

**APPLICATION OF LINEAR PROBABILITY MODEL IN THE STUDY OF  
HOME OWNERSHIP AND FAMILY INCOME FOR SOME SELECTED  
FAMILIES IN AFIKPO NORTH L.G.A**

**BY**

**<sup>1</sup>F. A. UJAH, <sup>2</sup>MADU CHUKWUNONSO**

**AND**

**<sup>3</sup>OKEEAHIALAM A. H**

**MATHEMATICS/STATISTICS DEPARTMENT  
AKANU IBIAM FEDERAL POLYTECHNIC, UNWANA**

**ABSTRACT**

*This paper employs linear probability model in the analysis of home ownership and family income for some selected Ebonyi state workers in Afikpo North LGA. The work aimed at demonstrating the application of linear probability model to the analysis of qualitative response regression model with the regressand or dependent variable having binary or dichotomous variable, to determine the probability of none home owners to own a house, And to ascertain the mean income of Ebonyi worker in Afikpo North to owning a house. Primary data of sample of 100 respondents were obtained using purposive method of a non-probability sampling technique. Out of the 100 non home owners sampled, the result showed that the mean increase rate of income of a family in Afikpo North to owning a house is 12%, which is far from 100%. The estimated value from Linear Probability Model (LPM) applied in this work indicates that the probability of non-home owners to own a house in Afikpo North Local Government Area is 0.12 which is very insignificant and defiles the possibility owning a house. We therefore recommend that the government of Ebonyi State should look into the staff welfare and increase their salaries to enable them own a house of their own.*

## INTRODUCTION

A model in which the dependent variable (regressand) itself is qualitative with binary response variable in nature, usually poses interesting difficult challenges in analysing. This type of regression model is seen as qualitative response regression model. Most or frequently used regression model have implicitly assumed that the regressand, the independent variable or the response variable  $Y$  is qualitative, whereas the explanatory variables are quantitative, qualitative or dummy or a mixture thereof. Although this model, as increasingly used in various areas of social sciences and medical research, qualitative response regression models pose interesting estimation and interpretation challenges.

This work studies qualitative response models by first considering the binary response regression model. According to Damoda N. Gujarati (2005), there are three approaches to developing a probability model for a binary response variable:

1. The Linear Probability Model (LPM)
2. The Logit model
3. The Probit Model

The linear probability model is chosen for the analysis and estimation of data maybe because of its comparative simplicity and its capability of being estimated with ordinary least square (OLS)

## 2. AIM/OBJECTIVE

The application of Linear Probability model to rate the relationship of the income of 100 Ebonyi state workers to their ownership of a house in Afikpo North

is aimed at

1. To demonstrate the application of LPM to the analysis of qualitative response regression model with the regressand or dependent variable having a binary or dichotomous variable.
2. To determine the probability of non-home owners to own a house.
3. To ascertain the mean rate of income of a family in Afikpo North to owning a house.

### 3. METHODOLOGY

The Linear Probability Model is used as a method of estimation because it has the ability to model a binary or dichotomous dependent variable as it is the case in this paper.

## The Linear Probability Model (LPM)

Consider the following regression model

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \dots \dots \dots eqn1$$

Where,

$$X_i = \text{family income}$$

$Y_i = 1$  if the family owns a house

*o, if the family does not own a house*

$\beta_1$  &  $\beta_2$  = parameter of LPM.

$$\mu_i = \text{error term of LPM}$$

This type of model (as in eqn1 above) is called Linear Probability Model (LPM), because its regressand or dependent variable is binary or dichotomous.

The conditional expectation  $X_i, E(Y_i/X_i)$ , can be interpreted as the conditional probability that the event will occur given  $X_i$  that is  $P_i(Y_i = 1/X_i)$

Thus in this paper  $E(Y_i/X_i)$  is given as the

probability of a family owning a house whose income is the amount of  $X_i$

Justifying the LPM as linear regression model, assuming  $E(\mu_i) = 0$ , as usual to obtain unbiased estimator. we have

$$E(Y_i/X_i) = \beta_1 + \beta_2 X_i + \mu_i \dots \dots \dots \text{eqn2}$$

Now if  $P_i = \text{probability that } Y_i = 1$  (that is the event occur)

and  $1 - P_i = \text{probability that } Y_i = 0$  (that is the event did not occur).

Then, the variable  $Y_i$  has the following probability distribution

$Y_i$	Prob
0	$1 - P_i$
1	$P_i$
Total	1

This implies that  $Y_i$  follows Bernoulli probability distribution.

Therefore, by the definition of mathematical expectation, we obtain

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \dots \dots \dots \text{eqn3}$$

Comparing eqn2 with eqn3, we can equate

$$E(Y_i/X_i) = \beta_1 + \beta_2 X_i = P_i \dots \dots \dots \text{eqn4}$$

This proves the fact that the conditional expectation of the model is interpreted as the conditional probability of  $Y_i$ . In general, the expectation of a Bernoulli random variable is the probability that random variable equals 1. The term success is defined in the context of the problem. Since the probability  $P_i$  must lie between 0 and 1, we have the restriction

$$0 \leq E(Y_i/X_i) \leq 1$$

That is, the conditional expectation or conditional probability must lie between 0 and 1.

## Draw Backs of LPM

The linear probability model poses several problems, which are as follows:

1. Non-Normality of the Disturbance  $u_i$
2. Heteroscedasticity Variance of the Disturbances

We know that in the presence of Heteroscedasticity, OLS estimators, although unbiased, are not efficient, that is they do not have minimum variance. But the problem of heteroscedasticity, like the problem of non-normality, is not insurmountable. Since the variance of  $u_i$  depends on  $E(Y_i/X_i)$ , one way to resolve the heteroscedasticity problem is to transform the model in (eqn1) by dividing it through by

$$\sqrt{E(Y_i/X_i)[1 - E(Y_i/X_i)]} = \sqrt{P_i(1 - P_i)} = \sqrt{w_i}$$

Non probability sampling procedure was adopted to select the number of families that were used during the study. A choice of purposive sampling technique was the best to study 100 workers of Ebonyi state Nigeria.

## Estimation Procedure Involved in LPM

1. Run the OLS regression and obtain  $\hat{Y}_i$  = estimate of the true  $E(Y_i/X_i)$

2. Then obtain  $\hat{w}_i = \hat{Y}_i(1 - \hat{Y}_i)$  the estimate of  $w_i$

3. Use the estimated  $w_i$  to transform the data and estimate the transformed equation by OLS (i.e weighted least squares).

TABLE 1  
HYPOTHETICAL DATA ON HOME OWNERSHIP (Y=1 if owns homes, 0 otherwise)  
AND INCOME X (in thousands of naira) OF SELECTED STATE WORKERS IN  
AFIKPO NORTH.

Famil	Y	X	Famil	Y	X
1	0	28	51	0	18
2	1	36	52	1	39
3	1	38	53	1	40
4	0	31	54	0	27
5	0	32	55	1	38
6	1	39	56	0	29
7	1	40	57	0	25
8	0	33	58	1	41
9	0	29	59	1	37
10	0	30	60	1	34
11	1	37	61	0	21
12	1	38	62	0	25
13	0	34	63	1	35
14	1	40	64	1	45
15	0	26	65	0	18
16	1	39	66	1	37
17	1	36	67	1	39
18	0	30	68	0	24
19	0	28	69	1	39
20	1	38	70	0	17

<b>21</b>	1	42	<b>71</b>	0	25
<b>22</b>	1	36	<b>72</b>	1	38
<b>23</b>	0	32	<b>73</b>	1	43
<b>24</b>	0	31	<b>74</b>	1	46
<b>25</b>	1	36	<b>75</b>	1	36
<b>26</b>	0	31	<b>76</b>	0	24
<b>27</b>	1	40	<b>77</b>	0	28
<b>28</b>	1	38	<b>78</b>	1	37
<b>29</b>	0	31	<b>79</b>	1	38
<b>30</b>	0	30	<b>80</b>	0	26
<b>31</b>	1	37	<b>81</b>	0	29
<b>32</b>	0	33	<b>82</b>	1	39
<b>33</b>	1	41	<b>83</b>	1	35
<b>34</b>	1	40	<b>84</b>	1	36
<b>35</b>	0	31	<b>85</b>	0	27
<b>36</b>	0	28	<b>86</b>	1	38
<b>37</b>	1	37	<b>87</b>	1	40
<b>38</b>	1	36	<b>88</b>	1	42
<b>39</b>	0	27	<b>89</b>	0	25
<b>40</b>	1	37	<b>90</b>	0	27
<b>41</b>	1	42	<b>91</b>	1	43
<b>42</b>	1	40	<b>92</b>	1	48
<b>43</b>	0	18	<b>93</b>	1	39
<b>44</b>	0	28	<b>94</b>	0	27
<b>45</b>	1	39	<b>95</b>	0	18
<b>46</b>	1	35	<b>96</b>	1	36
<b>47</b>	0	28	<b>97</b>	0	26
<b>48</b>	1	39	<b>98</b>	0	24
<b>49</b>	1	36	<b>99</b>	1	37
<b>50</b>	1	42	<b>100</b>	1	40

## RESULTS

Regression result before data transformation

Based on the data before me the linear probability model LPM estimated by OLS is as follows

parameters	constant	Slope[mean]
$\hat{y}$	-1.52	0.062
S	0.130	0.004
T	-11.7000	16.410
P	0.000	0.000
$R^2 = 0.733$ $n = 100$ DW = 2.092		

## INTERPRETATION:

Based on the output above, the intercept of **-1.52** gives the probability that a family with zero income will own a house. Since

this value is negative, and probability cannot be negative. We therefore treat this value as zero, which is sensible in the present instance. The slope or mean value of 0.06 implies that a unit change in income brings about 0.06 or 6% increase in the probability of owning a house by Ebonyi state worker in Afikpo North LGA. Both coefficients are statistically significant with absolute t-ratios (-11.700 and 16.410) greater than the approximate 5% two-tailed critical value of 2.000. There is approximately 73% explanatory power according to  $R^2$  which indicates a good model fit.

There is no evident (positive) autocorrelation since DW (=2.092) is substantially above 2.000.

## ACTUAL, ESTIMATED AND WEIGHTS FOR THE HOME OWNERSHIP SURVEY

$Y_i$	$Y_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$	$Y_i$	$\hat{Y}_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$
0	0.219	0.171	0.414	0	-0.404*		
1	0.716	0.203	0.451	1	0.898	0.201	0.303
1	0.839	0.135	0.368	1	0.960	0.029	0.196
0	0.405	0.241	0.491	0	0.154	0.175	0.361
0	0.467	0.249	0.499	1	0.836	0.242	0.370
1	0.902	0.088	0.297	0	0.278	0.029	0.448
1	0.964	0.035	0.186	0	0.030	0.228	0.171
0	0.529	0.249	0.499	1	1.022**		
0	0.281	0.202	0.450	1	0.774	0.175	0.418
0	0.343	0.225	0.475	1	0.588	0.092	0.492
1	0.778	0.173	0.416	0	-0.218*		

$Y_i$	$Y_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$		$Y_i$	$\hat{Y}_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$
<b>1</b>	<b>0.839</b>	<b>0.135</b>	<b>0.368</b>		<b>0</b>	<b>0.030</b>	<b>0.092</b>	<b>0.171</b>
<b>0</b>	0.591	0.242	0.492		1	0.650	0.029	0.477
<b>1</b>	0.964	0.035	0.186		1	1.270**		
<b>0</b>	0.095	0.086	0.293		0	-0.404*		
<b>1</b>	0.902	0.088	0.297		1	0.774	0.137	0.418
<b>1</b>	0.716	0.203	0.451		1	0.898	0.205	0.303
<b>0</b>	0.343	0.225	0.475		0	-0.032*		
<b>0</b>	0.219	0.171	0.414		1	0.898	0.169	0.303
<b>1</b>	0.839	0.135	0.368		0	-0.466*		
<b>1</b>	1.088**				0	0.030	0.175	0.171
<b>1</b>	0.716	0.203	0.451		1	0.836	0.137	0.370
<b>0</b>	0.467	0.249	0.499		1	1.146**		
<b>0</b>	0.405	0.241	0.491		1	1.332**		
<b>1</b>	0.716	0.203	0.451		1	0.712	0.084	0.453
<b>0</b>	0.405	0.241	0.491		0	-0.032*		
<b>1</b>	0.964	0.035	0.186		0	0.216	0.201	0.412
<b>1</b>	0.839	0.135	0.368		1	0.774	0.092	0.418
<b>0</b>	0.405	0.241	0.491		1	0.836	0.228	0.370
<b>0</b>	0.343	0.225	0.475		0	0.092	0.177	0.289
<b>1</b>	0.778	0.173	0.416		0	0.278	0.130	0.448
<b>0</b>	0.529	0.249	0.499		1	0.898	0.137	0.303
<b>1</b>	1.026**				1	0.650	0.038	0.477
<b>1</b>	0.964	0.035	0.186		1	0.771	0.177	0.420
<b>0</b>	0.405	0.241	0.491		0	0.154	0.130	0.361
<b>0</b>	0.219	0.171	0.414		1	0.836	0.137	0.370
<b>1</b>	0.778	0.035	0.416		1	0.960	0.038	0.196

$Y_i$	$Y_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$	$Y_i$	$\hat{Y}_i$	$\hat{w}_i$	$\sqrt{\hat{w}_i}$
1	0.716	0.171	0.451	1	1.084**		
0	0.157	0.088	0.364	0	0.030	0.029	0.171
1	0.778	0.226	0.416	0	0.154	0.130	0.361
1	1.088**			1	1.146**		
1	0.964	0.171	0.186	1	1.456**		
0	-0.402*			1	0.898	0.092	0.303
0	0.219	0.088	0.414	0	0.154	0.130	0.361
1	0.902	0.203	0.297	0	-0.404*		
1	0.654	0.092	0.476	1	0.712	0.205	0.453
0	0.219	0.038	0.414	0	0.092	0.084	0.289
1	0.902	0.130	0.297	0	-0.032*		
1	0.716	0.137	0.451	1	0.774	0.175	0.418
1	1.088**			1	0.960	0.038	0.960

\*Treated as zero to avoid probabilities being negative. Treated as unity to avoid probabilities exceeding one  $\hat{Y}_i(1 - \hat{Y}_i)$

## RESULTS

Regression result before data transformation

Based on the data before me the linear probability model LPM estimated by OLS is as follows:

Parameters	Constant	Slope[mean]
$\hat{Y}$	-1.2456	0.1195
S	0.1206	0.0069
T	-10.7000	17.410
P	0.000	0.000
$R^2 = 0.9214$ $n = 100$		
DW = 2.102		

## INTERPRETATION:

The result from this estimated weighted Least Square (WLS) regression showed that compared with the Ordinary Least Square (OLS) estimated regression equation in this work, showed smaller standard error, large t-ratios and better  $R^2$  model fit, though some observation were dropped (about 20 observations). Also, since  $W_i$  are estimated, the usual statistical hypothesis testing procedures are strictly speaking valid in the large samples.

Based on the estimated output above after data transformation, the intercept of -



1.2456 gives the probability that a family with 1.256 income will own a house. Since this value is negative, and probability cannot be negative. We therefore treat this value as zero, which is sensible in the present instance. The slope or mean value of 0.1195 implies that a unit change in income brings about 0.1195 or 11.95% increase in the probability of owning a house by Ebonyi state worker in Afikpo North LGA. Both coefficients are statistically significant with absolute t-ratios (-10.700 and 17.410) greater than the approximate 5% two-tailed critical value of 2.000. There is approximately 92% explanatory power according to  $R^2$  which indicates a good model fit.

There is no evident (positive) autocorrelation since DW (=2.102) is substantially above 2.000.

### **CONCLUSION:**

This work has demonstrated the use of linear probability model (LPM) to the

analysis of qualitative response regression model with the regressand or dependent variable having a binary or dichotomous variable. It also showed that the mean rate of income of a family in Afikpo North to owning a house is 12%, which is far from 100%. The estimated value from LPM applied in this work indicates that the probability of non-home owners to own a house in Afikpo North Local Government Area is 0.12 which is very far from 1. This could mean that the possibility of Ebonyi state worker, working under this current state of salary has a very low probability of owning a house out of his salary.

### **RECOMMENDATION:**

The researchers of this work, out of their findings hereby recommend that the state government should look into workers welfare and increase their salary. Also promotion for the staff should be conducted as at when due and the accompanying monetary entitlement paid on time.

## REFERENCES

Constantin Colonescu (2018). Using R for Principles of Econometrics: *A quick guide* (1<sup>st</sup> ed.)

James H. and Mark W. (2015). Introduction to Econometrics (new ed.), *Books.google.com*

Christian K. and Achim Z.(2008). *Applied Econometrics with R*. Hill Griffith and Lim (1996). *Principle of Econometrics* (2<sup>nd</sup> ed.) Wiley Books G.S. Maddala et al (1986). *Limited Dependent and Qualitative variable in Econometrics*. (2<sup>th</sup> edition).

John H. and Forrest D. (1984). *Linear Probability, Logit and Probit Models*, (new ed.), *Books.google.com*