Modelling Apiculture Production in Adamawa State, Nigeria: The Ordinary Least Squares Technique

Gabdo, B. H.

Department of Agricultural Economics and Extension, Adamawa State University, PMB 25, Mubi, Adamawa State, Nigeria

Abstract:- **The study adopted the famous Ordinary Least Squares (OLS) approach to model apiculture production in Adamawa State, Nigeria. The primary source method was employed for data collection from 108 apiarists in the state spread across the four agricultural zones in the state. The four OLS functional forms (Linear, Semilog, Exponential and Cobb-Douglas) were modelled as basis for selection of best fit on account of statistical, economic and econometric criteria. The apiculture data was adjudged normally distributed, free from heteroscedasticity, multi collinearity and autocorrelation, but, with model misspecification issue. The Cobb-Douglas model was adjudged the best fit owing to significance of all variables with respective *apriori* expected signs, highest R^2 = 92%, lowest RMSE = 0.286, AIC = 41.12 and BIC = 54.53. The study shows that labor, hives, farm distance and other costs adjudged significant variables affecting the production of honey. Together, they account for 92% of the variation in honey production. The study recommends timely and judicious utilization of resources for optimal honey yield in the area.**

Keywords: **Adamawa, Production, Apiculture, Cobb-Douglas, Beehives.**

I. INTRODUCTION

Apiculture is a sub-sector of the agricultural industry that deals with management of bees for the production of honey owing to its diverse utility to humanity. The significance of apiculture to agriculture and indeed the economy is further highlighted; NHB (1998) stated that apiculture is important because one-third of human diets are derived from insect pollinated plants and honey bees accounts for 80% of the entire plant pollination. Globally, apiculture is managed under two distinct systems: modern and traditional management. However, in Nigeria and indeed the study area, the traditional system of apiculture management predominates. Ojeleye (1999), Ja'afar-Furo (2005, 2007 and 2016) and Ja'afar-Furo and Madu (2017) asserted that Nigeria's apiculture management is largely traditional despite the advances in modern beekeeping introduced as far as 1914 in Zaria.

In terms of production, Shahbandeh (2019) stated that 2018 witnessed global peak in terms of number of beehives (92.27 million beehives), but peak global production was recorded in 2015 with a production index of 1.83 million metric tons; valued 7 billion USD in 2016. Thomas and Schumann (1992) stated that a massive 120,000 bees are needed to produce just a Kg of honey and on average, a bee produces just a teaspoon of honey in its entire life span. The study draws inspiration from the foregoing assertion. This scenario of large flock of bee population producing just a Kg of honey indicates low productivity; especially that bees are faced daily with threats to survival. The study therefore, employed the Ordinary Least Squares (OLS) technique to model four functional forms in order to dispel the production relationship in the apiculture industry and identify adjustable variables to mimic if meaningful productivity rise is to be met.

II. METHODOLOGY

The Study Area

The area of research was Adamawa State located in the North-eastern Nigeria; with geographical location between Latitudes 8.00^0N and 11.00^0N of the equator and Longitudes 11.50^0E and 13.50^0E of the Greenwich meridian and 39,742.12 square kilometers in land mass (Adebayo, 1999). The State is highly populated with population index 3,161,374 (NPC, 2006). It has two major seasons; dry season (between November and March) and rainy season (between April and October); it also has average monthly temperature range 26.7^0C to 27.8^0C and annual average rainfall range 700mm to 1000mm (Adebayo, 1999). The State is known for agricultural production such as crop production and livestock rearing, although apiculture is not one of the major agricultural enterprises deriving the economy, but it one of the supporting enterprises to so many farming families.

Method of Data Collection

The study used primary source of data with the aid questionnaire supported with personal interviews. The data was collected during the 2017/2018 production season.

Sampling Technique

The study used multi-stage sampling technique to collect apiculture data from apiarists. In the first stage, the state was stratified according to the four agricultural zones; Northeast, Northwest, Central and Southwest zones in the state. In the next stage, purposive sampling was employed to select two each most renown local government areas in the four zones selected. In the final stage, random sample was used to select 15 apiarists in each of the 8 selected local governments; thus, 120 apiarists were engaged in study. However, only data from 108 apiarists were adjudged reliable and hence used in analysis.

Table 1: Selected Local Government Areas of the study based on Agricultural Zones in Adamawa State, Nigeria

Agricultural Zone	Selected Local Government Area
Zone 1	Maiha and Mubi South
Zone 2	Hong and Song
Zone 3	Yola South and Fufore
Zone 4	Ganye and Toungo

Source: Computed from Field Survey (2018)

Analytical Techniques

Ordinary Least Squares (OLS) Regression

The study employed the method of Ordinary Least Squares (OLS) to estimate four functional forms as follows:

Linear → $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u_i$ eqn 1

Semi log → $Y = \beta_0 + \beta_1 LnX_1 + \beta_2 LnX_2 + \beta_3 LnX_3 + \beta_4 LnX_4 + u_i$ eqn 2

$Exponential \rightarrow LnY = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u_i$ eqn 3

Double log → $LnY = \beta_0 + \beta_1 LnX_1 + \beta_2 LnX_2 + \beta_3 LnX_3 + \beta_4 LnX_4 + u_i$ eqn 4

Where: X_1= Labor (Man hour), X_2= Hives (Number), X_3 = Farm Distance (Km), X_4 = Other Costs (₦).

Akaike Information Criteria (AIC)

Akaike (1974) introduced the Akaike Information Criteria (AIC) and it as:

$$AIC = -2ln\mathcal{L}_{max} + 2k \quad \text{eqn 5}$$

Where $\mathcal{L}_{max}$ represents the achievable maximum likelihood by the model and k represents number of parameters in the model. The best model is adjudged to have minimized value of AIC (Liddle, 2007).

Bayesian Information Criteria (BIC)

The Bayesian Information Criteria (BIC) was derived from Bayes factor (Kass and Raftery, 1995) introduced by Schwarz (1978) and expressed as:

$$BIC = -2ln\mathcal{L}_{max} + klnN \quad \text{eqn 6}$$

Where $\mathcal{L}_{max}$ represents the achievable maximum likelihood by the model, k represents number of parameters in the model and N represents the number of data points used to fit the model. The best fit is that model that minimizes the BIC (Mcquarrie and Tsai, 1998).

Durbin-Watson-d-Statistic

The Durbin-Watson-d-statistics was applied to detect the presence of autocorrelation in the apiculture data. Baum (2006) stated the Durbin-Watson-d-statistics old as it is, still relevant and widely used to detect presence of autocorrelation. The test is expressed below:

$$d = \frac{\sum_{t=2}^{t=n}(e_t - e_{t-1})^2}{\sum_{t=1}^{t=n} e_t^2}; \simeq 2(1-\rho);\ 0 \le d \le 4 \quad \text{eqn 7}$$

Where d= Computed Durbin-Watson-d value

The decision criteria for test is to compare the computed d value with the tabular d_L and d_U. If $d<d_L$ = presence of positive autocorrelation, if $d> d_U$ or $d< 4 - d_U$ = No autocorrelation and if $d> 4 - d_L$ = Presence of negative autocorrelation.

Variance Inflating Factor (VIF) Test

The variance inflating factor (VIF) was applied to detect collinearity among independent variables in regression models (Murray *et al.*, 2012). The VIF shows how the variance of an estimator is inflated under a scenario of multi collinearity (Gujarati and Porter, 2009). Chatterjee and Price (1977) and Belsley *et al.* (1980) stated that the VIF for multiple regression model with p predictors; X_i; $i = 1, \ldots. \rho$, are the diagonal elements (r^{ii}) of the correlation matrix R_{pxp} of the predictors. The VIF is described below for a given predictor variable:

$$\text{Variance Inflating Factor } (VIF_i) = r^{ii} = \frac{1}{1-R_i^2}; i = 1, \ldots\ldots\rho \quad \text{.eqn 8}$$

Where R_i^2 = Multiple correlation coefficient. If VIF of a variable > 10; usually occurs when R^2 > 0.90 shows that the variable is highly collinear; thus, the larger the VIF value the more worrisome.

Cook Weisberg Test

The Cook Weisberg test was also employed to measure the presence of heteroscedasticity. Yafee (2012), expressed the Cook Weisberg test as follows:

$$Var\ (e_i) = \delta^2 exp(Z_t) \quad \text{eqn 9}$$

Where e_i= error in regression model, $Z = X\hat{\beta}$ = variable list supplied by user

The test is whether t = 0

$$e_i^2 = \alpha + Z_i t + V_i \quad \text{eqn 10}$$

it forms a source test

$$S = \frac{SS\ of\ the\ model}{2} \quad \text{eqn 11}$$

$$h_o: S_{df=p} \sim X^2 \quad \text{eqn 12}$$

Where p = number of parameter

III. RESULTS AND DISCUSSION

Table 2 presented the Ordinary Least Squares (OLS) regression results for modelling the apiculture production in the research area. Four models are presented based on the four famous functional forms, namely: Linear, Semi log, Exponential and Cobb-Douglas functional forms. This was aimed to identify the model with the best fit that adequately suits the apiculture production in the area based on empirical criteria of the study. The study dwelled on statistic, economic and econometric criteria such as number of significant independent variables, F-value, R^2 value, Root Mean Square Error (RMSE), Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) as empirical evidences for the selection of the best fit of the four models tested for apiculture production. Evaluation of estimates on Table 2 on the basis of statistic, economic and econometric criteria revealed Cobb-Douglas (also known as Double Log or Log-Log Model or Model 4) as the best fit for apiculture production; thus, used to draw inference and derive conclusions of the study.

The basis for selection of the Cobb-Douglasmodel as the best fit for apiculture production are as follows: First, it returned all the four independent variables as significant; no other model of the four tested models returned that high. Labor (X_1) (0.284), number of hives (X_2) (0.597) and other costs (X_4) (0.131)all returned positive coefficient and significant at 1%, while farm distance (X_3) (-0.133) had negative coefficient and significant at 5%. This implies that a percent increase in labor, number of hives, other costs and farm distance, increases honey output by 0.284%, 0.597%, 0.131% and decreases honey output by 0.133% respectively. This finding is rational; increase in all the four factors of production (labor, number of hives and other costs) results to increase in the yield of honey, while farm distance (home to farm distance) decreases honey yield as distance increases. This means that the closer the honeybee farm is to home, the more the frequency of visits and the better management it gains. Note, number of hives had the highest coefficient; an indication that it is the most important of the four variables in apiculture production.

Secondly, the Cobb-Douglas also had the highest R^2 value (92.08%). This implies that 92% of the honey yield is explained by the four independent variables included in the model, while the remaining 8% accounts for error; including variables not embedded in the model. Thirdly, the Cobb-Douglas had the lowest RMSE value (0.286). The RMSE measures the dispersion of the residual around the predicted value; it is famously used in regression analysis to compare model selection; the lower the RMSE value, the better the model fit (Barnston, 1992).Fourthly, the model had the lowest AIC value (41.12). Lastly, it also had the lowest BIC value (54.53). Liddle (2007) and Mcquarrie and Tsai (1998) stated that both AIC and BIC are used as selection criteria for preferred or best fit models; models with lower values of AIC and BIC are more preferred or better fit than those with higher values.

Table 2: Ordinary Least Squares Regression Results across Four Functional Forms

Variable	Model 1 Linear Function (Lin-Lin Model)		Model2 Semi log Function (Lin-Log Model)		Model 3 Exponential Function (Log-Lin Model)		Model 4 Cobb-Douglas Function (Log-Log Model)	
	Coefficient (t value)	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Labor (X_1)	0.666*** (3.06)	0.217	-21.453 (-0.36)	59.077	0.003*** (4.65)	0.0006	0.2839*** (4.20)	0.068
Number of Hives (X_2)	8.209*** (9.26)	0.887	286.169*** (4.85)	59.013	0.015*** (5.68)	0.0025	0.5967*** (8.85)	0.068
Farm Distance (X_3)	-17.080 (-1.45)	11.787	-52.572 (-0.96)	54.542	-0.055 (-1.61)	0.0342	-0.1330** (-2.13)	0.062
Other Costs (X_4)	0.006*** (3.91)	0.002	125.801*** (4.01)	31.358	$2.5x10^{-6}$ (0.57)	$4.5x10^{-6}$	0.1307*** (3.65)	0.036
Constant	92.910* (1.86)	49.987	1296.215*** (-5.73)	226.114	4.951*** (34.09)	0.1452	1.6201*** (6.27)	0.259
N	108		108		108		108	
F (4, 103)	148.86		68.06		67.94		299.49	
P > F	0.000		0.000		0.000		0.000	
R^2	85.25%		72.55%		72.52%		92.08	
Root MSE	183.50		250.36		0.533		0.286	
AIC	1437.22		1504.31		175.53		41.12	
BIC	1450.63		1517.73		188.94		54.53	

Source: Computed from Field Survey (2018)

Table 3 presents 3 types of residuals to show its distribution in order to give the normality status of the apiculture data. All residuals show very low mean values and standard deviation. This indicates that the predicted residuals have less deviated insignificantly from its observed values (Gujarati and Porter, 2009)

Table 3: Predicted Residuals for Testing Normality and Other Diagnostic Checks

Variable (residual type)	**Observation**	**Min.**	**Max.**	**Mean**	**Standard Deviation**
Normalized Residual	108	$-1.1x10^3$	741.90	$3.8x10^{-7}$	180.04
Standardized Residual	108	-7.61	4.55	-0.009	1.13
Studentized Residual	108	-11.46	5.07	-0.04	1.43

Source: Computed from Field Survey (2018)

Table 4 shows the link test result for testing model specification. The _hatsq is significant at 5% ($P \leq 0.05$); this indicates a failure to reject the null hypothesis of no variable omission. This further implies the presence of variable bias in the model. Both exclusion of important variable and inclusion of non-important variable in a model may lead to variable bias. Frost (2019) states omitted variable bias is a component of model misspecification. The finding therefore detects a condition of model misspecification; the study did not test to find which model was misspecified (included or excluded). Hence, the need for further research to unravel the actual variable misspecified.

Table 4: Result of Link Test for Detecting Model Specification

Y	Coefficient	Standard Error	T-value	P-value	95% Confidence Interval
_hat	1.2796	0.130	9.87	0.000	1.023 1.537
_hatsq	-0.0002	$7.3x10^{-5}$	-2.27	0.025	$-3.1x10^{-4}$ $-2.1x10^{-5}$
_Cons	-65.4812	38.613	-1.70	0.093	-142.043 11.081
N = 108					
F (2, 105) = 320.90					
P > F = 0.0000					
R^2 = 85.94					
Root MSE= 177.46					

Source: Computed from Field Survey (2018)

Table 5 presents Durbin-Watson test for autocorrelation. Given the tabular d-values: d_L = 1.461 and d_U = 1.625, then, $4 - d_U$ = 2.375. the computed d-value (1.7067) is greater than d_U, but lower than $4 - d_U$. This implies failure to reject both Ho and Ho*; thus, indicates the absence of autocorrelation in the apiculture data. Gujarati and Porter (2006) states the decision criteria for Durbin-Watson statistics includes absence of autocorrelation if the d-computed is greater than d_U, but lower than $4 - d_U$.

Table 5: Result of Durbin-Watson d-Statistics for Testing First Order Autocorrelation

Observation	k	Computed d-statistics (value)
108	4	1.7067

Source: Computed from Field Survey (2018)

Table 6 shows results of Variance Inflating Factor (VIF) and tolerance for detecting multi collinearity in the apiculture data. The VIF for all variables and the mean VIF all situated away from the value 10; which indicates high evidence for absence of multi collinearity. Similarly, the tolerance for Other cost (X_4) tends to one (1); indicating high evidence of absence of multi collinearity, while, labor (X_1), number of hives (X_2) and farm distance (X_3) show less evidence for absence of multi collinearity. Gujarati and Porter (2009) states as the tolerance tends to one (1), the less chances of existence of multi collinearity and vice versa.

Table 6: Result of Variance Inflating Factor for Detecting Multi collinearity

Variable	Variance Inflating Factor (VIF)	Tolerance (1/VIF)
Labor (X_1)	3.25	0.3082
Number of Hives (X_2)	2.18	0.4578
Farm Distance (X_3)	2.04	0.4895
Other costs (X_4)	1.02	0.9808
Mean of Variance Inflation Factor (VIF)	2.12	

Source: Computed from Field Survey (2018)

Table 7 shows the Cook-Weisberg result for detection of heteroscedasticity in the apiculture data. The result shows significant chi-square; which indicates absence of heteroscedasticity; but, absence of heteroscedasticity also means presence of homoscedasticity. If the chi-square is insignificant, indicates homoscedasticity (Yafee, 2012) and if the chi-square is significant, it indicates heteroscedasticity (Baum, 2006). Thus, result in the table conspicuously indicates homoscedasticity in the apiculture data.

Table 7: Results of Cook-Weisberg Tests for Detecting Heteroscedasticity

Test Type	**Chi^2**	**$P > Chi^2$**
Cook-Weisberg Test for Heteroscedasticity	28.18	0.0000

Source: Computed from Field Survey (2018)

IV. CONCLUSION

The study estimated apiculture production using the linear, semi log, exponential and Cobb-Douglas models of OLS regression. Diagnostic checks revealed the apiculture data as normally distributed and free from heteroscedasticity, multi collinearity and autocorrelation. Model comparison on the basis of statistic, economic and econometric criteria indicated the Cobb-Douglas model as the best fit. The study revealed that all the variables included in the model with their respective *apriori* expected signs significantly affect honey production and the variables explained 92% variation in honey. In the end, the study recommends timely and judicious utilization of production inputs if the aim of optimum honey production is to be met.

REFERENCE

[1]. Adebayo, A. A. (1999). *Climate I and II*. In: Adamawa State in Maps (eds. A.A. Adebayo and A.L. Tukur), Paraclete Publishers, Yola, Nigeria, 1: 10-26.

[2]. Akaike, H. (1974). *A New Look at the Statistical Model Identification*. Institute of Electrical and Electronic Engineering, Transaction on Automatic Control, 19 (6): 716-723.

[3]. Barnston, A. G. (1992). *Correspondence among the Correlation, RMSE and Heidke Forecast, Verification Measures; Refinement of the Heidke Scores*. Notes and Correspondence, 699-709

[4]. Baum, C. F. (2006). *An Introduction to Modern Econometrics Using Stata*. Department of Economics, Boston College, USA, Stata Press Publication, 341Pp.

[5]. Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*. John Wiley and Sons, New York, 292Pp.

[6]. Chatterjee, S. and Price, B. (1977). *Regression Analysis by Example*. John Wiley and Sons, New York, 19-22.

[7]. Frost, J. (2019). Regression Analysis: An Intuitive Guide for using and Interpreting Linear Models.340Pp.

[8]. Gujarati, D. N. and Porter, D. C. (2 009). *Basic Econometrics.* 5th Edition, India, 922Pp.

[9]. Ja'afar-Furo MR, 2005. Economics of smallholder beekeeping in Adamawa State, Nigeria. Thesis for PhD, Abubakar Tafawa Balewa University, Bauchi, Nigeria (Ng).

[10]. Ja'afar-Furo MR, 2006. Analysis of costs and returns, and major constraints to beekeeping industry in Adamawa State, Nigeria. Journal of Research in Agriculture, 5(1), 14-19.

[11]. Ja'afar-Furo MR, 2007. Appraising the perception of farming communities towards adoption of apiculture as viable source of income in Adamawa State, Nigeria. Apiacta, 42, 1-15.

[12]. Ja'afar-Furo MR and Madu UA, 2017. Awareness campaign on beekeeping in institutions of learning and rural communities in Nigeria: The need for policy intervention. International Journal of Agricultural Research, Sustainability and Food Security, 3(1), 32-34.

[13]. Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of American Statistical Association*, 90 (430): 773-795.

[14]. Liddle, A. R. (2007). Information Criteria for Astrophysical Model Selection. Monthly Notices of the Royal Astronomical Society, 377: 74-78.

[15]. Mcquarrie, A. D. R. and Tsai, C. L. (1998). Regression and Time Series Model Selection.World Scientific Publication Company, Singapore, 455Pp.

[16]. Murray, L., Nguyen, H., Lee, Y., Remmenga, M. and Smith, D. (2012). *Variance Inflating Factors in Regression Models with Dummy Variables. C*onference on Applied Statistics in Agriculture, Kansas State University, 19Pp.

[17]. NHB (1998). National Honey Board. http//www.honey.com.

[18]. Ojeleye B, 1999. Foundation of Beekeeping in the Tropics. Center for Beekeeping Research and Development (CEBRAD), Ibadan, Nigeria, pp. 186-197.

[19]. Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, 6 (2): 461-464

[20]. Shahbandeh, M. (2020). Agriculture: Farming, Number of Beehives Worldwide from 2010 to 2018. Statista. http//www.statista.com.

[21]. Thomas, M. G. and Schumann, D. R. (1992). Non-wood Forest Products from Temperate Broad- leaved Trees: In William M. C. (2002). Food and Agricultural Organization, Technical Report.

[22]. Yafee, R. A. (2012). *Robust Regression Modelling with Stata.* Lecture Note. Statistics, Social Science and Mapping Group Academic Computing Services, 93Pp.