Journal of Agricultural Economics and Development Vol. 4(7), pp. 113-119, December 2015
Available online at http://academeresearchjournals.org/journal/jaed
ISSN 2327-3151 ©2015 Academe Research Journals

Full Length Research Paper

Model misspecification and factors affecting land use decisions using linear probability model (LPM)

Ochalibe, A. I.*, Ogah, M. O. and Obo, V.

Department of Agricultural Economics, University of Agriculture Makurdi, Benue State, Nigeria.

Accepted 18 November, 2015

This study examined linear probability approach and misspecification of econometric model on land fallow decisions. It utilised primary data with sample size of 4187 agricultural households. The data were analyzed using linear probability model (LPM). Specification error test, robustness, goodness of fit and significance test were performed after estimation. Ramsey RESET test using powers of fitted values is significant at 5% (p>f=0.000), indicating that there is no serious signs of misspecification. The inktest (P=<0.05) is not significant indicating normality. In the same vein, diagnostic tests further indicated that Breusch-pagan/Cook-weisberg test for heteroskedasticity with the null hypothesis of constant variance was rejected (p>chi2=0.000). From the results, LPM passed only some of the tests suggesting that LPM may be less suited when dependent variable is dichotomous. The study revealed that distance and transaction costs along with socioeconomic and institutional factors significantly (p=>0.000) affect the decision to leave land fallow. It was concluded that the model was correctly specified, and the relevant variables in the economic relationship of land use decision were clearly identified with well defined pattern of relationship. It was recommended that a combination of econometric techniques like binary choice model as well as policy measures to ensure accurate estimates and reduce possibility of leaving agricultural land unutilised is required.

Key words: Linear probability model (LPM), distance, transaction costs, land use.

INTRODUCTION

The model provides the medium for expressing economic theory in a form easily ready for empirical analysis. In empirical analysis, different models are stated to measure a specified economic relationship and results are statistically tested, with stipulated criteria in order to gain some useful insight into the true characteristics of the economic relationship. One way in which a model can be specified is linear probability model (LPM). Linear probability model is Ordinary Least Square (OLS) applied to dichotomous dependent variable – that is the observable phenomenon to be explained can take only discrete, not continuous values. According to Ray (1991), models or theories that are at variance with observed behaviour should be revisited. Most of today's farmers are involved in the cultivation of crops and rearing of animals on few plots of land, and farming basically forms part of their livelihood. According to World Bank (2007), an estimated 2.5 billion of the 3 billion rural inhabitants are involved in agriculture: 50% of them are living in smallholder households and 27.7% of them are working in smallholder households. Empirical analysis of land use decision by small holder farmers has being variously modelled. Previous studies found mixed results. Economic and institutional structure, alongside general feeling of insecurity has been found by Sauer et al. (2011) as the factors that influence decisions to leave land fallow. However, according to Hamer (2008) land fallow periods were traditionally used by farmers to maintain the natural productivity of their land. Latruffe et al. (2009) argued that the main barriers for full integration of household in the market, influencing land use, are related to transaction costs and imperfections in land and labour markets. Misspecification of model occurs when either some relevant variables are included, some irrelevant variables are omitted from a given model and

*Corresponding author. E-mail: alexee2007@yahoo.com. Tel: +2348101342902, +2348053362445.

or using a linear model instead of a non-linear model. Given the nature of agricultural data, it is important to explore a model technique to affirm or the otherwise the applicability of recent analytical technique used especially the LPM model.

Problem statement and significance of the study

Observation shows that in economic analysis of farmers, relevant variables are most often excluded from models. This has its attendance consequences of heteroskadasticity with biasedness in the resultant estimate. However, various studies especially those related to land utilization hardly carry out post-estimation test of misspecification of model. Discovering it and attempting to remedy it is often difficult. These often result in undefined pattern of relationship. There are empirical evidences showing that farmers still leave their land unutilized but the question is whether model specification constitutes a serious concern in empirical analysis of agricultural house data. Is land left fallow due to the interplay among various factors, for example, household endowment, socioeconomic characteristics or due to factors that determine the heterogeneity in households behaviour like plot size, number of plots operated? Several studies have been carried out using different models - from regression model to non-linear models approach, but none has being done using LPM particularly in land use decision for agricultural purposes. This study is important because the ability to understand and predict changes in land use pattern is essential for agricultural development.

Research questions

1. What are the factors that affect farmers' land use decision using LPM?
2. Is the model heteroskadastic?
3. Are relevant variables omitted estimating land use decision using the method of LMP?

Aims and objectives

The broad objective of this study is to examine LPM model and misspecification of model on what drives the farm household decision to utilize land or not. Specifically, the study aims at the following:

1. Examines the factors affecting farmers' land use decision using LMP.
2. Estimates the test of homoscadasticity of LPM in land use decision.
3. Estimates misspecification test of LPM on land use decision.

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

According to Uwakonye (2007), investment in land improvement and increases in productivity are hindered by land decisions. Equitable access to productive opportunities on the land and security of such access once gained may help farmers to utilise more land for cultivation. Various studies have been carried out on determinants of land use. These studies generally found economic and socio-economic characteristics as factors that influence land use decision. For example, Mmopelwa (1998) examined the proportion and factors causing fallowing in Botswana. The study revealed that farmers left land uncultivated either for it to regain soil fertility and/or due to biophysical, social and economic factors. Using the binomial probit model, Grisley and Mwesigwa (1995) investigated the socio economic factors that influence seasonal fallowing in Kigezi highlands. The study found out that households' size contributed an average of 26% of land under fallow and land fragmentation was highly associated with the land fallowing decision. Economic and institutional structure, alongside general feeling of insecurity has been found by Sauer et al. (2011) as the factors that influence decisions to leave land fallow.

The distance to the plots and markets will also result in differences in labour costs, transport costs related to farmers output in terms of moving them to the market and even risks of not having access to market which will affect land use decisions (Sauer et al., 2011). Bergeron and Pender (1999) noted that a new road may increase farmers' incomes and their incentive to invest in land and suggested that the final outcome is premised upon the interplay among three levels of determination: the community, the farm, and the plot. According to Barret (2008), the central role played by physical (for example, roads) and transaction costs (for example, market information services) is often under-appreciated in economic analysis of market related behaviour. Therefore, efforts at investigating distance and transaction costs cannot be overemphasized.

The decision of farm household in land use is normally due to the interplay between various factors (Bergeron and Pender, 1999); these could be human capital like age, sex, educational qualification size of household as well as household consumption needs. According to Gosmani et al. (2010), the average education is generally low in Kosovo with just 8.5% of them having a university degree and a proportion of 84.6% have completed secondary/high school education. This has the potential to influence the decision to leave land fallow. Household farmers who are educated are expected to have gain useful skills as it relates to improved methods of land management. They have the easy access to loans that could help them acquire labour and capital to improve the land but they may be constrained by institutional factors. They are also capable of diverting the resources to non-agricultural activities perceived to offer higher returns and thus compounding liquidity constraints. Furthermore, the age of farm household can influence land use decisions. According to Bergeron and Pender (1999), older farmers with more experience are likely to have more opportunity to dispose of unwanted plots and may even have accumulated more wealth to invest in land thereby leaving less land fallow. On the other hand, they may have acquired many plots and found that they cannot adjust their land holdings as quickly as other factors like labor or capital have changed thus resulting in more land left fallow. Therefore, a priori expected age as well as education are not certain.

Institutional factors like land tenure may influence farmers' decision to leave certain piece of land uncultivated. Chisholm et al. (1997) noted that uncertainty in land ownership usually leads to fragmentation of land into uneconomic size, reduction of farm sizes making cultivation of land very difficult. This portends uncertainty about the use of land. It may result in low or decreased agricultural productivity and consequently, low utilisation of land. Discussions on land tenure literature have always being on the hypothesis that tenure insecurity impacts negatively on the probability of investing in land improvements. Private land ownership is considered as failing to provide farmers with adequate incentives or means to make land improvements or adopt new technologies that could enhance more use of land.

Liquidity constraint can hamper households from the use of land. According to Bergeron and Pender (1999), lack of credit may constrain adoption of new technologies and land use change. Where there is insufficient income along with costly access to farm inputs and outputs, the incentive to use more land would be lacking thereby hindering farmers from fully utilizing agricultural land. So, increase in farmers income should decrease the probability of land fallow indicating that increase income (high output) gives farmers the incentive to use more land as it means more investment in farm activities. However it is also possible that farmers may invest in other productive activities like livestock production as income rises or even in non-agricultural ventures which is believed to yield more returns per investment compared to agricultural sector (Bergeron and Pender, 1999).

Theoretical framework

Theoretically with LPM, the ordinary least square estimation will appear as follow:

$$Pr(Y = 1) = \beta_1 + \beta_2 X_i + u_i \qquad (1)$$

Where X_i is the vector of independent variable and u_i is the error term. From the above representation, the interpretation of the coefficients is straightforward: A unit increase in X is associated with a β increase in the probability of an event occurring. The relationship is linear so the impact of X on y is constant. However the disadvantages far outweigh the advantages, for example the variance of u_i is a function of the value of X. Hence, OLS estimator of LPM is inefficient and standard errors are biased resulting in incorrect hypothesis tests t & F statistics which are invalid (Gujarati, 2003). Linear probability model has a lot of limitations: firstly, heteroscedasticity of error term; secondly, the non-normality of the disturbance term; thirdly, the possibility of having estimate lying outside [0, 1]; and fourthly, low R^2. However, with large sample, the non-normality can be minimised and weighted least square can be used to resolve heteroscadasticity issue (Gujarati, 2003). Even at that the basic problem is, LPM assumes that the Pr=E(Y=1|X) increases linearly with X, that is, the marginal effect remains constant throughout. This is so whether or not the level of a particular variable changes. This, according to Gujarati (2003) makes the use of LPM very unattractive. However, its simplicity in use makes it very relevant in empirical analysis. Additionally, since the null hypothesis of homoscadasticty can be tested, heteroscadasticity which is usually a penalty for fitting wrong model is not likely to be a problem in this analysis.

Hypotheses

1. Null hypothesis H_0: The model specified is homoscedastic (constant variance)
2. Null hypothesis H_0: Farmers are not influenced by transaction costs to leave land fallow.

If we can reject the null hypotheses above at a sufficiently statistically significant level of t-values ($p<0.05$), then it is an indication that the variables are significant and the alternative hypothesis will be taken.

METHODOLOGY

This study used primary data from a survey carried out by the Statistical Office of Kosovo (SOK) - Agricultural Household Survey, 2005. The choice of this country results from the opportunity to use a complete dataset of Agricultural Household Survey. The data contain land utilization and output data and agricultural households' perceptions of barriers to land use with sample size of 4187 agricultural households. The data contain binary dependent variable (farmers leave land fallow=1 or not=0) and it is not observationally censored but rather is defined only over the interval [0, 1]. Hence a binary choice model is chosen for analysis (LPM). The limited dependent variable, dummy=1 if land fallows, otherwise=0 was regressed over a set of explanatory variables using linear probability model (LPM). This was motivated by the fact that individual model may not be able to account for the specificity of the data distribution. The null hypothesis of homeskadasticty will be tested, using estest ovtest in *stata*, heteroskadasticity which is usually a penalty for fitting wrong model is not likely to be a problem in this analysis. Descriptive statistics and analysis such as frequency distribution, percentages were used to describe the distribution of farm household characteristics in the study area and to obtain information averages, standard deviations, proportions and other tabular results. Given the composition of the dataset, some likely problems were envisaged, for instance, missing of some observations; this was mitigated by cleaning up the dataset. Consequently before estimation the model was checked and diagnosed to ensure that the model is correctly specified. For instance, according to Greene (2000), if x_2 is omitted from a model containing x_1 and x_2, (that is, $\beta_2 \neq 0$) then plim $\hat{\beta}_1 = c_1\beta_1 + c_2\beta_2$, where c_1 and c_2 are complicated functions of the the unknown parameters, the coefficient on the included variable will be inconsistent. Similarly, if the disturbances in a given regression is heteroscedastic, then the maximum likelihood estimators are inconsistent and the covariance matrix is inappropriate. The Stata command *linktest* was used to carry out misspecification test, after the probit command.

Expected sign of explanatory variables

The expected signs of all the explanatory variables are shown in Table 1. The sign as shown in Table 1 represents a decrease if negative (-) and an increase if positive (+) in the coefficient of parameter estimates. Based on the theoretical framework, the expected sign of coefficient of the variables are as indicated: If we assume that land fallow Y is dependent on a number of variables as discussed and it is normally distributed with the same mean and variance; the model for estimating whether or not the land fallows is:

$$\begin{aligned}\Pr[falw = 1|Xi] &= \emptyset(\beta_0 + \beta_1 h_{se} + \beta_2 age + \beta_3 n_{plt} + \beta_4 \mathrm{sm}_{sz} + \beta_5 t_{arl} \\ &+ \beta_6 \mathrm{RND}_s + \beta_7 lfmi_{ha} + \beta_7 t_{lbor} + lab_w + \varepsilon_i)\end{aligned}$$

Where $Pr[falw = 1|X_i]$ means the probability that the land will be fallowed given the determinants *Xi* as stated and Ø is the cumulative standard normal density of the distribution, β is the coefficient and δ is the coefficient of categorical variable education[1].

DISCUSSION OF RESULTS

The variables, number of plots *(n_plt)*, total arable land *(t_arl)*, farm income *(lfmi_ha)*, total labour *(t_lbor)*, rented plots from state (R_ND_S), as well as labour per unit land area cultivated *(lab_w)* were individually statistically significant at 1% level, while age of farm household (age), size of the smallest plot size (sm_sz), the dummy variable (*distance)*, rented plots from private individual (*R_NDI)* as well as the constant were both significant at 5% level. However, household size (h_se) as well as categorical variable education (educ) were not statistically significant.

In order to interpret the estimates from LPM, it must be noted that a change in an independent variable changes the probability of dependent variable, *fawl*=1 by a constant amount. For example, the results from the LPM

[1] δ_{18}educ2 is omitted from the equation because it is used as the reference category

Table 1. Expected sign of explanatory variables.

Variable	Definition	Sign
h_se	Household size	-
age	Age of farm household head in years	?
n_plt	Number of plots	+
sm_sz	Size of the smallest plot size in (ha)	+
tar_l	Total arable land in hectares (ha)	-
lfmi_ha	Log of farm income per hectare in Euros	-
l_vfequip	Log of value of equipment per hectare in Euros	-
irrg_lnd	Irrigated area of land in (ha)	-
t_lbor	Total labor (total number of workers)	-
R_LNDI	Rented plots from private individual %	?
R_LNDs	Rented plots from private state %	?
distance	Dummy=1 if municipality is farther from high way	+
Educ	Education (categorical)	?
lab_w	Labour per land area	?

revealed that holding all other variables fixed, a rise in farm income decreases the probability of land fallow by 0.8% *ceteris paribus*. However, an increase in the age of household head is associated with 0.15% increase in the probability of leaving land fallow *ceteris paribus*. Transaction proxies, number of plots and smallest size of plot were positively related to land fallow decisions. An additional number of plots increases the probability of land fallow by 3.5%. In the same vein, size of the smallest plot size increases the probability of land fallow by 13.1%. However, factor endowment variables decrease the probability of land fallow. For instance, an increase in total arable land*,* total labour and irrigated area of land accounted for approximately 3.1%, 2.7% and 0.27% decrease respectively in the probability of land fallow *ceteris paribus*. The dummy variables *distance,* is associated with approximately 3.8% higher land fallow compared to those in the municipalities with highways *ceteris paribus*. Institutional factors like rented plots from private individual and rented plots from state as well as labour per farm household accounted for approximately 2.1% and 2.9% decrease respectively in the probability of land fallow *ceteris paribus*. Table 3 shows that Ramsey RESET test using powers of fitted values is significant at 5% (p>f=0.000) indicating that the model is correctly specified and there is no omitted variable, and no irrelevant variable was included. The diagnostic tests, as shown in Table 4, further indicated that Breusch-pagan/Cook-weisberg test for heteroskedasticity with the null hypothesis of constant variance is rejected (p>chi2=0.000).

The interpretation of the LPM above highlights one basic problem with LPM. It assumes constant marginal effects (for example, *age educ n_plt* etc). For instance, as shown in Table 2, an additional year of age is estimated to increase the probability of land left fallow by approximately 0.15%, *ceteris paribus*, regardless of how many plots the household was initially operating and regardless of other levels of other explanatory variables like number of plots and education. This is better imagined than real. One would expect that the probability is non-linearly related to year of age. At a very small age, of say under 16 years, a household will not leave land fallow (cannot even own a land let alone leaving a land fallow), but at working age of say 16-64 years (from the summary statistic in Table 1), it is most likely that a land will be left fallowed. Any increases beyond 90 years, which is the maximum age of household, will have little effect on the probability of leaving a land fallow. Thus at both end of the distribution, the probability of leaving a land fallow will be virtually unaffected by a marginal increase in years of age.

Another related issue is that LPM violates the assumption of normal distribution of the error term and heteroscadasticity condition, meaning the variance of the disturbance is not constant. As a result, OLS is inefficient and the t and F statistics is generally invalid. Perhaps this explains why the LPM failed the normality test (skewness and kurtosis) and heteroscadasticity test (Table 5). This suggests that LPM models might not be suitable for estimation when the dependent variable is dichotomous but may only provide an insight into a binary regression model.

CONCLUSION AND RECOMMENDATIONS

The study econometrically examined what drives the farm household decision to leave land fallow or not using Kosovo as a case study. The study found that farm households are driven by the influence of combination of factors to leave their land fallow in Kosovo. The test for homoscadasticity given the null hypothesis that the model

Table 2. LPM estimate of land fallow decision.

Dependent variable *falw*			
Independent variable	**Parameter estimate**	**Robust standard error**	**P>|t|**
h_se	- 0.001	0.0015	0.511
age	0.0015	0.0006	0.007
n_plt	0.035	0.003	0.000
sm_sz	0.131	0.040	0.001
t_arl	-0.031	0.006	0.000
lfmi_ha	-0.008	0.002	0.000
l_vfequip	-0.004	0.002	0.019
irrg_lnd	-0.0027	0.005	0.579
t_lbor	-0.027	0.006	0.000
R_LNDI	0.021	0.009	0.018
R_LNDs	0.029	0.006	0.000
distance	0.038	0.014	0.006
Lab_w	0.0005	0.0001	0.000
educ0	-0.015	0.091	0.867
educ1	-0.018	0.019	0.338
educ3	0.029	0.021	0.166
educ4	-0.004	0.015	0.800
constant	0.017	0.024	0.479
Number of observation =		4187	
R-squared =		0.091	p>F=0.000

Note on categorical variable; *

Table 3. Misspecification test.

Estat ovtest
Ramsey RESET test using powers of the fitted values of falw
H_0: Model has no omitted variables
F (3, 4166) = 8.23
Prob > F = 0.0000

Table 4. Heteroskedasticity test.

Estat hettest
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
H_0: Constant variance
Variables: fitted values of falw
chi2 (1) = 307.89
Prob > chi2 = 0.0000

is homscadastic is rejected at (p>0.005) level of significance. Additionally, since the variables included are statistically individually significant at p=<005 from their respective t-values, we reject the null hypothesis that these variables are not statistically significant. The model was correctly specified, and the relevant variables in economic relationship of land use decision were clearly identified with well defined pattern of relationship. Thus, this econometric technique (LPM) that is well specified is recommended for studies of this nature. Given the issues raised, institutions concern with land tenure arrangement should be strengthened, and credit programmes should

Table 5. Normality test.

Shapiro-Wilk W test for normal data						
Variable	**Obs**	**W**	**V**	**z**	**Prob>z**	
dis	4187	0.73128	621.387	16.782	0.00000	
Source	SS	df	MS		Number of obs = 4187	
Model	57.222003	2	28.6110015		F (2, 4184) = 206.85	
Residual	578.724976	4184	0.138318589		Prob > F = 0.0000	
					R-squared = 0.0900	
Total	635.946979	4186	0.151922355		Adj R-squared = 0.0895	
					Root MSE = 0.37191	
falw	Coef.	Std. Err.	T	P>|t|	[95% Conf. Interval]	
_hat	0.7910148	0.096452	8.20	0.000	0.6019177	0.980112
_hatsq	0.4430757	0.1754453	2.53	0.012	0.0991098	0.7870416
_cons	0.0176143	0.0129353	1.36	0.173	-0.0077457	0.0429743

be encouraged to increase liquidity. A combination of policy measures that will reduce the possibility of leaving agricultural land fallowed is required.

REFERENCES

Bergeron G, Pender J (1999). Determinants of land use change: evidence from a community study in Honduras. Environment and Production Technology Division Int. Food Policy Res. Inst. (EPTD) Washington. Discussion paper no. 46.

Chisholm M (1995). Rural Settlement and Land Use. London: Hutchinson University Library.

Grisley W, Mwesigwa D (1995). Socio Economic Determinants of Seasonal

Fallowing. J. Environ. Manage. 42: 81-89.

Gujarati D (2003). Basic Econometrics, 4th Edition, McGraw-Hill.

Latruffe L, Davidova S, Desjeux Y (2009). Perpetuation of Subsistence Farming in Kosovo: The Role of Factor Market Imperfections, Paper presented in the mini-symposium organised by IAMO, 'Structural change in Europe's rural regions – farm livelihoods between subsistence orientation, modernisation and non-farm diversification' (27th Congress of the International Association of Agricultural Economists (IAAE), Beijing, China, 16–22August.

Hamer E (2008). http://www.the ecologist.org/fallow and fertile.html

Mmopelwa G (1998). Factors Contributing to Land Fallowing in a Permanent Cultivation System. J. Arid Environ., 40 (2) 211-216.

Ray P (1991). Economics For Professional and Business Studies. DP Publications Limited, London. ISBN 0-905435-97-4.

Sauer J, Davidova S, Lautruffe L (2011). Determinants of Smallholder's Decisions to Leave Land Fallow. J. Agric. Econs.

Uwakonye N (2007). An Econometric Analysis Of Two Possible Land Reform Strategies In Nigeria. J. Bus. Econs. Res., Vol. 5(3).

World Bank (2007). Kosovo Poverty Assessment. Report No. 39737-XK (Poverty Reduction and Economic Management Unit, Europe and Central Asia Region.