (Smita S., 2009). In this study, the dummy variables like DAP and Urea used to fit the regression model. Therefore, the linear regression model with the dummy variable given as:

$$y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \qquad (2)$$

This can be interpreted as an intercept shift

If $x_{i1} = 0$, then $y = \beta_0 + \beta_2 x_{i2} + \beta_3 x_3 + \beta_4 x_4$

If $x_{i1} = 1$, then $y = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_3 + \beta_4 x_4$

If $x_{i2} = 0$, then $y = \beta_0 + \beta_1 x_{i1} + \beta_3 x_3 + \beta_4 x_4$

If $x_{i2} = 1$, then$y = \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \beta_1 x_3 + \beta_1 x_4$

where: $x_{i1}\ and\ x_{i2}$ are dummy variables (binary) so that $i = 1$ if the farmer use DAP and $i = 0$ otherwise, and $i = 1$ if the farmer use Urea and $i = 0$ otherwise.

Coefficient $\beta_1$for x_{i1} and $\beta_2$for x_{i2} are interpreted as the *increase or decrease* in the mean when comparing farmers maize production in quintal who use DAP or UREA and farmers who do not use DAP or UREA (yes vs no)

Assumptions of the linear regression model

- ✓ The relationship between response variable and predictors are linear.
- ✓ The explanatory variable and error term are independent.
- ✓ The error term follows normal distribution with mean (μ) zero and variance (σ^2).
- ✓ The variance of error term is constant.

2.2.4. Analysis of variance (ANOVA)

ANOVA is the statistical method used to test the equality of two or more coefficient of estimates. In this paper, analysis of variance (ANOVA) used to test the equality of coefficient of parameters.

The statistical model for ANOVA written as follows:

$$y_{ij} = \mu + \beta_i + \varepsilon_{ij} \tag{3}$$

Where:

y_{ij}: the jth observation

μ: the overall mean

β_i: the ith regression coefficient or parameter

ε_{ij}: the error term for the jth observation

Model Assumption of ANOVA

- Within classes the observation y_{ij} all assume normally and independent.
- The error assumed to be normal and independently distributed with mean zero and variance σ^2. $\varepsilon_{ij} \sim N(0, \sigma^2)$
- Homogeneity of the variance.

2.2.5. Overall significance test of the model

ANOVA table used to check an overall significance of the model.

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

$$SSReg = \frac{\sum_{i=1}^{n} X_i Y_i}{\sigma_x^2}$$

$$SSRes = SST - SSReg$$

Table 1: Analysis of variance (ANOVA) Table

Source of variation	Sum square(ss)	df	Mean square(MS)	F_{cal}
Regression	Sum square Regression	k-1	Mean Square Regression (MSReg)	$\frac{MSReg}{MSRes}$
Residual	Sum square Residual	n-k	Mean square Residual (MSRes)	
Total	Sum square Total	n-1		

df: degree freedom, F_{cal}: F calculated value, k:number of parameter, n: number of observations

2.2.6. Coefficient of determination

The coefficient of determination (R^2) is the ratio of sum square regression to sum square total and, used to indicate the degree to which the model explains the observed variation in the dependent variable, relative to the mean.

$$R^2 = \frac{\text{Sum square regression}}{\text{sum square total}}$$

$$R^2 = \frac{SSReg}{SST} \tag{4}$$

The R^2 always lies between 0 and 1, where the higher R^2 indicates a better model fit. When interpreting the R^2, higher values indicate that more of the variation in y explained by variation in x, and therefore; the Sum square Residual is low relative to the Sum square Regression.

2.2.7. Adjusted coefficient of determination

An adjusted coefficient of determination (R^{-2}) used to test whether the model explains the observed variation strongly or not. To avoid a bias towards complex models, the adjusted R^2 used to select regression models. The adjusted R^2 only increases if the addition of another independent variable explains a substantial amount of variance.

$$R^{-2} = 1 - \frac{\frac{\sum_{i=1}^{n} ei^2}{n-k}}{\frac{\sum_{i=1}^{n} yi^2}{n-1}}$$

$$R^{-2} = 1 - \frac{\sum_{i=1}^{n} ei^2}{\sum_{i=1}^{n} yi}\left(\frac{n-1}{n-k}\right)$$

$$\text{Adjusted } (R^2) = R^{-2} = 1 - (1 - R^2)\left(\frac{n-1}{n-k}\right) \tag{5}$$

2.2.8. Multicollinearity

Multicollinearity is a statistical phenomenon in which two or more independent variables in a multiple regression model are highly correlated. Variance Inflation Factor (VIF) used to check Multicollinearity. The variance inflation factor (VIF) is the reciprocal of tolerance. When the VIF is greater than 10 the Collinearity is an issue. It means that the tolerance is below 0.10 (Fornell C.et al., 1996). Variance Inflation Factor (VIF) computed as follows:

$$VIF(\hat{\beta}_i) = \frac{1}{1-R_i^2} \tag{6}$$

Where: $\hat{\beta}_i$ is i^{th} regression coefficient or parameter

R_i^2 *is* i^{th} *coefficient of determination*

3. RESULTS AND DISCUTION

3.1. Descriptive Statistics

Descriptive statistics is method of collecting, organizing, presenting and analyzing data.

Table 2: Descriptive Statistics

Variables	Mean	Std. Deviation
Maize production in quintal	58838.80	474.761
farm size in hectare	2112.58	183.142
Number of participant	4456.75	241.669

Based on the table 2 above this study revealed that the mean of maize production in ten year in Yeki woreda was 58838.80 quintal, in 2112.58-hectare mean cultivated area and with mean of 4456.75 number of participants.

3.2. Multiple Linear Regression analysis

Table 3: Analysis of variance (ANOVA) table

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	26765574612.050	4	6691393653.013	73.788	.000
Residual	1360262936.729	15	90684195.782		
Total	28125837548.780	19			

df: degree freedom, F: calculated value, sig.: significance (p-value)

ANOVA table used to test an overall significance of the parameters in the study. According to table 3, the p-value of an overall test is 0.000. Therefore, the researcher conclude that at least one coefficient of regression (parameter) is different from zero. To identify the regression coefficient that made an overall test different from zero, an individual test used.

Table 4: Individual regression coefficient estimates

variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% CI for β	
	β	Std. Error	Beta			Lower	Upper
(Constant)	28170.306	877.718		3.173	.006	9247.90	47092.71
DAP (Fertilizer)	-21572.711	467.307	-.286	-2.889	.011	-37488.90	-5656.52
Urea (Fertilizer)	-27671.091	747.176	-.361	-3.927	.001	-42691.80	-12650.39
farm size in hectare	16.368	2.029	.674	8.067	.000	12.04	20.69
Number of participant	5.510	1.299	.335	4.242	.001	2.74	8.28

t: calculated value, sig.: significance (p-value), CI: confidence interval (95%)

Based on the table 4 above the constant or intercept is 28170.306 and the p-value is 0.006 with confidence interval of (9247.90, 47092.71), this implies that the intercept is statistical significant and is interpreted as assuming other variables zero the mean maize production in quintal is 28170.306. The variables DAP and Urea are dummy variables, so in case of dummy variables it needs reference categories. The coefficient for variable DAP by considering "no" as reference category is -21572.711, but by referring to the dummy variable model, the researcher can put the correct coefficient for variable DAP as 28170.306-21572.711= 6597.289. Since the intercept is the average value of all reference categories. The estimate for variable DAP is 6597.289. Based on the above table 4 the p-value is 0.011 with confidence interval (-37488.90, -5656.52), it implies that DAP has significant effect on the maize production. In addition, is interpreted as using other variables fixed the maize production in quintal is increase by 6597.289 when increasing the use of DAP fertilizer by a unit. The coefficient for variable Urea by considering "no" as reference category is -27671.091, as referred in regression model with dummy variable, the correct estimate for variable Urea is 28170.306-27671.091= 499.215.Therefore, an estimate for variable Urea is 499.215. According to above table 4 the p-value is 0.001 with Confidence interval (-42691.80, -12650.39), this implies that Urea has significant effect on the maize production. Moreover, is interpreted as using other variables fixed the maize production in quintal is increased by 499.215when Urea fertilizer was increased by a unit.

The coefficient for variable farm size in hectare is 16.368; from the table 4 above the p-value is 0.000 with confidence interval (12.04, 20.69). Holding other variables constant the maize production in quintal increased by 16.368, when farm size in hectare increased by a unit. The coefficient for variable Number of participant is 5.510; from the table 4 above the p-value is 0.001 with confidence interval (2.74, 8.28). Holding other variables constant the maize production in quintal increased by 5.510, when Number of participant increased by a unit.

3.3. Goodness of fit

Table 5: Coefficient of determination $(\boldsymbol{R^2})$

Model	R	R Square $(\boldsymbol{R^2})$	Adjusted R Square $(\boldsymbol{R^{-2}})$	Std. Error of the Estimate
	.976	.952	.939	522.825

Based on the table 5; $R^2 = 0.952$, implies that 95.2% variation of dependent variable is observed or explained by independent variables in the model. Adjusted R Square $(R^{-2}) = 0.939$, implies that the dependent variable is strongly explained by independent variables.

Table 6: Multicollinearity checking

Model	Collinearity Statistics	
	Tolerance	VIF
DAP (Fertilizer)	.329	3.044
Urea (Fertilizer)	.380	2.629
farm size in hectare	.463	2.162
Number of participant	.516	1.938

VIF: variance inflation factor

According to table 6 above the tolerance for all variables in the study is greater than 0.1 and VIF is less than 10. Therefore; the researcher concluded that there is no multicollinearity in the fitted model.

Assumption of linear regression model

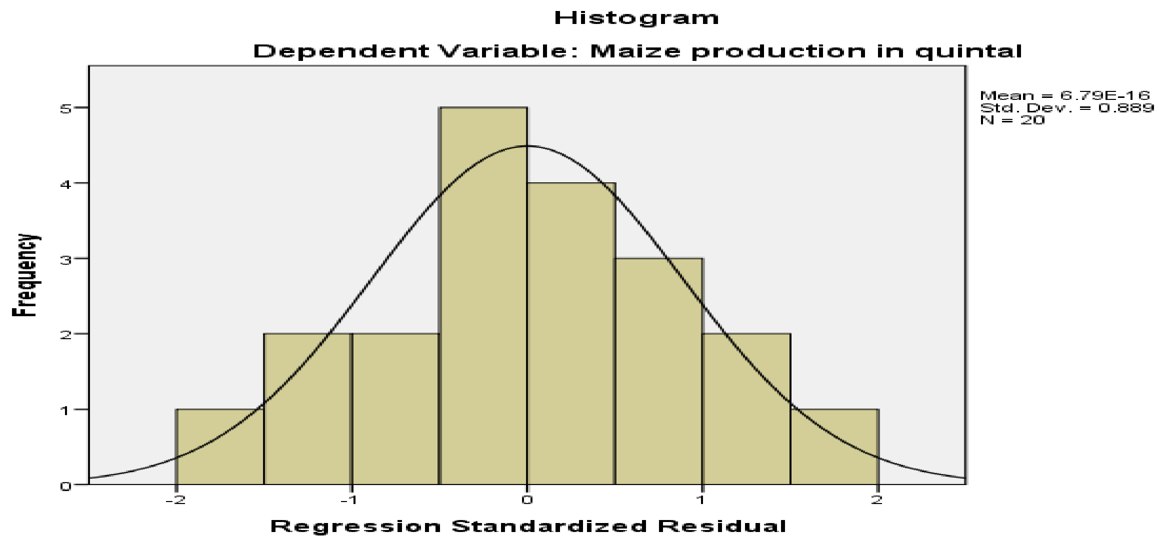

Figure 1: Histogram plot for assumption of normality checking

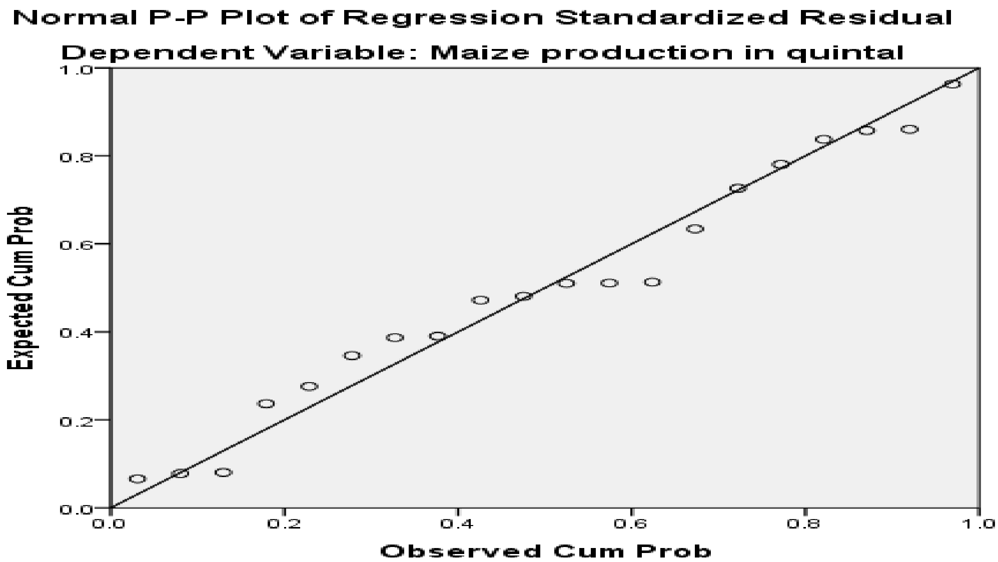

Figure 2: Normal P-P plot for assumption of linearity checking

According to figure 1 and figure 2 above the assumption of the linear model is satisfied; Means that from figure 1 above we can observe that the data was approximately normal distributed with mean zero and constant variance. In addition, from figure 2 we can observe that there is linearity between expected cumulative probability and observed cumulative probability.

3.4. Discussion

In this study descriptive and inferential statistical analysis used to identify factors that determine the production of maize in Yeki woreda, Sheka zone southwest Ethiopia. To fit data multiple linear regression model with dummy variable used. In addition, ANOVA table and goodness of fit test such as coefficients of determination, adjusted coefficient of determination and assumptions of linear regression model applied. Based on ANOVA table result an overall test is statistical significant, which means that at least one coefficient of regression or parameter was different from each other. To identify the parameters those are responsible for the significance of an overall test an individual test of parameters used. From variables included in this study DAP and Urea considered as dummy variables. According to an individual test, DAP, Urea, farm size and number of participant parameters was statistical significantly responsible for the significance of an overall test. The goodness of fit was checked by coefficient of determination and adjusted coefficient of determination, and the result showed that more than 90% variations of dependent variable explained by independent variables; this implies that the model is well fitted.

According to the results in this study the parameter DAP had significant effect on the production of maize. In addition, this study is in line with the report of Situation and outlook of Maize in Ethiopia which implied that using DAP fertilizer increased the production of maize (Dawit A. C. Y., 2014) and Fertilizer plays a vital role in maize production as explained by (Rebecca B. , 2011) in South Africa, and this statement supports the result. Urea is one of an important fertilizer in the production of crops in different areas of crop producing countries including Ethiopia, as stated in the different literatures (Dawit A. C. Y., 2014). In this study, the results showed that Urea fertilizer has statistical significant relationship with the production of maize; supported by study done in Sidama Zone by (Bealu T., 2013). In addition, is in line with the study conducted in Ethiopia by (Tsedeke A. et al., 2015); Ethiopia has one of the fastest growth rates of fertilizer usage in Sub-Saharan Africa (SSA). Farm size in hectare or cultivation area was one of the factors that determine the production of crops. The result of this study revealed that farm size had statistical significant effect on the production of maize and supported by the study conducted in the Nigeria by (Ibitola et al, 2019) farm size is significant and positively related to Maize output. In addition, it was in line with study done in Ethiopia by (Endrias G., 2013) farm size is highly significant for positively affecting the technical efficiency of smallholder maize producers. The number of participants had significant relationship with the production of maize and it is in line with the study done in Sidama Zone by (Bealu T., 2013).

3.5. Conclusion

Multiple linear regression model with dummy variable used to fit the data set in this study, with the general objective of identifying the determinants of production of maize in Yeki Woreda, Sheka zone south west Ethiopia. Based on the fitted model DAP, Urea, farm size and number of participant were identified as the major determining factors of the production of maize. The concerned body should give attention to the identified factors, to increase the production of maize in Ethiopia.

Competing interest: The author declare that no competing interests exist

Reference

Abdolreza A. (2006). Competitive Commercial Agriculture in Sub–Saharan Africa (CCAA) Study,Food and Agriculture Organization of the United Nations. Maize International Market Profile.

Ammani A., A. J. (2010). Effects of fertilizer liberalization on maize production in Nigeria. Journal of Development and Agricultural Economics, 2(11): 401-405.

Bealu T., E. G. (2013). Factors Affecting Economic Efficiency in Maize Production: The Case of Boricha Woreda in Sidama Zone, Southern Ethiopia.

CSA. (2009). Maize Value Chain Potential in Ethiopia. International food policy Research Institute.

CSA. (2012). Agricultural Sample Survey 2011-2012. Volume I, Area and Production of Crops. Ethiopia: Central Statistical Agency.

Dawit A., C. Y. (2014). Situation and Outlook of Maize in Ethiopia. Addis Abeba: Ethiopian Institute of Agricultural Research.

Dawit A., M. W. (2008). The Maize Seed System in Ethiopia: Challenges and Opportunities in Drought Prone Areas. African Journal of Agricultural Research, Vol. 3 (4), Pp. 305-314.

De Janvry A. and Sadoulet E. (2002). World Poverty and the Role of Agricultural Technology: Direct and Indirect Effects. . Journal of Development Studies, 38(4), 1-26.

Endrias G., A. B. (2013). Productivity and Efficiency Analysis of Smallholder Maize Producers in Southern Ethiopia. Journal of Hum Ecol, 41(1): 67-75.

Fornell C.et al. (1996). The American customer satisfaction index: Nature, purpose, and findings. Journal of Marketing, ,60(4), 7–18.

Huffnagel H. P. (1961). Agriculture in Ethiopia. . Rome: FAO.

Ibitola et al. (2019). Determinants of Maize Framers' Productivity among SmallholderFarmers in Oyo State, Nigeria. Greener Journal of Agricultural Sciences, Vol. 9(2), 189-198, 2019.

Johnson, M. D., & Fornell, C. et al. (n.d.).

Rebecca B. . (2011). Technical Efficiency in Maize Production by Small-Scale Farmers in Ga-Mothiba, Limpopo Province, South Africa. University of Limpopo Private Bag X1106, Sovenga, South Africa. .

Shahidur R., K. G. (2010). Maize Value Chain Potential in Ethiopia Constraints and Opportunities for Enhancing the System International Food Policy Research Institute, Ambo College, Ethiopia.

Smita S. (2009). The Use of Dummy Variables in Regression Analysis. http://www.moresteam.com.

SNNPR Livelihood Woreda Report. (2009). Key Parameters for Monitoring Livelihoods at Woreda Level.

Tsedeke A. et al. (2015). Factors that transformed maize productivity in Ethiopia. Food Sec. DOI 10.1007/s12571-015-0488-z.