MD

WP 89-41

WORKING PAPERS

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS
SYMONS HALL
UNIVERSITY OF MARYLAND
COLLEGE PARK 20742

[redacted]

AGGREGATION OF PROFIT-MAXIMIZING LAND USE DECISIONS FOR THE ANALYSIS OF NEW LAND USE POLICIES

by

Ian Hardie
Agricultural and Resource Economics
University of Maryland
College Park, Maryland 20742

Peter Parks
School of Forestry and Environmental Studies
Duke University
Durham, North Carolina 27706

Working Paper 89-41

March 1990

Funded in part by USDA Forest SErvice, Southeastern Forest Experiment Station, Project No. 01-4-33508.

AGGREGATION OF PROFIT-MAXIMIZING LAND USE DECISIONS FOR THE ANALYSIS OF NEW LAND USE POLICIES

Ian Hardie
Agricultural and Resource Economics
University of Maryland
College Park, Maryland 20742

Peter Parks
School of Forestry and Environmental Studies
Duke University
Durham, North Carolina 27706

Working Paper 89-41

March 1990

Funded in part by USDA Forest Service, Southeastern Forest Experiment Station, Project No. 01-4-33508.

AGGREGATION OF PROFIT-MAXIMIZING LAND USE DECISIONS FOR THE ANALYSIS OF NEW LAND USE POLICIES

Policies designed to change land use patterns often depend upon economic incentives to alter landowner behavior. Examples include farm land tax programs to preserve agricultural lands, cost subsidies to regenerate forest lands after harvest, and conservation payments to remove cropland from active use. Policies designed to affect farmer behavior can also alter land use. A new target price, for example, may change the acreages devoted to corn, soybeans and other farm crops. The distinguishing feature of these policies is that they elicit two interrelated responses, one from the landowner and one that manifests as a geographic change in the way land is used.

Analysts interested in measuring the effect of proposed land use policies tend to focus either on the landowner response (Royer, Esseks and Kraft) or on the acreage response (Alig and Healy, McIntosh and Skideed). There seems to be a tacit belief that if one is known, the other may be deduced. We believe, however, that this possibility is ruled out by the aggregation problem inherent in relating micro-production decisions to macro-production functions (Sato 1975). As in the case of capital, land is a heterogeneous input which may be used differently by owners who have the same economic objective and who face the same economic incentives.

In this paper, we will review this aggregation problem and indicate why it is present in land use policy analysis. A way to solve the problem using an area-frame sampling procedure will then be presented. Use of the sampling procedure to solve the aggregation problem is consistent with use of a discrete-continuous choice model to explain landowner behavior. One such model, a random-variable profit maximization model developed for multi-plant firms by McFadden (1981), is adapted to the land use problem in the middle sections of the paper.

The usefulness of the adapted model for land use policy analysis is then illustrated with an empirical study of the effect of government cost-sharing on the replanting of harvested southern pine land. The empirical study is somewhat unique in that it results from a rare instance in which the requisite area-frame sample data were collected and made available for economic analysis. The paper ends by drawing some conclusions about the usefulness of further developing the proposed approach.

The Aggregation Problem

To analyze how landowners react to a new economic incentive, we need to model micro-behavior. If a very high value is put on simplicity, we can postulate a landowner who produces y_i units of a product i, using land and f units of some other input. Under suitable product separability conditions (Chambers 1988), this landowner would have the micro-production function $y_i = g_i(a_i, f)$, where a_i is the acreage of land in use i. If y_i and f are homogeneous products, returns to scale are constant, profits are maximized, the markets for y_i and f are perfectly competitive, and a_i is less than or equal to some fixed stock of land, we can obtain a well behaved function that relates the quasi-rent of a_i to a change in either the price of product i or the cost of the nonland input. Given a similar situation for another product j, we could deduce how a change in an economic incentive that operates through the output price or input cost would affect land use; essentially the landowner would change land use whenever the economic incentive caused the quasi-rent from use i to exceed or to fall below the quasi-rent from use j.

One might think that such a stringent set of behavioral assumptions would make it simple to determine the effect of the economic incentive on aggregate land use in a region. Yet if land varies in soil fertility, topography, plot size and location, this aggregate change can be obtained only through laboriously adding acreages in a particular land use across micro-units. Sato (1975, pp. 4-5) shows that in this case, the region's aggregate production function must be written as $Y_i = G_i(a_{i1},\ldots,a_{iK},F)$, where $a_{i1},\ldots,a_{iK}$ are the land inputs of the K landowners in the region, F is the sum of profit-maximizing nonland inputs across landowners, and Y_i is the corresponding sum of outputs. This function results from the heterogeneity of land quality across owners. It implies that there will be K different quasi-rent functions and K potentially different responses whenever the economic incentive changes.

Analyses that begin with aggregate land quantities necessarily assume that the aggregate production function is $Y_i = G_i(A_i,F)$, where A_i is a measure of total acreage in use i in the given region. But Solow, Fisher and others have proven that this function is a true aggregate of the micro-functions only if (1) land is a homogeneous input, or (2) the land and nonland inputs are separable in production and the marginal products of the land remain in constant proportion as land use changes (Sato, Chapter 1). The infeasibility of this formulation is illustrated by noting that when land is heterogeneous, all landowners are required to produce all outputs in all situations. Else marginal productivity would be positive for some uses and zero for others, violating the proportionality requirement.

Area Frame Statistics

In an area-frame sample, a given region is divided into blocks of known acreage. A probability sample is selected from these blocks on the basis of physical land characteristics. All tracts of land within the sample block or "segment" that are owned by members of the population of interest are then located. Multiplication of the acreage in these tracts by expansion factors derived from the sampling procedure produce total acreage estimates for the region.

Since owners are surveyed, owner characteristics, attitudes and choices can be recorded. Domain estimation then produces descriptive statistics of the form (Fecso et al.):

$$S_i = \sum_{k=1}^{K} N_k(i) \cdot A_k \cdot M_k$$

where i represents a particular characteristic, attitude or choice, and k represents the k^{th} individual surveyed in a sample of K landowners. The indicator $N_k(i)$ equals one if the owner has the characteristic, attitude or makes the choice i, and equals zero otherwise. A_k is the acreage owned in the sample segment by the k^{th} owner, and M_k is the expansion factor for the sample segment. The statistic S_i is an estimate of the total acreage in the region owned by individuals who have the characteristic, attitude, or make the particular choice.

A behavioral relationship between landowner choice and factors affecting that choice must exist if the owner's response to a change in economic incentives is to be predicted. If the choice determinants could be varied and the survey repeated for T mutually-independent trials, the observed average total acreage associated with the i^{th} choice would be:

$$\bar{S}_i = \frac{1}{T^2} \sum_{k=1}^{K} \left(\sum_{t=1}^{T} N_{kt}(i,X) \cdot A_{kt}(X) \right) M_k ,$$

where $X = \{x_{kt}\}$ is the set of possible vectors of choice determinants and x_{kt} is the vector for owner k in trial t. As T becomes large, $\bar{S}_i$ approaches the expected value:

$$\hat{S}_i = \sum_{k=1}^{K} P_{ik} \cdot E(A_k) \cdot M_k$$

where P_{ik} is the probability that individual k will choose alternative i, and $E(A_k)$ is the expected acreage used by owner k.

The inferential statistic $\hat{S}_i$ suggests the development of a discrete-continuous choice model that simultaneously explains P_{ik} and $E(A_k)$ as a function of X. Such a model can be calibrated from a single survey if the sample observations are treated as a set of mutually-independent trials for a representative owner. The representative owner hypothesis requires that all owners in the population have the same set of choice determinants and the same functional relationship between choice determinants and observed choices.

A Profit Maximization Model

The obvious choices for a behavioral hypothesis are profit and utility maximization. Utility maximization is more general, but more complicated. Here the more simple profit maximization hypothesis is used.

The use of profit maximization as a behavioral objective requires the assumption that the land is used as a factor of production and not as a homesite or other consumption good. When this assumption can be applied, we may follow McFadden (1981) by specifying a production possibilities set and defining a dual profit function. The unique aspect of this specification is that it defines land use choice as a discrete input. Although the discrete input makes the initial technology set nonconvex, convexity remains a property of the dual technology set implied by the profit function. McFadden's specification of an initial technology set and profit function are adapted to the land use problem in the immediately following subsections. Dual and original technology sets are then compared to verify the consistency of the specification.

<u>A Technology Set</u>

Let:

$i = 1,\ldots,I$ be an integer index of a set of mutually exclusive, exhaustive and feasible revenue-generating land use alternatives,

$q = [q_1,\ldots,q_I]$ be a vector of choice indicators, with $q_i = 1$ if land use alternative i is chosen and $q_i = 0$ otherwise,

x be a vector of observed exogenous variables including measures of owner characteristics and site attributes,

$a = \sum_i a_i$, where $a_i \geq 0$ is the number of acres devoted to land use i,

$\epsilon = [\epsilon_1, \ldots, \epsilon_I]$, where ϵ_i is an exogenous random variable representing unobserved quantities of inputs or outputs associated with land use i,

$\nu = [E(\nu_1), \ldots, E(\nu_I)] + [\mu_1, \ldots, \mu_I]$, where ν_i is a J^{th} order column vector of exogenous random yields of j possible products gained from land use i, $E(\nu_i)$ is a corresponding vector of expected yields, and μ_i is a vector of zero-mean random yield components.

The feasible set for the discrete input q, which represents the input of a land use choice, can be represented as:

$$R(x, \epsilon) = \left\{ q \in \mathbb{R}^I \right\} .$$

Thus this input can depend upon owner characteristics, site attributes, and upon the availability of unobserved inputs or outputs such as milking parlors, harvest equipment, and wild game. Examples of restrictions on the set of land use alternatives might include a farmer nearing retirement who rules out tree crops, a farmer who eliminates dairying as a land use because there is no milking parlor, or an individual who includes participation in the Conservation Reserve Program because he or she is an avid hunter.

Examples of observable owner characteristics include measures of wealth, occupation, age and education. Examples of site characteristics include soil fertility, topography, presence or absence of diseases and insects, and location relative to output or input markets. From the representative owner perspective, these observed exogenous variables control for systematic differences between sampled owners.

The definition of a_i requires total acreage to be divided into fields or plots devoted to a single land use. Thus practices such as double cropping, crop rotations, etc. must be defined as separate land uses with multiple outputs. This definition of homogeneous use areas parallels the identification of sample tracts within the survey segments of the area frame.

With these definitions, the land input requirement set may be defined as follows. Let:

$$S(q,x,\nu) = \left\{ a \in \mathbb{R}_+ \mid \text{some feasible } q \text{ is given} \right\}$$

and let $\{\nu\}$ be the set of all possible realizations of the matrix ν of exogenous random yield effects. Then:

$$T(x,\epsilon,\nu) = \left\{ (q,a) \in \mathbb{R}^{I+1} \mid q \in R(x,\epsilon), \text{ and } a \in S(q,x,\nu) \right\}.$$

is the input requirements set for the land use problem.

This set characterizes the production possibilities of the landowners if the output from the i^{th} land use is defined as $y_i = \nu_i a_i$, or as yield per acre times the number of acres devoted to a particular land use. The matrix of stochastic outputs is consequently defined as

$$y = [y_1,\ldots,y_I] = [a_1\nu_1,\ldots,a_I\nu_I]$$

and, since ν is exogenous, the set of outputs is determined by the input requirements set.

Since q is discrete, this technology set will not be convex. Thus, we shall need to use duality principles in formulating the profit maximization model.

The Profit Function

If a land use decision is made each time revenue is generated from the land asset, profit may be defined as:

$$\Pi(\gamma,p,w,x,\nu,\epsilon) = \alpha(\gamma + \beta(p,w,x,\nu),x,\epsilon)$$

where

$p = [p_1,\ldots,p_J]$, and $p_j > 0$ is the expected price of product j,

$w = [w_1,\ldots,w_H]$, $w_h < 0$, and w is a vector of expected prices of the productive factors $[f_1,\ldots,f_H]$,

$\gamma = [\gamma_1,\ldots,\gamma_I]$, and $\gamma_i \leq 0$ is a function representing the cost of shifting from the current land use to land use i,

$\beta = [\beta^1,\ldots,\beta^I]$, and $\beta^i \geq 0$ is a function measuring the net returns from land use i,

$\alpha \geq 0$ is a function measuring the total net returns from the land asset during a given production period.

The functions β and α are assumed to be convex, increasing in p, w and γ, linearly homogeneous in these values, and bounded from above. Net returns received from the output of the land in use i are assumed to be independent of the costs of shifting to land use i.

Since the land use decision typically precedes the realized output from the land, we assume that the owner makes this decision on the basis of expected prices. Hence p_j and w_h represent exogenous expected prices while γ_i represents a current value. Current values may also be used for the expected price variables if one accepts the naive price expectation hypothesis. Past prices and yields may be included in x as predetermined variables if an extrapolative or adaptive expectation hypothesis is preferred. One might also use computed net present values for p, such as has been suggested by Brown and Brown, or would be obtained from the Faustmann soil rent formula (Newman).

<u>Dual Technology Set</u>

The properties of the profit function permit characterization of the dual technology set. Following McFadden (1981, p. 10), we can note that the function $\alpha(\gamma,x,\epsilon)$ is dual to

$$R(x,\epsilon) = \left\{ q \in \mathbb{R}^I \mid \gamma\cdot q \leq \alpha(\gamma,x,\epsilon) \ \forall \ \gamma \in \mathbb{R}^I \right\}.$$

and the function $\beta^i(p,w,x,\nu)$ is dual to

$$S^i(x,\nu) = \left\{ a \in \mathbb{R}_+ \mid p\cdot y - w\cdot f \leq \beta^i(p,w,x,\nu) \ \forall \ p \in \mathbb{R}^I_{++}, \text{ and } \nu \in \mathbb{R}^{I+J} \right\}.$$

Hence, $\gamma_i + \beta^i(p,w,x,\nu)$ is dual to

$$\left\{ (q,a) \in \mathbb{R}^{I+1} \mid a \in S^i(x,\nu),\ q_i = 1,\ q_j = 0 \text{ for } j \neq i \right\}.$$

Composition rules (McFadden 1978, pp. 95-100) then allow the overall technology set to be defined for the profit function $\alpha(\gamma + \beta(p,x,\nu),x,\epsilon)$ as:

$$\bigcup_{q \in R(x,\epsilon)} \sum_{i=1}^{I} q_i \left\{ (q^i,a_i) \in \mathbb{R}^{I+1} \mid a_i \in S^i(x,\nu),\ q^i_i = 1,\ q^i_j = 0, \text{ for } j \neq i \right\}$$

$$= \bigcup_{q \in R(x,\epsilon)} \left\{ (q, \sum_{i=1}^{I} q_i a_i) \in \mathbb{R}^{I+1} \mid a_i \in S^i(x,\nu) \text{ for } i = 1,\ldots,m \right\}.$$

This dual technology set is the convex hull of

$$T(x,\epsilon,\nu) = \left\{ (q,a) \in \mathbb{R}^{I+1} \mid q \in R(x,\epsilon) \text{ and } a \in S(q,x,\nu) \right\}$$

when $S(q,x,\nu) = \sum_i q_i \cdot S^i(x,\nu)$.

Hence the defined profit function has a plausible technology set that is consistent with the initial input and output definitions. Furthermore, this dual technology set is convex despite the discrete nature of q.

Derived Demands for Land Uses

Formulating the dual profit function allows derivation of a derived demand system for land use choice. This system can be obtained through application of the derivative property. It can be estimated from area-frame sample data using standard discrete variable econometric techniques. The estimated system can be used to obtain simultaneous predictions of landowner land use choice probabilities.

Estimation of the system is simplified by restricting q to a unit vector. Little or no generality is lost by doing this, since mutually-exclusive alternatives can always be designated by a set of unit vectors. We shall also assume that ϵ, the measures of unobserved inputs and outputs associated with each land use, can be translated into a vector of monetary measures: $z = \psi(x,\epsilon)$. In this case, the technology set R may be specified as the particular set

$$R(x,\epsilon) = \left\{ (q,z) \in \mathbb{R}^{2I} \mid q \text{ is a unit vector, } z_i = q_i\psi(x,\epsilon_i) \right\}.$$

Since z is expressed in monetary terms, the price of each z_i may be taken to be one. Then the dual profit function corresponding to the particular set R is

$$\alpha(\gamma,x,\epsilon) = \max_i \left[\gamma_i + \psi(x,\epsilon_i) \right] \qquad i = 1,\ldots,I.$$

To obtain a useful profit function that includes returns from use of the land, we may use the random yield definition to further specify β^i:

$$\beta^i(p,w,x,\nu) = \bar{\beta}^i(p,w,x;E(\nu_i)) + \delta(p,\mu_i)$$

The conditional nature of the expected net returns function $\bar{\beta}^i$ reflects the assumption that decisions are made on the basis of expected yields. Actual returns may differ from expected values because of random yield variation. This random variation is represented by the function δ.

Given this definition of β^i, the overall profit function for the representative landowner may be written as

$$\Pi(p,w,\gamma,x,\nu,\epsilon) = \max_{1 \leq i \leq m} \alpha\left[\gamma_i + \psi(x,\epsilon_i) + \bar{\beta}^i(p,w,x;E(\nu_i)) + \delta(p,\mu_i),x \right].$$

Application of Hotelling's Lemma to this particular function gives the derived demand for land use i as

$$Q^i(\gamma,p,w,x,\epsilon,\nu) = \begin{cases} 1 \text{ if } \psi(x,\epsilon_m) + \delta(p,\mu_m) - \psi(x,\epsilon_i) - \delta(p,\mu_i) \leq \gamma_i - \gamma_m \\ \quad + \bar{\beta}^i(p,w,x,E(\nu_i)) - \bar{\beta}^m(p,w,x;E(\nu_i)) \text{ for } m = 1,\ldots,I \\ 0 \text{ otherwise.} \end{cases}$$

This equation is similar to the random utility formulation underlying most probabilistic choice models (Maddala pp. 59-78). When calibrated, it can be used to compute the probability that the landowner will choose a particular land use.

Further application of Hotelling's Lemma will give supply functions for the j^{th} product:

$$Y^j(q,p,w,x,\nu) = \begin{cases} \bar{\beta}^i_{p_j}(p,w,x;E(\nu_i)) + \delta_{p_j}(p,\mu_i) & \text{if } Q^i(\gamma,p,w,x,\epsilon) = 1 \\ 0 \text{ otherwise} & \end{cases}$$

$$j \in (1,\ldots,J).$$

Derived demand functions for the observed non-land inputs can be obtained in a similar manner. However, estimates of the expected acreage devoted to each land use is not a direct result of the presently-constituted model. To obtain such estimates, one would have to know $E(\nu_i)$, estimate $E(Y^j)$, and then divide the expected output of product j by its expected yield. In the next subsection, we suggest an alternative way to obtain estimates of these expected acreages.

An Expected Acreage Equation

One way to develop direct estimates of the expected acreages used by a profit-maximizing landowner is to assume that the net returns functions for the various land uses are homothetic. This is a restrictive assumption (Chambers pp. 149-152). It implies that returns to scale are constant and that profit-maximizing plans are identical for each acre devoted to a particular use. Despite these restrictions, the homotheticity assumption is commonly employed. It is used, for example, by farm management specialists when they develop crop budgets or linear programming models of alternate farm enterprises. It is also implicitly employed whenever a per-acre figure is used to represent optimum returns or costs for a field or plot. In our case, the assumption implies that landowners use per-acre figures to determine optimum production plans for a given land use, and that they treat acreage devoted to any particular use as homogeneous in quality and yield.

Given homotheticity, total expected net returns from land use i may be specified as;

$$\beta^{i}(p,w,x,\nu) = a_{i}\cdot\tilde{\beta}^{i}(p,w,x;E(\nu_{i})) + a_{i}p\cdot\mu_{i},$$

where $\tilde{\beta}^{i}$ is the expected per-acre net returns from land use i. Similarly, costs of shifting to another land use can be written as:

$$\gamma_{i} = a_{i}c_{i}$$

where c_i is the per-acre cost of adopting land use i. Supply of product j by a profit-maximizing owner is, in this instance,

$$Y^{j}(q,p,w,x,\nu) = \begin{cases} a_{i}[\tilde{\beta}^{i}_{p_{j}}(p,w,x;E(\nu_{i})) + \mu_{i}] & \text{if } Q^{i}(\gamma,p,w,x,\epsilon) = 1 \\ 0 \text{ otherwise} & \end{cases} \qquad j \in (1,\ldots,J)$$

and $\tilde{\beta}^{i}_{p_{j}}$ is, by definition, the revealed expected yield of this product from land use i. Comparison of revealed expected yields to observed yields produced by landowners who choose use i therefore becomes one possible way to test the validity of this form of the profit maximization model.

When homotheticity is assumed, the implicit quasi-rent accruing to the land when it is in use i and when profits are maximized can also be obtained from application of Hotelling's Lemma. This rent is

$$R^i = \begin{cases} -\,[\tilde{\beta}^i(p,w,x;E(\nu_i)) + c_i] & \text{if } Q^i(\gamma,p,w,x,\nu,\epsilon) = 1 \\ 0 \text{ otherwise} & \end{cases}$$

for land uses $i = 1,\ldots,I$. Rent is defined in this case as $-\delta\Pi/\delta a_i$, noting that profit may be expressed as $p\cdot y - w\cdot f - \sum_i r_i a_i$ where r_i is the cost of the land input that is devoted to use i (Chambers pp. 126-127).

The expected return to the land input when profits are maximized may then be determined as:

$$E(R\cdot A) = E(R)E(A) = \sum_{i=1}^{I} a_i E(R^i \mid Q^i = 1)(\text{Prob } Q^i = 1)$$

where R is the quasi-rent of the land across all uses and A is the acreage employed in all uses by the representative landowner. The expected acreage used by this landowner is consequently:

$$E(A) = \sum_{i=1}^{I}\left(\frac{-1}{E(R)} \left(\beta^i(p,w,x;E(\nu_i)) + \gamma_i\right)\left(\text{Prob } Q^i = 1\right)\right).$$

This equation will be used to predict the expected acreage employed in different land uses by a profit-maximizing landowner.

An Econometric Procedure

As noted earlier, a suitably-specified derived demand function for the discrete input, $Q^i(\gamma, p, w, x, \nu, \epsilon)$, may be estimated by discrete choice methods. This estimated function will provide predictions of P_{ik}, the probability that landowner k chooses land use i.

One of the simplist examples of this type of estimation is the case where the land use choice is dichotomous. Suppose the error terms μ_1, μ_2, and $\psi(x, \epsilon_1) + p_1\mu_1 - \psi(x, \epsilon_2) - p_2\mu_2$ are assumed to be jointly normally distributed with variance-covariance matrix:

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{01} \\ & \sigma_{22} & \sigma_{02} \\ & & 1 \end{bmatrix}.$$

The probability that landowner k chooses the second of the two land uses then would be

$$\Phi_k = \Phi(d_0 + d_1[\beta^2(p_k, w_k, x_k) - \beta^1(p_k, w_k, x_k) + \gamma_{2k} - \gamma_{1k}] + d \cdot x_k)$$

where Φ is the standard normal cumulative distribution function, d_0, d_1 and $d = [d_2, \ldots, d_{L+2}]$ are scale and location parameters, and p_k, w_k and x_k are vectors of observed values for the k^{th} owner. This probability may be estimated using a probit model if we have given values of β^1, β^2, γ_1 and γ_2. Since the land use choice is dichotomous, the probability that the landowner would choose the first use can be estimated as $1 - \hat{\Phi}_k$.

When there are more than two land use choices, computational burden may encourage substitution of a conditional logit model for the probit specification. Such a substitution will not necessarily dictate a change in model error specification, for multinomial logit analysis methods may be used to estimate a model with jointly-normal errors (Maddala pp. 272-276). The introduction of logit methods may introduce an "independence of irrelevant alternatives" assumption, however, and this assumption may be violated if crop rotations, equipment capabilities, complementary feed-livestock enterprises, or other factors create correlations between land use alternatives. When this assumption is violated, the more complex Generalized Extreme Value method would have to be used to estimate the probabilistic choice component (Maddala pp. 70-72).

Obtaining exogenously-determined values for γ_i should be feasible, since these costs are contemporaneous to the land use decision. Exogenous values for β^i are less feasible, since there is no way to assure that such estimates will maximize profits or match the values used by the landowner. We propose instead to approximate the nonlinear net returns functions with second-order flexible functions and to simultaneously estimate the discrete input derived-demand and expected acreage functions. The suggested procedure is an adaptation of the E-M method for obtaining maximum likelihood estimates for a Tobit model (Maddala p. 222).

An illustration will clarify this proposed approach. Let us omit non-land costs for simplicity and approximate the expected returns from land use i with a generalized Leontief function. Then:

$$\beta^i(p,x;E(\nu_i)) \cong E(R)\cdot B^i(p,x) =$$

$$E(R)\cdot\left[b^i_{00} + \sum_{m=1}^{J}\left(\sum_{n=1}^{J} b^{0i}_{mn}\sqrt{p_m p_n} + \sum_{n=1}^{L} b^{1i}_{mn}\sqrt{p_m x_n}\right) + \sum_{m=1}^{L}\sum_{n=1}^{L} b^{2i}_{mn}\sqrt{x_m x_n}\right]$$

where $i \in (1,\ldots,I)$, J is the number of outputs, and L is the number of elements in x. We shall employ regression methods to estimate:

$$A_k = \sum_{i=1}^{m}\left[P_{ik}\left(- B^i(p,x_k) - b^i_{cc}c_{ik} - \omega_{ik}\right)\right]$$

(where $\omega_{ik} = E(R)^{-1}a_{ik}p_k\cdot\mu_{ik}$ is a random error term) using estimated probabilities from an ad hoc probabilistic choice model as instruments for the P_{ik}. The probabilistic choice model would then be re-estimated, using estimates of $B^i(p,x)$ from the acreage model as instruments. Iteration of this process would be continued until the estimates of P_{ik} and parameters of $B^i(p,x)$ stabilize. The resulting equations would respect the cross-equation parameter restrictions of the profit maximization model.

Parameters estimated for the discrete input demand equations using this procedure would differ from those of the theoretical model by the factor E(R). Relative relationships would be maintained, however, and substitution of $B^i(p,x)$ for $\beta^i(p,x)$ should have negligible effect on the probability predictions, for the logit and probit analysis models are specified only up to a linear transformation. Profit function properties should also be maintained for $B^i(p,x)$. Hence b^{0i}_{mn} should be negative whenever $m \neq n$ and positive whenever $m = n$.

Sample selection bias is a potential estimation problem of the procedure, since the error terms ω_{ik} and μ_{ik} are interrelated. A way to correct for this bias is shown by Maddala (pp. 224-227). The error term ω_{ik} is also expected to be heteroscedastic, since it would increase in magnitude with the total value of the random output from land use i. Standard errors of the estimated parameters of the regression equation accordingly would be adjusted using White's procedure (1978). However heteroscedastic estimation procedures would not be employed because of the iterative estimation method.

This procedure would provide estimated equations that could be employed to obtain predictions of expected acreages and land use choice probabilities for each of the sampled landowners in an area-frame survey. Given these estimates for $E(A_k)$ and P_{ik}, prediction of the total land area devoted to use i within a defined region would be a straightforward matter of applying the definitional equation for $\hat{S}_i$.

An Illustrative Application

The Problem

Leaving a harvested southern pine site to regenerate without planting, seeding or leaving seed trees often will result in the site converting to hardwood species. Federal and state governments have sought to stop this unwanted conversion by encouraging nonindustrial private forest (NIPF) landowners to invest in pine regeneration practices such as planting pine seedlings. Methods have included government sharing of the costs of planting or seeding a harvested site. The presence of cost sharing has, in turn, created an interest in measuring the effects of cost-share programs on regeneration choices and acreages regenerated.

Research on this issue concentrated initially on the determination of economically feasible regeneration investments (USDA Forest Service 1981, 1988). Focus then shifted to the prediction and explanation of regeneration choices made by typical owners (Boyd, de Steiguer, Hyberg, Royer). As part of the study of landowner choice, the USDA Forest Service commissioned a survey of NIPF landowner management and reforestation practices for harvested Southern pine land (Fecso et al.). The survey utilized an area sample frame maintained by the USDA Agricultural Statistics Service, and produced data suitable for the estimation of inferential statistics of the form $\hat{S}_i$. Here we use this data to predict acreages of Southern pine land receiving post-harvest regeneration treatments under different cost-share programs.

The land use choice for this analysis is defined as that of either leaving the site to regenerate on its own ($q_{2k} = 0$) or of actively investing in its reforestation through planting, seeding, or leaving seed trees ($q_{2k} = 1$). Determinants of this choice, and of the acres devoted to each land use, include pulpwood and sawtimber stumpage prices (p_1, p_2), the per acre net cost (after any cost sharing) of planting or seeding a plot (c_2), the landowner's annual household income (x_1), and five dichotomous categorical variables which take the value of one if the landowner is in the category and zero otherwise. The categories include: (1) the landowner is a farmer (x_2), (2) the site is located in a survey sample segment classified as forest land (x_3), (3) the landowner received public forester assistance (x_4), (4) the landowner retained the services of a consulting forester (x_5), and (5) the landowner dealt with an industrial forester (x_6). Measures of yields are not available, and costs other than the costs of planting or seeding are assumed to be negligible. Estimates of present net returns from future harvests are available, but show no correlation with observed regeneration choices or acreages harvested.

Since measures of harvest yields are not present in the data, probabilities of regeneration choice and expected total pine acreage demanded by the landowner are estimated using the previously discussed iterative procedure. The derived demand function for the discrete land use input is calibrated using a dichotomous probit maximum likelihood model. Expected acreages are obtained from the model:

$$a_k = -B^1(p_k,x_k)(1-\hat{\Phi}_k) - [B^2(p_k,x_k) + b^2_{cc}c_{2k}]\hat{\Phi}_k - \phi_k(\sigma_{10}-\sigma_{20}) + \eta_k$$

$$= -B^1(p_k,x_k) - [B^2(p_k,x_k) - B^1(p_k,x_k) + b^2_{cc}c_{2k}]\hat{\Phi}_k + \phi_k(\sigma_{20}-\sigma_{10}) + \eta_k.$$

The variable $\hat{\phi}_k$ is the value from the standard normal density function corresponding to $\hat{\Phi}_k$, the estimated probability that landowner k chooses to plant, seed or leave seed trees. This variable is introduced to correct for sample selection bias. The error term

$$\eta_k = \hat{\phi}_k(\sigma_{10}-\sigma_{20}) - \omega_{1k} - \omega_{2k}$$

is zero-mean but heteroscedastic, requiring adjustment of the standard errors of the estimated parameters.

Pine Regeneration Results

Descriptive statistics for the data used in this analysis are given in Table 1. Estimation results are presented in Tables 2 and 3. All parameters that could not be shown to be different from zero at a 30 percent level of significance are assumed to be zero.

Estimates in Table 3 indicate that when landowners invest in pine land regeneration, they seek to maximize profits. Estimated price parameters for this choice are consistent with a dual profit function that increases in output prices and is convex in these prices. Estimated demand for acreage to be regenerated by planting, seeding or leaving seed trees also increases with output price and with income. For a given acreage and income, estimated profits are higher for landowners who are farmers or whose sites are located in a forested area.

When landowners do not invest in pine regeneration, the acreage left to regenerate naturally cannot be explained by the profit maximization hypothesis. Since a conservative criterion is used to infer that estimated parameters are equal to zero, we can assert with reasonable confidence that prices and income do not explain the harvested acreage that is not planted, seeded or left with seed trees. Of the variables available in the data set, only two categorical measures of landowner information (x_4 and x_6) can be used to predict this acreage.

Despite this inability to explain untreated acreage by using the profit maximization hypothesis, 54 of the 68 NIPF owners observed to plant, seed or leave seed trees are predicted to actively invest in pine regeneration, and 148 of the remaining 159 owners are predicted not to invest. Eleven owners are predicted to actively invest who did not, and 14 who invested are predicted to omit planting, seeding or leaving seed trees. Overall, 89 percent of the observed sample choices are correctly predicted.

The expected presence of heteroscedasticity in the error term of the acreage model is verified by the Breusch-Pagan test, which rejects the homoscedasticity hypothesis at a high level of significance (cf. Table 3). As is also expected, sample selection bias is present in the sample. This bias is corrected by the significant parameter for $\hat{\phi}$ (Maddala p. 227).

Since the error term is heteroscedastic, we have no measure of the predictive power of the acreage equation. We expect, however, that this is moderate, since the biased $\bar{R}^2$ for the model with unadjusted standard errors is 0.35. That the equation provides significant predictions for the acreage that is actively regenerated is evident from the large Student t statistics for the parameters comprising $B^2 - B^1$. We suspect predictive power is reduced primarily by the inability of profit maximization behavior to explain the acreage harvested by landowners who did not plant, seed or leave seed trees.

Mean probability and expected acreage estimates for the survey period ("Base" predictions in Table 4) are virtually identical to the values observed for the subsample of 248 owners used to make the predictions. Predicted acreage for the region is 1.098 million acres, slightly above the 1.077 million acre sample value, but below the 1.260 million acre value cited as correct in Fecso et al. In Table 4, the total acreage predictions are scaled up to match the cited value.

Values are also predicted for four policy options, which are identified at the bottom of Table 4. As can be seen from that table, varying the size of the cost share and the number of qualifying NIPF landowners can have a large effect on the probability that a landowner will choose to actively invest in pine regeneration. It can also substantially affect both acreages treated by a representative owner and total acreage planted, seeded or left with seed trees in the South. Extending the existing subsidy to all NIPF owners who harvested pine during the survey period, for example, would have increased the total acreage receiving regeneration investments by 47 percent, and increasing the subsidy rate to 80 percent would have increased this acreage by an additional 58 percent. Cost sharing consequently appears to be an effective way to affect pine regeneration in the South.

Elasticities reported in the table show the predicted percentage response of $\hat{\Phi}$, the probability that the owner will plant, seed or leave seed trees, E(A), the expected acreage that the representative owner will harvest and regenerate, and $\hat{S}_i$, the predicted acreage that will be planted, seeded or left with seed trees in the South, to a given percentage increase in c_2, the owner's costs of reforestation. These elasticities are not calculated at the mean sample values; they are averages of 248 individual elasticities calculated for each survey respondent.

Tabled results show an elastic response in both the probability of treatment and the acreage regenerated by a typical owner when cost sharing is at low levels. This response becomes inelastic when subsidy levels are high. The size of regenerated plot first increases as owner's costs increase and then decreases as these costs continue to increase. Total acreage planted, seeded or left with seed trees always increases as owner's costs decrease, however, and the response to a further cost decrease is always elastic.

The results given in Table 4 are typical of the type of predictions that can be obtained from a model based on an area-frame sample. Hypothetical policy options can be explored, but results will pertain to the period in which the sample is obtained. Results are enriched by the decomposition of the land use decision into land use choice and acreage components. Consistency is maintained between the representative landownerand the aggregate land use for a given region by the sample-based aggregation procedure.

Conclusions

Employment of an area-frame sampling procedure to solve the land use aggregation problem allows development of a relatively straightforward profit maximization model of landowner behavior. When estimated from disaggregated area-frame sample data, the model allows prediction of both representative landowner and regional land area response. Since the model is constructed from the area-frame statistic, data and method comprise an integrated approach to land use policy analysis. The random yield formulation of the model allows estimation to proceed without output or yield data.

The pine regeneration model used to illustrate the modelling approach indicates that some nonindustrial private forest landowners do not behave as profit-maximizing agents. Yet the model provides useful predictions of regenerated acreage and landowner decisions for the owners who do exhibit profit maximization behavior. Derivation of such results when there is a behavioral mis-specification for part of the landowners is possible because the model exhibits a property similar to the "weak complementarity" property of discrete-continuous choice demand models (Hanneman p. 542): attributes of the choice which is not made do not affect the acreage component of the chosen alternative.

The "weak complementarity" property also suggests use of Hanneman's conditional utility function as a way to extend the model of the paper. The conditional utility formulation would allow some land use choices to be motivated by profit maximization behavior and others to be driven by consumption. Acreage would relate to land use choice in a more complex manner when land is a consumption good, for Roy's Identity would have to replace Hotelling's Lemma. A good example of such an application to the demand for fuel is provided by Dubin and McFadden.

Further development of the profit maximization behavioral model is also feasible. An obvious extension is the introduction of risk-averse behavior on the part of the landowner facing random yields. Investment of such conceptual effort should perhaps be deferred, however, until further empirical verification of the modeling approach is obtained.

Given the profit maximization hypothesis of the paper, one might wonder why the empirical illustration did not deal with farmers' choice between farm land uses. Such an application would be more consistent with the presumption that land is used for income-generating purposes, and could be quite useful in predicting acreage response to proposed changes in price policies. This option was not chosen because the area-frame data currently collected on farm land use is restricted to observations on acreages in different uses, numbers of livestock, numbers of paid farm workers, and total sales value. Thus the data on characteristics needed to implement the described modeling approach are not available from the current surveys.

At present, area frame sample data are used mostly to develop descriptive statistics of regional land use. These statistics are being employed in economic models without regard to the land use aggregation problem. The described modeling approach offers a way to obtain aggregate inferential statistics derived from explicit behavioral hypotheses. We believe this approach could be quite useful in analyzing land use and landowner response to changes in economic incentives.

References

Alig, R. J. and R. G. Healy. "Urban and Built-up Land Area Changes in the United States." *Land Economics* 63(1987):215-226.

Boyd, R. "Government Support of Nonindustrial Production: The Case of Private Forests." *Southern Economic J.* 51(1984):89-107.

Brown, K.C. and D.J. Brown. "Heterogenous Expectations and Farmland Prices." *American J. Agricultural Economics* 66(1984):164-169.

Chambers, Robert G. *Applied Production Analysis*, Cambridge University Press, Cambridge, 1988.

de Steiguer, J.E. "Impact of Cost-share Programs on Private Reforestation Investment." *Forest Science* 30(1985):697-704.

Dubin, J.A. and D.L. McFadden. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(1984):345-362.

Esseks, J. D. and S. E. Kraft. "Why Eligible Land Owners Did Not Participate in the First Four Sign-ups of the Conservation Reserve Program." *J. Soil and Water Conservation* 43(1988):251-255.

Fesco, R. S., H. F. Kaiser, J. P. Royer and M. Weidenhamer. "Management Practices and Reforestation Decisions for Harvested Southern Pinelands." USDA Statistical Reporting Service, Staff Report AGES82130, 1982.

Houseman, E. E. "Area Frame Sampling in Agriculture." USDA Statistical Reporting Service, Report 20, 1975.

Hyberg, B. *The Development and Testing Of a Utility Maximizing Model for NIPF Landowner Forest Management Behavior*. Ph.D. Dissertation, North Carolina State University, Raleigh, 1986.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics.* Econometric Society Monographs No. 3. Cambridge University Press, Cambridge, 1983.

McFadden, D. "The General Linear Profit Function." In M. Fuss and D. McFadden (eds.), *Production Economics, Volume 1*, North-Holland Publishing, Amsterdam, 1978.

McFadden, D. "Econometric Net Supply Systems for Firms with Continuous and Discrete Commodities." Unpublished working paper, Department of Economics, Massachusetts Institute of Technology, 1981.

McIntosh, C. S. and K. H. Shideed. "The Effect of Government Programs on Acreage Response Over Time: The Case of Corn Production in Iowa." *Western J. of Agricultural Economics* 14(1989):38-44.

Newman, D.H. "The Optimal Forest Rotation: A Discussion and Annotated Bibliography." USDA Forest Service General Technical Report SE-48, 1988.

Royer, J. P. "Determinants of Reforestation Behavior Among Southern Landowners." *Forest Science* 33(1987):657-667.

Sato, Kazuo. *Production Functions and Aggregation*, North-Holland Publishing Co., Amsterdam, 1975.

USDA Forest Service. *Economic Opportunities to Increase Softwood Production on Private Lands*, Vols 1 (Projections) and II (Investments). State and Private Forestry, Washington D.C., 1981.

USDA Forest Service. *The South's Fourth Forest: Alternatives for the Future*, Forest Resources Report No. 24. Washington D.C., 1988.

White, H. "A Heteroscedasticity Consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 46(1978):817-838.

TABLE 1: Description of Data Used in Southern Pineland Regeneration Analysis

Variable Symbol	Variable Description	Mean Value	Standard Deviation
p_1	Pulpwood price: 6 year average for 35 *Timber Mart South* regions, $ per mbf.	11.76	4.870
p_2	Sawtimber price: 6 year average for 35 *Timber Mart South* regions, $ per mbf.	125.7	27.08
c_2	Cost of reforestation, net of any subsidy, $ per acre, 1983.	77.36	41.98
x_1	Landowner 1980 annual household income, 1000's of $.	25.45	17.65
x_2	1 if landowner is a farmer, 0 otherwise.	0.207	0.406
x_3	1 if site is classified as *forest* in sample, 0 otherwise.	0.502	0.501
x_4	1 if landowner was assisted by a service forester, 0 otherwise.	0.216	0.412
x_5	1 if landowner retained a consulting forester, 0 otherwise.	0.093	0.290
x_6	1 if landowner dealt with an industrial forester, 0 otherwise.	0.062	0.241
q	1 if landowner planted, seeded or left seed trees during harvest, 0 otherwise.	0.300	0.459
a	Acreage harvested at site from Jan. 1971 to May 1981.	64.15	80.20

TABLE 2: Estimated Parameters for Model of NIPF Landowner Pineland Regeneration Choice

Variable	Parameter Estimate (Std. Error of Est.)	t Statistic (Signif. Level)	Mean of Variable (Std. Dev. of Var.)
Constant	.74157 (.29113)	2.55 (0.01)	1.000 0.000
$B^2 - B^1$	-.00234 (.00089)	-2.63 (0.01)	119.9 (236.1)
c_2	.02173 (.00380)	5.72 (0.00)	-77.36 (41.98)
x_4	1.0130 (.28093)	3.61 (0.00)	.2158 (.4123)
x_5	.84015 (.51884)	1.62 (0.11)	.0925 (.2904)

Estimated using probit MLE model with q as choice variable. Variable $B^2 - B^1$ is computed from estimated acreage equation to insure cross-equation parameter restrictions. Sample of 227. Model is significant at 0.000 level.

Table 3: Estimated Parameters for Model of NIPF Landowner Demand for Regenerated Pineland Acreage

Variable	Parameter Estimate (Std. Error of Est.)	t Statistic (Signif. Level)	Mean of Variable (Std. Dev. of Var.)
Intercept	-51.373 (6.9501)	-7.39 (0.00)	-1.000 (0.000)
$-x_4$	-145.87 (36.318)	-4.02 (0.00)	-.2143 (.4112)
$-x_6$	56.253 (25.849)	2.18 (0.03)	-.0625 (.2426)
$-\hat{\Phi}$	416.52 (87.916)	4.74 (0.00)	-.3028 (.3369)
$-\hat{\Phi} \cdot p_1$	118.66 (40.666)	2.92 (0.00)	-3.716 (4.779)
$-\Phi \cdot p_2$	9.8435 (3.5278)	2.79 (0.01)	-36.30 (40.94)
$-\hat{\Phi} \cdot (p_1 p_2)^{1/2}$	-70.373 (24.622)	-2.86 (0.00)	-11.47 (13.60)
$-\hat{\Phi} \cdot (p_1 x_1)^{1/2}$	-20.633 (3.4637)	-5.96 (0.00)	-5.685 (7.755)
$-\hat{\Phi} \cdot (x_1 x_2)^{1/2}$	20.340 (4.0038)	5.08 (0.00)	-.7314 (1.582)
$-\hat{\Phi} \cdot (x_1 x_3)^{1/2}$	19.488 (4.5587)	4.28 (0.00)	-.4407 (1.359)
$-\hat{\Phi} \cdot x_5$	-179.69 (61.885)	-2.90 (0.00)	-.0676 (.2338)
$-\hat{\Phi} \cdot x_6$	-172.43 (49.628)	-3.47 (0.00)	-.0407 (.1711)
$\hat{\phi}$	362.66 (107.38)	3.38 (0.00)	.1737 (.1378)

$\hat{\Phi}$ and $\hat{\phi}$ are probability and density estimates from the probit MLE model. Sample of 224. Chi-squared statistic for Breusch-Pagan heteroskedasticity test = 71.66 (Signif. level = 0.005). Standard errors are corrected by White's Procedure (1978).

Table 4: Predictions from Southern Pineland Regeneration Model

Estimate	Policy Alternative Base	1	2	3	4
Probability owner will reforest	.303	.105	.497	.579	.175
Average acreage treated by owners who reforest	64.6	60.7	68.2	81.3	54.2
Total acreage reforested in south (Thousand acres)	1,260	228	1,858	2,587	395
Elasticity of probability owner actively reforests	-3.44	-5.28	-0.66	-0.35	-4.22
Elasticity of acreage response - typical owner	0.46	1.18	-0.69	-0.87	1.64
Elasticity of acreage response - total region	-2.98	-4.10	-1.35	-1.22	-2.58

Base is 69% cost sharing for some owners. New policies are: (1) remove cost share program, (2) extend current subsidy to all NIPF owners who harvest, (3) increase rate of cost sharing to 80% and extend to all landowners, (4) extend program to all landowners but reduce rate to 30%.