Modelling farmers' land use decisions

NICHOLAS C. BALTAS [1] **and** **OURANIA KORKA** [2]

This paper aims at analyzing and predicting farmers' land use decisions. For this purpose, farmers' decision processes are described by a nested discrete choice model. At all levels of estimation, the values of the dependent variable indicate the proportion of land allocated to each agricultural activity. Then, the share of land assigned to each activity is estimated by employing the random utility model of the logit specification.

I. INTRODUCTION

This study presents a framework for analyzing land use allocation under risk. A discrete choice model of acreage decisions that includes price uncertainty was developed and applied to the choice among land use alternatives. While farmers' acreage decisions are usually estimated in a linear regression framework, in this paper they are estimated within a stochastic utility theoretic model. Conceptualizing land allotment as a two-stage process allows the acreage decisions to be characterized in a novel way.

Acreage decisions are analyzed separately based on different irrigation states. At the first decision level, under the irrigated state, the major agricultural activity such as planting annuals or starting a tree orchard is first selected. At the second decision level, once the choice of a major agricultural activity has been made, the individual chooses a particular annual or perennial crop. Under the non-irrigated state, at the first level, the major agricultural activity of starting/retaining livestock or a tree orchard is selected and at the second level, the decision concerning varieties of olive trees and grazing and non grazing animals is taken.

The paper is organized as follows. The next section presents the modelling of individual decision processes. In the third section the employed data is discussed. In the fourth section the empirical results are analyzed. Finally, the paper closes with some concluding remarks.

II. THE ESTIMATED MODEL

Estimation of the entire decision process is handled by a nested logit model (McFadden, 1983). This model breaks down simultaneous choices into a number of sequential steps and it is based upon the concept of feedback mechanisms from lower-to higher-level decisions through inclusive values. Let the probability of an individual choosing a specific crop or animal j (i.e. alfalfa, corn, lemons, oranges, olives, olive-oil, sheep, goats, and hogs) among N_i number of alternatives available in activity i, (i.e. annuals, trees and livestock) be $P_{j/i} = P_{ij} / P_i$. Conditional probability can be written as

[1] Professor of Economics and Jean Monnet Chair, Department of Economics, Athens University of Economics and Business, 76, Patission Str., 104 34 Athens, Greece, E-mail: baltas@aueb.gr

[2] Research Associate at the Research Centre, Athens University of Economics and Business.

$$P_{j/i} = \frac{e^{\frac{1}{1-\sigma}(V_{ij})}}{\sum_{k=1}^{Ni} e^{\frac{1}{1-\sigma}(V_{ik})}} = \frac{e^{\frac{1}{1-\sigma}(\beta' X_{ij})}}{\sum_{k=1}^{Ni} e^{\frac{1}{1-\sigma}(\beta' X_{ik})}} \tag{1}$$

and marginal probability

$$P_i = \sum_{j=1}^{Ni} e^{V_{ij}} / \sum_{m=1}^{c} \sum_{n=1}^{Nm} e^{V_{mn}} , \tag{2}$$

where σ is an estimated parameter and V stands for the utility function. We define $V_{ij} = \beta' X_{ij} + \alpha' Y_i$, where X_{ij} is the vector of observed characteristics associated with alternative j in activity i and Y_i is the vector of characteristics associated only with activity i; α and β are vectors of unknown parameters. Then, following Maddala (1983), the inclusive value I_i is defined as:

$$I_i = \ln[\sum_{j=1}^{Nj} \exp(\beta' X_{ij}/(1-\sigma))] , \tag{3}$$

where I_i is an index of variables which is generated from the estimation of the model's second level. Given this definition, marginal probability (3) can be rewritten as

$$P_i = e^{(\alpha' Yi + (1-\sigma)Ii)} / \sum_{m=1}^{c} e^{(\alpha' Ym + (1-\sigma)Im)} . \tag{4}$$

The model would be estimated by first estimating the parameters from the conditional logit model (1), calculating the values of the inclusive value in (3), and then using these in estimating the parameters from the marginal logit model (4).

The utility, V_{ij}, obtained from each alternative choice is assumed to be some function of wealth (W), which is comprised of an initial level of wealth (W_o) and the random returns from the agricultural alternative (R). The Taylor series expansion of this utility function around the expected value, W_o + E(R), is given by

$$U(W) = U(W_0 + E(R)) + \sum_{k=1}^{\infty} [U_k (W_0 + E(R)) \frac{(W - W_0 - E(R))^k}{k!}] , \tag{5}$$

where U_k is equal to the k^{th} derivative of the utility function with respect to wealth W. Taking the expectation of (5), the expected utility is

$$\begin{aligned} E[U(W)] = U(W_0 + E(R)) + (1/2)U_2(W_0 + E(R))M_2(R) + \\ + (1/6)U_3(W_0 + E(R))M_3(R) + \ldots, \end{aligned} \tag{6}$$

where $M_k(R)$ is equal to the k^{th} moment of R around the mean. Assuming that terms beyond those involving the second moment add insignificantly to the precision of the approximation, the expected utility can be written as

$$E[U(W)] = U(W_0 + E(R)) + (1/2)U_2(W_0 + E(R))V(R) \ , \tag{7}$$

where V(R) denotes the variance of returns.

In order to apply this approach and to incorporate uncertainty in the discrete choice model an expression which relates the decision problem of individuals with their risk preferences is required. Let the utility function be of a semi-logarithmic form; U = ln(W). Then, the first derivative, $U_1=1/W>0$, implies a monotonically increasing function of wealth and the second derivative $U_2=-1/W^2<0$ implies that marginal utility decreases with increasing wealth. Making the appropriate substitutions on (7), expected utility can be approximated by

$$E[U(W) \cong \ln[W_0 + E(R)] - \frac{1}{2}\frac{V(R)}{[W_0 + E(R)]^2} \ . \tag{8}$$

Note that a positive coefficient on the first term implies that utility function is increasing in expected wealth and a negative on the second shows that utility function is decreasing in the variability of wealth. A log formulation and the above characteristics ensure decreasing absolute risk aversion (Anderson, Dillon, Hardaker; 1977).

Acreage decisions involve uncertainty about prices since there exists a natural time lag from initial productive activities to final output. It is assumed that farmers' current expectations about the product price is given by:

$$p_t^* = E(p_t/\Omega_{t-1}) \ , \tag{9}$$

where E_{t-1} denotes the expectation formed given all the information available at t-1, i.e. Ω_{t-1} and p_t is the price in period t. It is postulated that expected prices follow an ARMA procedure because they build their expectations about future prices according to the difference between the prices obtained and intervention prices, whenever they exist,

$$\begin{aligned} p_t^* &= \Theta_0 + \Theta_1\bar{p}_t + \Theta_2(p_{t-1} - \overline{p_{t-1}}) + \ldots + \Theta_q(p_{t-q} - \overline{p_{t-q}}) + \\ &\varepsilon_1 + \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + \ldots + \Phi_p\varepsilon_{t-p}, \end{aligned} \tag{10}$$

where p denotes actual prices, $\bar{p}$ supported prices (i.e. withdrawal or intervention price), the Θ's and the Φ's are the coefficients of the autoregressive and moving averages respectively and ε_t is white noise.

The estimated expected prices are then introduced to the following equation to calculate the variance of the price

$$var(p_t) = E[p_t - p_t^*]^2 \ . \tag{11}$$

Expressions (10) and (11) are further employed to calculate the expected returns, E(R), and the second moment of returns, V(R), as follows

$$E(R) = p_t^* * y_t + Inc_t - C_t \, V(R) = var(p_t) * y_t^2 \ , \tag{12}$$

where Inc_t, C_t and y_t denote the direct income support, the cost and the yield of the product per land unit in time t respectively.

Substitution of specifications (12) into expression (8) gives the expected utility of a specific agricultural choice

$$E[U(W)] \cong \ln[W_0 + p_t^* * y_t + Inc_t - C_t] - (\frac{1}{2})\frac{var(p_t) * y_t^2}{[W_0 + p_t^* * y_t - Inc_t - C_t]^2} . \tag{13}$$

By incorporating expression (13) into (1), the conditional probability takes the following form

$$P_{j/i} = \frac{\exp\frac{1}{1-\sigma}[\beta_1 \ln(W_{0j} + p_j^* * y_j + Inc_j - C_j) + \beta_2(\frac{1}{2})\frac{var(p_j) * y_j^2}{[W_{0j} + p_j^* * y_j + Inc_j - C_j]^2}]}{\sum_{k=1}^{Ni}\exp\frac{1}{1-\sigma}[\beta_1 \ln(W_{0k} + p_k^* * y_k + Inc_k - C_k) + \beta_2(\frac{1}{2})\frac{var(p_k) * y_k^2}{[W_{0k} + p_k^* * y_k + Inc_k - C_k]^2}]}, \tag{14}$$

where β are the estimated parameters.

Then, the marginal probability (4) of choosing an activity i, under irrigated state, takes the following form

$$P_i = \frac{\exp[\alpha_1 Intplnt_i + \alpha_2 L\alpha g\alpha c_{ij} + \alpha_3 L\alpha g\alpha c_{ij'} + (1-\sigma)I_i]}{\sum_{m=1}^{c}\exp[\alpha_1 Intplnt_m + \alpha_2 L\alpha g\alpha c_{mj} + \alpha_3 L\alpha g\alpha c_{mj'} + (1-\sigma)I_m]} \tag{15}$$

and under non irrigated,

$$P_i = \frac{\exp[\alpha_1 Fnlrt_i + \alpha_2 L\alpha g\alpha c_{ij} + \alpha_3 L\alpha g\alpha c_{ij'} + (1-\sigma)I_i]}{\sum_{m=1}^{c}\exp[\alpha_1 Fnlrt_m + \alpha_2 L\alpha g\alpha c_{mj} + \alpha_3 L\alpha g\alpha c_{mj'} + (1-\sigma)I_m]} , \tag{16}$$

where *Intplnt* denotes the cost of planting a tree orchard, $Lagac_j$ is the percentage of land allocated to choice j in last period, and *Fnlrt* denotes the returns from selling a livestock herd or a tree stock.

For the estimation of parameters β, α and σ of the nested model (14), (15) and (16) the two-stage sequential approach is employed. First, the parameters â of the conditional model (14), and then, in the second stage, after the inclusive values, σ, of the alternatives in each activity are calculated, the parameters α of the marginal probabilities (15) and (16) are estimated. According to McFadden (1983), a sufficient condition for a nested logit model to be consistent with stochastic utility maximization is that the coefficient of the inclusive value should fall into the unit interval.

III. DATA

The data consisted of 76 villages of the the Amvrakikos Gulf Area (AGA), an important agricultural region which lies in Northwestern Greece, and are grouped by crop and animal per village. The sample includes the farms which are listed by the National Statistical Service in every village. Farms have been aggregated and choices in every village are represented by the number of land units allocated to each choice. All choices are considered available to every farm.

The analysis of this study is based on the 1989/90 farming season. For this season, output, producer prices, and direct income support for each crop have been collected. For the same period, the number of animals, stocking rates, private and public pasture land acreage, producer prices, headage payments and meat output are also available. In addition, information concerning public investment for crop and livestock farming (irrigation and drainage projects, abattoirs and pasture land) and agricultural cost data (seeds, the quantity and value of labour hours for people and machinery, fertilizer and pesticide rates and the land value per stremma) has been collected by the Agricultural Bank of Greece. Acreage (irrigated and non-irrigated) of crops as well as acreage of private and public pasture land has been collected for two consecutive years (1988 and 1989).

IV. ESTIMATION AND EMPIRICAL RESULTS

The empirical analysis is based on the assumption that any given land unit is allocated to only one crop or animal. Regardless of the irrigation state, the number of alternative choices available to farmers are four. Usually, in the discrete choice framework, the observed dependant variable consists of an indicator reflecting the respondent's most preferred alternative. However, in this case, there is no access to individual farm - level data. Thus, at all levels of estimation, the data have been grouped and the values of the dependant variable indicate the proportion of land allocated to each agricultural activity in each village. Use of aggregate proportions instead of the traditional indicator produces the same results but different standard errors (Greene, 1990). The definitions of the variables used as arguments in the utility function of the model are presented in Table 1.

As a first step to the estimation procedure, the estimated coefficients of the price function from each agricultural product are shown in Table 2.

In the second level on the decision tree (3.1), the probability of a specific irrigated crop (i.e. alfalfa, corn, lemons or oranges) to be chosen is predicted. For each of the variables the estimated coefficients were found to be significant. The positive coefficient for the first variable indicates that as the expected returns from alternative i increases relative to that of the other alternatives, the probability of allocating more land to alternative i increases relative to the other alternatives. The negative coefficient of the second variable, the variance of returns suggests that farmers allot their land to crops with less variation around their respective expected returns, reflecting Greek farmers' risk aversion. The test statistic for the joint hypothesis that all coefficients are zero is significant at the 90 percent confidence level. This estimation postulates the assumption that land is allocated to each crop with equal probability. There are no universally accepted measures of goodness of fit in discrete choice models. One measure[3] suggested by Greene. The value of this statistic is 0.67.

In the next stage of the estimated model which predicts how the probability of allocating land to annual or tree crops is affected by economic and lagged acreage variables (3.2). *Intplnt* is negative and statistically highly significant. This suggests that land is more

---

[3] LRI = 1-(lnLu/lnLr), where lnLu and lnLr are the maximized values of the log-likelihood from the unrestricted model and from the model in which all slopes are zero, respectively. The measure is bounded in the unit interval and LRI increases as the fit of the model improves.

likely to be allotted to the crop with the lower planting cost and with less restrictions regarding the future cultivations. The variables *Lagann* and *Lagper* are intended to capture the effect of the last period's allocation of acreage to annual and tree crops respectively. In other words, we try to capture the notion of habit persistence (Heckmann, 1981). Both coefficients are statistically significant and positively related to the probability of repeating last year's land use. The last variable of the estimated model is the inclusive value, which is significant and justifies the nested logit structure of the model. This, in turn, means that there are similarities among specific annual crops or citrus trees but sufficient differences between the major crop groups. The model, with likelihood value L_i, postulates that farmers formulate their decisions concerning the allocation of their land without taking into account the lagged allocated acreage. The inclusion of variables *Lagann* and *Lagper* make a significant contribution to the predictive power of the model (LRI = 0.52).

Estimates of the non irrigated land shares between the alternative choices, which include olive trees cultivated for olive oil or olive, grazing animals (i.e. sheep or goat) and non grazing animals, hogs, show that the signs of both coefficients were found as *a priori* expected (3.3). Again, this suggests that land is more likely to be allotted to the agricultural choice with the higher expected returns and the lower variability around the returns. The estimated coefficients were found significant and the LRI is equal to 0.71.

The results of the top stage of the estimated model, under the non irrigated state, in which the probability of land allocation to olive trees or livestock is predicted. All estimated coefficients were found significant. The sign of the first coefficient indicates that it is more likely for land to be allotted to the major agricultural activity which has the higher returns from selling the tree stock or the animals. The second and third parameter are introduced to capture the impact of the last period's decision concerning land allotment to current choices are statistically significant and positively related to the probability of retaining the same agricultural activity. Thus, the hypothesis of no habit persistence can be rejected. Lastly, the inclusive value which is generated by the second level of estimation falls in the unit interval justifying the nested structure. The value of the LRI is equal to 0.53.

V. CONCLUSIONS

The empirical results reflect some important aspects of farmers' behaviour. They indicate that risk and returns effects are important in land allocation decisions. They also provide empirical documentation of farmers' risk-averse behaviour and of a positive response to increases in direct income support and yield and a negative one to higher production costs. Lastly, they point out the effect of last period's land use on the current choice.

In general, the developed nested approach of decision making, in the context of agricultural activity choice, performed fairly well and established a framework within regional land allocation. However, the analysis was stymied by the limited data. More information regarding farmers and agricultural fields specific characteristics would shed more light on the role of risk and government support in farmers' decisions. For example, further research could consider the effect of individuals' sociodemographic characteristics on their attitude towards risk. The analysis could also incorporate past land use decisions on the current probability of choosing a specific agricultural product, in an attempt to determine preferences for one particular product over time.

Table 1 : Definitions of Variables For the Nested Logit Model

Variable	Definition
Expret	Expected returns (drs/stremma[a])
Varret	Variance of returns (drs/stremma)
Intplnt	Initial cost of planting a tree orchard (drs/stremma)
Lagann	Percentage of land allocated to annual crops in last period
Lagper	Percentage of land allocated to tree crops in last period
$Incl_{A/T}$	Inclusive value obtained from the estimation of the lower level decisions among irrigated crop choices
Fnlrt	Returns from selling a tree stock or a livestock herd (drs/stremma)
Lagolv	Percentage of land allocated to olive groves in last period
Laggraz	Percentage of pasture land in last period
$Incl_{OL/LV}$	Inclusive value obtained from the estimation of the lower level decisions among non irrigated choices

[a] 1 stremma is equal to 0.1 hectares.

Table 2 : Expected Price Functions Using ARMA Procedures [a]

Dependent Variable	Constant	ΔP_{t-1}	ΔP_{t-2}	$\overline{\Delta}^2 P_{t-1}$	$\overline{\Delta}^2 P_{t-2}$	$\overline{\Delta} P_{t-1}$	R^2	DW
ALHALFA								
$\Delta \hat{P}_t$	0.050 (0.065)	–0.622 (3.234)	–0.360 (2.197)				0.288	2.157
CORN								
$\overline{\Delta} \hat{P}_t$	–0.020 (0.359)				0.379 (2.531)	0.387 (2.744)	0.235	2.363
LEMONS								
$\overline{\Delta}^2 \hat{P}_t$	0.067 (0.116)			0.913 (9.641)			0.770	2.059
ORANGES								
$\overline{\Delta}^2 \hat{P}_t$	0.141 (0.748)					–0.998 (2.546)	0.690	2.018
OLIVE OIL								
$\overline{\Delta}^2 \hat{P}_t$	0.049 ((0.006)			–0.587 (2.618)			0.215	2.191
OLIVES								
$\Delta \hat{P}_t$	–1.050 (0.314)	–0.587 (5.664)	–0.816 (2.727)				0.572	2.041

[a] $\Delta P = P_t - P_{t-1}$

$\overline{\Delta} P_t = P_t - \overline{P}_t$

$\overline{\Delta}^2 P_t = \overline{\Delta} P_t - \overline{\Delta} P_{t-1}$

Figures in parentheses denote t-statistic.

Table 3 : Empirical Results

Variable	*Coefficient*	*t-statistic*	*Test*	*Statistic*	*Degrees of Freedom*
3.1 Logit Estimates of Irrigated Land Shares among Irrigated Crops (i.e. alfalfa, corn, lemon and orange trees)					
Expret	0.129-4	4.53	–2ln(Lr/Lu)	19.15	2
Varret	–1.662-8	8.79	LRI=0.67		
3.2 Logit Estimates of Irrigated Land Shares among Annual and Tree Crops					
Intplnt	–1.584-7	10.41	–2ln(Lr/Lu)	41.71	4
Lagann	0.587-4	17.70	–2ln(Li/Lu)	21.86	2
Lagper	0.218-4	6.35	LRI=0.52		
$Incl_{A/T}$	0.120	19.29			
3.3 Logit Estimates of Irrigated Land Shares among Non Irrigated Crops (i.e. olive trees cultivation and animals)					
Expret	1.0097	16.94	–2ln(Lr/Lu)	28.06	2
Varret	–2.1689	19.96	LRI=0.71		
3.4 Logit Estimates of Non Irrigated Land Shares among Olive Trees and Livestock					
Fnlrt	5.916-6	10.45	–2ln(Lr/Lu)	43.33	4
Lagolv	0.601-4	31.40	LRI=0.53		
Lagranz	2.058-7	15.43			
$Incl_{OL/LV}$	0.156	33.26			

Lr = value of likelihood at zero parameters
Lu = value of likelihood with all parameters estimated.
Li = value of likelihood with first and last parameters estimated.

ACKNOWLEDGEMENTS

The financial support of the European Communities is gratefully acknowledged.

REFERENCES

Anderson, J.R., Dillon J.L. Hardaker, B. (1977). *Agricultural Decision Analysis.* The Iowa State University Press, Iowa.

Greene, W.H. (1990). *Econometric Analysis*. MacMillan Publishing Company, New York.

Heckman, J. (1981). Statistical Models for Discrete Panel Data. In Manski, C.F. and McFadden D., *Structural Analysis of Discrete Data with Econometric Applications,* M.I.T. Press, Cambridge Mass.

McFadden, D. (1983). Quantitative Response Models. In Griliches, Z. and Intriligator, D.M. *Handbook of Econometrics*, North-Holland Publishing Company, Amsterdam.

Maddala, J.S. (1983). *Limited Dependent and Qualitative Variable in Econometrics,* Cambridge University Press, Cambridge.