301

Working Paper No. 301

PESTICIDE USAGE AND THE CHOICE OF PEST CONTROL STRATEGY: A SWITCHING REGRESSION ANALYSIS

by

W. Michael Hanemann, Yacov Tsur, and David Zilberman

California Agricultural Experiment Station
Giannini Foundation of Agricultural Economics
June 1984

PESTICIDE USAGE AND THE CHOICE OF PEST CONTROL STRATEGY: A SWITCHING REGRESSION ANALYSIS

I. Introduction

The adoption of integrated pest management (IPM) as a substitute for conventional chemical pest management (CPM) is an important issue in current agricultural and environmental policy. The factors which influence growers' decisions to adopt IPM and the effects of IPM adoption on agricultural productivity and the intensity of pesticide usage are fundamental questions which must be resolved before environmental control strategies can be evaluated. The purpose of this paper is to develop a methodology for analyzing these issues which is suitable for empirical application, using data that can readily be obtained from grower surveys. After describing the theoretical model and the procedures for estimating it, we will present an example of an application using data on a sample of cotton growers in the San Joaquin Valley in California.

The key to the model is the recognition that the "discrete" choice of whether to employ IPM or CPM and the "continuous" choice of how much pesticides and other agricultural inputs to apply are interrelated. The outcome of the one choice affects the outcome of the other, and both flow from a single underlying profit maximization decision on the part of the grower. This linkage between discrete and continuous choices shapes both the formulation of the theoretical model and the statistical procedure which is used to estimate the model. It also provides a way of unifying two separate strands of literature on IMP--the literature on new technology adoption in agriculture [for a survey, see Feder, Just, and Zilberman (   )], which tends to focus exclusively on the discrete choices, and the literature on pesticide demand functions [see, for example, Miranowski (   )] which tends to focus exclusively on the continuous choices. To be sure, the empirical mathematical programming models of grower behavior [for example, Musser and Stamoulis (1981) and Kaiser and Robinson (1979)] explicitly recognize the discrete/continuous nature of the choices faced by growers. But these are normative rather than positive analyses, and there is a serious problem in validating them. What is needed is a positive, statistical model of discrete/continuous choices which is capable of being validated by data on actual grower behavior. This is provided by the model developed in this paper.

The literature on statistical techniques for estimating switching regression models emerged somewhat ahead of the literature on microeconomic models of discrete/continuous choices. The basic statistical techniques were worked out by Amemiya (   ), Heckman (1979 and 1980), and Lee and Trost (1978); another type of estimation methodology which we will also consider was developed by Tsur (1983a, b). Early examples of theoretical microeconomic models of simultaneous discrete/continuous choices include Just and Zilberman (1979) and Just, Zilberman, and Rausser (1980) on the production side and Novshek and Sonnenschein (1979) and Lancaster (1976 and 1979) on the consumer demand side. However, these are deterministic, rather that stochastic, models of micro behavior. Stochastic models combining utility- or production-theoretic models of discrete/continuous choices with switching regression estimation techniques were developed by Hanemann (1978 and 1984) and Dubin and McFadden (1984) for consumer choices and by Duncan (1980) and Hanemann and Tsur (1982 and 1984) for producer choices. The present model is firmly in this latter tradition.

The paper is organized as follows: Section II outlines the theoretical model of discrete/continuous choices by growers. Section III describes some alternative estimation techniques. The model is applied to data on cotton growers in the San Joaquin Valley in section IV. Section V contains the conclusions, including a discussion of the empirical results and suggestions for how this type of model can be applied to other data sets in order to evaluate the consequences of alternative environmental control policies.

II. Model Specification

In this section, we develop a model of a grower's decisions on whether to use IPM or CPM and how much pesticides to apply which is suitable for empirical estimation. The model is tailored to the specific limitations of data that are available to us--constraints which might arise in other data sets on pesticide usage. However, in order to place the model in a broader context and indicate how one could proceed if a more complete data set were available, we present, first, a somewhat more general decision model and then specialize it to the data at hand.

The grower's decision variables are pest control strategy, represented by a binary variable T, where $T = 1$ indicates use of IPM and $T = 2$ indicates use of CPM; the quantity of pesticides applied per acre, x; and the quantity of all other nonpesticides per acre, represented by the vector z. There is a production function which depends, in principle, on the type of pest control strategy: $y = f(x, z; T)$, where y is the output of cotton per acre. Let the price of cotton be p, the cost of pesticides w, and the cost of nonpesticide inputs q. For the sake of generality, we allow for the possibility of other fixed costs associated with the use of the two technologies represented by k(T). Given that the growers acreage is A, which is taken as exogenous to the model, his profit is

$$\pi = \pi(x, z, T; p, w, q, A) \equiv [pf(x, z; T) - wx - qz]\, A - k(T). \tag{1}$$

The grower selects (x, y, T) so as to maximize profits (1). This generates a set of per-acre input demand functions,

$$x = h^{x}(p, w, q, A) \tag{2}$$

$$z = h^{z}(p, w, q, A); \tag{3}$$

a pest control strategy decision function,

$$T = h^{T}(p, w, q, A); \tag{4}$$

a per-acre output supply function,

$$y = h^{y}(p, w, q, A) \equiv f[h^{x}(p, w, q, A), h^{z}(p, w, q, A); h^{T}(p, w, q, A)]; \tag{5}$$

and a (maximized) total profit function,

$$\begin{aligned} \pi &= \pi(p, w, q, A) \\ &\equiv \pi[h^{x}(p, w, q, A), h^{z}(p, w, q, A), h^{T}(p, w, q, A); p, w, q, A]. \end{aligned} \tag{6}$$

By Hotelling's lemma,

$$h^{x}(p, w, q, A) \cdot A = \frac{\partial\pi(p, w, q, A)}{\partial w} \tag{7a}$$

$$h^{z}(p, w, q, A) \cdot A = \frac{\partial\pi(p, w, q, A)}{\partial q} \tag{7b}$$

and

$$h^{y}(p, w, q, A) \cdot A = \frac{\partial\pi(p, w, q, A)}{\partial p}. \tag{7c}$$

Observe that this optimization simultaneously embodies a continuous and a discrete choice; the continuous choice is the level of pesticide and other inputs while the discrete choice is the selection of a pest control strategy. Moreover, the continuous and discrete choices are interdependent and both flow from a single, underlying profit maximization. For example, the amount of pesticides to be applied depends on which pest control strategy is adopted and vice versa.

In order to illuminate this interdependence, it is convenient to think of the maximization of (1) as occurring in two stages. Suppose that the grower has decided to adopt IPM ($T = 1$). Conditional on this decision, his per-acre production function may be written as: $y_1 = f_1(x, z) \equiv f(x, z; 1)$, where the subscript "1" indicates output conditional on $T = 1$, and his profit is

$$\pi_1 = \pi_1(x, z; p, w, q, A) \equiv [pf_1(x, z) - wx - qz] A - k(1). \tag{8}$$

The levels of pesticide and other input usage conditional on the decision to adopt IPM are determined by maximizing (8). This yields the conditional input demand functions,

$$x_1 = h_1^x(p, w, q) \tag{9}$$

$$z_1 = h_1^x(p, w, q); \tag{10}$$

the conditional output supply function,

$$y_1 = h_1^x(p, w, q) \equiv f_1[h_1^x(p, w, q), h_1^z(p, w, q)]; \tag{11}$$

and the conditional (maximized) profit function,

$$\pi_1 = \pi_1(p, w, q, A) \equiv \pi_1[h_1^x(p, w, q), h_1^z(p, w, q); p, w, q, A]. \quad (12)$$

Note that these functions possess the standard properties; in particular, they satisfy Hotelling's lemma--$h_1^x(p, w, q) \cdot A = \partial\pi_1(p, w, q, A)/\partial w$, etc. Suppose, alternatively, that the grower has decided to adopt CPM (T = 2). In a similar manner, we can define a conditional production function, $y_2 = f_2(x, z) \equiv f(x, z; 2)$; a conditional profit function analogous to (8); conditional profit-maximizing input demand functions, $x_2 = h_2^x(p, w, q)$ and $z_2 = h_2^z(p, w, q)$; a conditional profit-maximizing output supply function, $y_2 = h_2^y(p, w, q)$; and a conditional maximized profit function, $\pi_2 = \pi_2(p, w, q, A)$.

All of these conditional functions represent the grower's continuous choices conditional on his discrete choice. They can be related to his discrete choice and to the unconditional decision functions (2)-(6) as follows:

$$T = h^T(p, w, q, A) = \begin{cases} 1 \text{ if } \pi_1(p, w, q, A) \geq \pi_2(p, w, q, A) \\ 2, \text{ otherwise} \end{cases} \quad (13)$$

$$\pi = \pi(p, w, q, A) = \max\,[\pi_1(p, w, q, A), \pi_2(p, w, q, A)] \quad (14)$$

$$x = h^x(p, w, q, A) = \begin{cases} h_1^x(p, w, q) \text{ if } \pi_1(p, w, q, A) \geq \pi_2(p, w, q, A) \\ h_2^x(p, w, q), \text{ otherwise} \end{cases} \quad (15)$$

$$z = h^z(p, w, q, A) = \begin{cases} h_1^z(p, w, q) \text{ if } \pi_1(p, w, q, A) \geq \pi_2(p, w, q, A) \\ h_2^z(p, w, q), \text{ otherwise} \end{cases} \tag{16}$$

$$y = h^y(p, w, q, A) = \begin{cases} h_1^y(p, w, q) \text{ if } \pi_1(p, w, q, A) \geq \pi_2(p, w, q, A) \\ h_2^y(p, w, q), \text{ otherwise.} \end{cases} \tag{17}$$

If stochastic terms are added to the various conditional functions, (13)-(17) will be recognized as a set of switching regression equations, and they can be estimated by the statistical techniques recently developed for use with such regression models.

In our case, however, the available data on input prices, particularly for nonpesticide inputs, are somewhat limited; this forces us to take a slightly different tack. In effect, the system (13)-(17) represents a solution to the underlying profit maximization of the form,

$$\max_{T} \{\max_{x,z} \pi(x, z, T; p, w, q, A)\}. \tag{18}$$

As an alternative, we can seek a solution to the underlying profit maximization of the form,

$$\max_{z} [\max_{T} \{\max_{x} \pi(x, z, T; p, w, q, A)\}]. \tag{19}$$

The maximizations within the square brackets in (19) are conditional on the level of nonpesticide inputs, z. Since we have better data on the quantities of these nonpesticide inputs than on their prices, it is convenient to derive the discrete choice decision function for T and the continuous choice decision function for x by focusing on the inner maximizations in (19).

Accordingly, maximization of (8) with respect to x for fixed z yields a conditional profit-maximizing demand function for pesticide inputs, conditional on both the quantity of nonpesticide inputs and the adoption of IPM, of the form $x_1 = \tilde{h}_1^x(p, w, z)$, a conditional output supply function, $y_1 = \tilde{h}_1^y(p, w, z)$, and a conditional maximized profit function, $\pi_1 = \tilde{\pi}_1(p, w, z, A)$, analogous to (9), (11), and (12) above. Similarly, conditional on the quantity of nonpesticide inputs and the adoption of CPM, we obtain a profit-maximizing input demand function $x_2 = \tilde{h}_2^x(p, w, z)$ output supply function, $y_2 = \tilde{h}_2^y(p, w, z)$, and profit function $\pi_2 = \pi_2(p, w, z, A)$. The solution to the inner maximizations in (19) can then be written in the form

$$T = \tilde{h}^T(p, w, z, A) = \begin{cases} 1 \text{ if } \tilde{\pi}_1(p, w, z, A) \geq \tilde{\pi}_2(p, w, z, A) \\ 2, \text{ otherwise} \end{cases} \tag{20}$$

$$\pi = \tilde{\pi}(p, w, z, A) = \max\,[\tilde{\pi}_1(p, w, z, A), \tilde{\pi}_2(p, w, z, A)] \tag{21}$$

$$x = \tilde{h}^x(p, w, z, A) = \begin{cases} \tilde{h}_1^x(p, w, z) \text{ if } \tilde{\pi}_1(p, w, z, A) \geq \tilde{\pi}_2(p, w, z, A) \\ \tilde{h}_2^x(p, w, z), \text{ otherwise} \end{cases} \tag{22}$$

$$y = \tilde{h}^y(p, w, z, A) = \begin{cases} \tilde{h}_1^y(p, w, z) \text{ if } \tilde{\pi}_1(p, w, z, A) \geq \tilde{\pi}_2(p, w, z, A) \\ \tilde{h}_2^y(p, w, z), \text{ otherwise.} \end{cases} \tag{23}$$

The system (20)-(23), rather than (13)-(17), generates the switching regression model that we propose to estimate here.

In particular, we assume that the conditional production functions underlying (8) have the form

$$y_j = f_j(x, z) = G_j(z)\, x^{\alpha_j} \qquad j = 1, 2 \tag{24}$$

for some functions $G_j(z)$, $j = 1, 2$. In our empirical work, we will employ the following specification

$$G_j(z) = \theta_j \left(e^{\sum_{\ell=1}^{L} s_\ell \delta_{j\ell}} \right) \left(\prod_{m=1}^{M} z_m^{\gamma_{jm}} \right) \qquad j = 1, 2 \tag{25}$$

where $z = (z_1, \ldots, z_M)$ are the nonpesticide inputs and $s = (s_1, \ldots, s_L)$ is a vector of observable grower characteristics, such as education and farming experience, which, in effect, shift the intercept of the conditional production functions. Using (24), the maximization of (8) with respect to x yields

$$x_j = \tilde{h}_j^x(p, w, z) = \left[\alpha_j\, G_j(z)\, \frac{p}{w} \right]^{\frac{1}{1-\alpha_j}} \qquad j = 1, 2 \tag{26}$$

$$\begin{aligned} y_j = \tilde{h}_j^y(p, w, z) &= G_j(z) \left[\alpha_j\, G_j(z)\, \frac{p}{w} \right]^{\frac{\alpha_j}{1-\alpha_j}} \\ &= \left(\alpha_j \frac{p}{w} \right)^{\frac{\alpha_j}{1-\alpha_j}} G_j(z)^{\frac{1}{1-\alpha_j}} \qquad j = 1, 2 \end{aligned} \tag{27}$$

$$\pi_j = \tilde{\pi}_j(p, w, z, A) = [py_j - wx_j - qz]\, A - k(j)$$
$$\equiv [py_j - wx_j]\, A - F_j \qquad j = 1, 2 \tag{28}$$

where $F_j \equiv k(j) + q\, z\, A$, $j = 1, 2$. In order to simplify (28), we substitute (26) and (27) and make use of the fact that, for the Cobb-Douglas formulation in (24), the share of pesticide input costs in total revenues is α_j

$$w\, \tilde{h}^x_j(\cdot) = \alpha_j\, p\, \tilde{h}^y_j(\cdot)$$

or

$$w\left[\alpha_j\, G_j\, \frac{p}{w}\right]^{\frac{1}{1-\alpha_j}} = \alpha_j\, p\left[\alpha_j\, \frac{p}{w}\right]^{\frac{\alpha_j}{1-\alpha_j}} G_j^{\frac{1}{1-\alpha_j}} = [\alpha_j\, G_j\, p]^{\frac{1}{1-\alpha_j}}\, w^{\frac{\alpha_j}{\alpha_j - 1}}$$

to obtain

$$\pi_j = (1 - \alpha_j)\, py_j\, A - F_j$$
$$= (1 - \alpha_j)\, [G_j\, p]^{\frac{1}{1-\alpha_j}} \left(\frac{\alpha_j}{w}\right)^{\frac{\alpha_j}{1-\alpha_j}} A - F_j \qquad j = 1, 2. \tag{29}$$

In order to generate a statistical model, we need to add some stochastic elements to (26), (27), and (29). We introduce six stochastic terms, υ_1, υ_2, υ_3, υ_4, η_1, and η_2 into the conditional input demand, output supply, and profit functions as follows

$$x_j = \left[\alpha_j\, G_j(z)\, \frac{p}{w}\right]^{\frac{1}{1-\alpha_j}} e^{\upsilon_j} \qquad j = 1, 2 \tag{26'}$$

$$y_j = \left[\alpha_j \frac{p}{w}\right]^{\frac{\alpha_j}{1-\alpha_j}} G_j(z)^{\frac{1}{1-\alpha_j}} e^{\upsilon_{2+j}} \qquad j = 1, 2 \tag{27'}$$

$$\pi_j = (1 - \alpha_j) [G_j p]^{\frac{1}{1-\alpha_j}} \frac{\alpha_j}{w}^{\frac{\alpha_j}{1-\alpha_j}} A - F_j + \eta_j \qquad j = 1, 2. \tag{29'}$$

These random terms may be thought of as representing errors of measurement, unobservable or omitted variables, and random errors in the grower's optimization process. Actually, the two terms, η_1 and η_2, are not needed separately; we only need their difference, $\upsilon_5 \equiv \eta_1 - \eta_2$, since we do not observe π_1 and π_2 directly but only the sign of their difference.

$$\Delta\pi \equiv \pi_1 - \pi_2 = \left\{(1 - \alpha_1) [G_1 p]^{\frac{1}{1-\alpha_1}} \left(\frac{\alpha_1}{w}\right)^{\frac{\alpha_1}{1-\alpha_1}} - (1 - \alpha_2) [G_2 p]^{\frac{1}{1-\alpha_2}} \left(\frac{\alpha_2}{w}\right)^{\frac{\alpha_2}{1-\alpha_2}}\right\} A - (F_1 - F_2) + \upsilon_5. \tag{30}$$

To simplify the notation, it will be convenient to rewrite (26'), (27'), and (30) as

$$x_j = e^{\mu^x_j + \upsilon_j} \qquad j = 1, 2 \tag{31a}$$

$$y_j = e^{\mu^y_j + \upsilon_{2+j}} \qquad j = 1, 2 \tag{31b}$$

$$\Delta\pi = \mu^\pi + \upsilon_5 \tag{31c}$$

where μ^{π} contains all the nonstochastic terms on the right-hand side of (30) and similarly for μ_j^x and μ_j^y. Thus, making use of (25),

$$\mu_j^x = \left[\frac{1}{1-\alpha_j}(\ln\alpha_j + \ln\theta_j)\right] + \frac{1}{1-\alpha_j}\left[\sum_{\ell=1}^{L} s_j\,\delta_{j\ell}\right] + \frac{1}{1-\alpha_j}\left[\sum_{m=1}^{M}\gamma_{jm}\ln z_m\right] + \frac{1}{1-\alpha_j}\ln\left(\frac{p}{w}\right) \qquad j = 1, 2 \tag{32}$$

$$\mu_j^y = \left[\frac{\alpha_j}{1-\alpha_j}\ln\alpha_j + \frac{1}{1-\alpha_j}\ln\theta_j\right] + \frac{1}{1-\alpha_j}\left[\sum_{\ell=1}^{L} s_j\,\delta_{j\ell}\right] + \frac{1}{1-\alpha_j}\left[\sum_{m=1}^{M}\gamma_{jm}\ln z_m\right] + \left(\frac{\alpha_j}{1-\alpha_j}\right)\ln\left(\frac{p}{w}\right) \qquad j = 1, 2 \tag{33}$$

and

$$\mu^{\pi} = A\left\{(1-\alpha_1)\left[p\sigma_1\, e^{\sum_{\ell=1}^{L} s_\ell\delta_{1\ell}} \prod_{m=1}^{M} z_m^{\gamma_{1m}}\right]^{\frac{1}{1-\alpha_1}}\left(\frac{\alpha_1}{w}\right)^{\frac{\alpha_1}{1-\alpha_1}} - (1-\alpha_2)\left[p\sigma_2 e^{\sum_{\ell=1}^{L} s_\ell\delta_{2\ell}} \prod_{m=1}^{M} z_m^{\gamma_{2m}}\right]^{\frac{1}{1-\alpha_2}}\left(\frac{\alpha_2}{w}\right)^{\frac{\alpha_2}{1-\alpha_2}}\right\} - (F_1 - F_2). \tag{34}$$

Accordingly, taking logarithms, the statistical switching regression may be written as

$$T = \begin{cases} 1 \text{ if } -\upsilon_5 \leq \mu^{\pi} \\ 2, \text{ otherwise} \end{cases} \tag{35a}$$

$$\ln x = \begin{cases} \mu_1^x + \upsilon_1 \text{ if } T = 1 \\ \mu_2^x + \upsilon_2 \text{ if } T = 2 \end{cases} \tag{35b}$$

$$\ln y = \begin{cases} \mu_1^y + \upsilon_3 \text{ if } T = 1 \\ \mu_2^y + \upsilon_4 \text{ if } T = 2. \end{cases} \tag{35c}$$

The model is closed by specifying a joint distribution for $\upsilon \equiv (\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_4, \upsilon_5)$. In our application, we will assume that υ is independently and identically distributed (i.i.d.) across observations as multivariate normal with mean zero and some variance-covariance matrix $\Sigma \equiv \{\sigma_{ij}\}$. Accordingly, the parameters to be estimated are $\beta \equiv (\alpha_1, \alpha_2, \delta_{11}, \ldots, \delta_{1L}, \delta_{21}, \ldots, \delta_{2L}, \gamma_{11}, \ldots, \gamma_{1M}, \gamma_{21}, \ldots, \gamma_{2M}, \theta_1, \theta_2)$ and the elements of Σ.

Two features of this model are worth commenting on. First, we have allowed attributes of the grower to influence the production functions, and we have left open the possibility that their influence differs according to the type of pest control strategy. One hypothesis we propose to test is that $\delta_{1\ell} = \delta_{2\ell}$ for all ℓ--i.e., the effect of grower characteristics, such as age or farming experience, on the output levels attained by the grower is the <u>same</u> regardless of which pest control strategy he adopts. Similarly, one can test the hypothesis that the other components of the production function--θ, α, or the γ_M's--are the same across pest control strategies. In this way, we can pinpoint how the difference in pest control strategy affects agricultural productivity. Second, it will be shown in the next section that four off-diagonal elements of Σ (namely, σ_{12}, σ_{14}, σ_{23}, and σ_{24}) are not identifiable and cannot be estimated; all of the other elements of Σ are, in principle, estimable. The presence of covariance terms such as σ_{13}, σ_{15}, and σ_{35} allows for the possibility that unobservable factors (such as unmeasured grower characteristics) that influence a grower's output levels also affect his input choices and his decision to adopt IPM versus CPM; we can also test the hypothesis that these cross-effects are zero. Thus, the model (35) allows for considerable flexibility in modeling observed grower behavior.

III. Estimation

Given observations on a sample of N growers of whom N_1 adopt IPM and $N_2 \equiv N - N_1$ adopt CPM, the log-likelihood function for the model (35) is

$$\mathcal{L}(\beta, \Sigma) = \sum_{i=1}^{N_1} \log \ell_i^1(\beta, \Sigma) + \sum_{i=N_1+1}^{N} \log \ell_i^2(\beta, \Sigma) \tag{36}$$

wherefor $i = 1, \ldots, N_1$

$$\ell_i^1(\beta, \Sigma) = b\left(\frac{\ln x_i - \mu_{1i}^x}{\sigma_{11}^{1/2}}, \frac{\ln y_i - \mu_{1i}^y}{\sigma_{33}^{1/2}}; \rho_{13}\right)^*$$

$$* \, \Phi \; \frac{\dfrac{\mu_i^{\pi}}{\sigma_{55}^{1/2}} + \left(1 - \rho_{13}^2\right)^{-1} \left(\rho_{15} - \rho_{35}\,\rho_{13}\right) \dfrac{\ln x_i - \mu_{1i}^x}{\sigma_{11}^{1/2}}}{\left(1 - \rho_{13}^2\right)^{-1/2} \left\{1 - \rho_{13}^2 - \rho_{15}^2 - \rho_{35}^2 + 2\rho_{13}\,\rho_{15}\,\rho_{35}\right\}^{1/2}} \tag{37a}$$

$$* \, \Phi \; \frac{+ \left(\rho_{35} - \rho_{13}\,\rho_{15}\right) \dfrac{\ln y_i - \mu_{1i}^y}{\sigma_{33}^{1/2}}}{\left(1 - \rho_{13}^2\right)^{-1/2} \left\{1 - \rho_{13}^2 - \rho_{15}^2 - \rho_{35}^2 + 2\rho_{13}\,\rho_{15}\,\rho_{35}\right\}^{1/2}}$$

and for $i = N_1 + 1, \ldots, N$

$$\ell_i^2(\beta, \Sigma) = b\left(\frac{\ln x_i - \mu_{2i}^x}{\sigma_{22}^{1/2}}, \frac{\ln y_i - \mu_{2i}^y}{\sigma_{44}^{1/2}}; \rho_{24}\right) *$$

$$\Phi \left[\frac{\dfrac{\mu_i^{\pi}}{\sigma_{55}^{1/2}} - \left(1 - \rho_{24}^2\right)^{-1} \left\{ \left(\rho_{25} - \rho_{45}\,\rho_{24}\right) \dfrac{\ln x_i - \mu_{2i}^x}{\sigma_{22}^{1/2}} \right.}{\left(1 - \rho_{24}^2\right)^{-1} \left\{1 - \rho_{24}^2 - \rho_{25}^2 - \rho_{45}^2 + 2\rho_{24}\,\rho_{25}\,\rho_{45}\right\}^{1/2}} \right] \tag{37b}$$

$$\frac{\left. + \left(\rho_{45}\,\rho_{24}\,\rho_{25}\right) \dfrac{\ln y_i - \mu_{2i}^y}{\sigma_{44}^{1/2}} \right\}}{\left(1 - \rho_{24}^2\right)^{-1} \left\{1 - \rho_{24}^2 - \rho_{25}^2 - \rho_{45}^2 + 2\rho_{24}\,\rho_{25}\,\rho_{45}\right\}^{1/2}}$$

where $\rho_{ij} \equiv \sigma_{ij}/(\sigma_{ii}\,\sigma_{jj})^{1/2}$, $b(z_1, z_2; \rho)$ is the density function of a standard bivariate normal with correlation coefficient ρ, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Observe that the terms σ_{12}, σ_{14}, σ_{23}, and σ_{34} (or ρ_{12}, ρ_{14}, ρ_{23}, and ρ_{34}) do not appear in the likelihood function (37a, b); therefore, they are not identifiable from the observed data. One approach to estimation is to maximize (36) directly with respect to β and the remaining elements of Σ. Following the argument of Amemiya ( ), it can be shown that the maximum likelihood estimator (MLE) is consistent and asymptotically normal and efficient. In practice, however, the normal equations may have multiple roots and, unless one starts from an initial consistent estimator, there is no guarantee of convergence to the global maximum. Moreover, because (37a, b) is extremely nonlinear in the parameters, it is often computationally burdensome to obtain the MLE directly. Also, in small samples there is no guarantee that one will obtain an estimate of Σ which is finite or positive definite [for a technical explanation, see Tsur (1983b)].

As an alternative, one can employ the two-stage estimation procedure originated by Heckman (1976) and Lee and Trost (1978). The first stage is maximum likelihood estimation of the probit model for the discrete choice of pest control technology. From (35a), the likelihood function for this probit model is

$$\ell'(\beta, \Sigma) = \prod_{i=1}^{N_1} \Phi\left(\frac{\mu_i^\pi}{\sigma_{55}^{1/2}}\right) \prod_{i=N_1+1}^{N} \Phi\left(\frac{-\mu_i^\pi}{\sigma_{55}^{1/2}}\right). \tag{38}$$

Unlike conventional probit models, this model is nonlinear in the parameters--see (34)--and, in principle, it permits one to obtain a separate estimate of σ_{55}. Direct maximization of (38) will yield estimates of σ_{55} and all the elements in β, which we denote σ_{55} and β. However, these estimates ignore the information about β which is contained in the data on the continuous choices--the levels of pesticide input and output supply. This information is incorporated in the second stage which is based on the observation that

$$\mathcal{E}\{\ln x | T = 1\} = \mu_1^x + \mathcal{E}\{\upsilon_1 | -\upsilon_5 \leq \mu^\pi\} \tag{39a}$$

$$\mathcal{E}\{\ln x | T = 2\} = \mu_2^x + \mathcal{E}\{\upsilon_2 | -\upsilon_5 > \mu^\pi\} \tag{39b}$$

$$\mathcal{E}\{\ln y | T = 1\} = \mu_1^y + \mathcal{E}\{\upsilon_3 | -\upsilon_5 \leq \mu^\pi\} \tag{39c}$$

$$\mathcal{E}\{\ln y | T = 2\} = \mu_2^y + \mathcal{E}\{\upsilon_4 | -\upsilon_5 > \mu^\pi\} \tag{39d}$$

where

$$\mathcal{E}\{\upsilon_j | -\upsilon_5 \leq \mu^\pi\} = \frac{\sigma_{j5}\,\phi\left(\frac{\mu^\pi}{\sigma_{55}^{1/2}}\right)}{\Phi\left(\frac{\mu^\pi}{\sigma_{55}^{1/2}}\right)} \equiv \sigma_{j5}\,\lambda^+ \qquad j = 1, 3 \tag{39e}$$

and

$$\mathcal{E}\{\upsilon_j | -\upsilon_5 > \mu^\pi\} = \frac{-\sigma_{j5}\,\phi\left(\frac{\mu^\pi}{\sigma_{55}^{1/2}}\right)}{\Phi\left(\frac{\mu^\pi}{\sigma_{55}^{1/2}}\right)} \equiv -\sigma_{j5}\,\lambda^- \qquad j = 2, 4 \tag{39f}$$

where $\phi(\cdot)$ is the standard normal density. Accordingly, we can set up the following regression models:

$$\ln x_i = \mu^x_{1i} + \sigma_{15}\,\lambda^+_i + \omega_{1i} \tag{40a}$$

$$\ln x_i = \mu^x_{2i} - \sigma_{25}\,\lambda^-_i + \omega_{2i} \tag{40b}$$

$$\ln y_i = \mu^y_{1i} + \sigma_{35}\,\lambda^+_i + \omega_{3i} \tag{40c}$$

$$\ln y_i = \mu^y_{2i} - \sigma_{45}\,\lambda^-_i + \omega_{4i} \tag{40d}$$

where $\mathcal{E}\{\omega_{1i}|T = 1\} = \mathcal{E}\{\omega_{2i}|T_i = 2\} = \mathcal{E}\{\omega_{3i}|T = 1\} = \mathcal{E}\{\omega_{4i}|T = 2\} = 0$.

In order to fit these regressions, we make use of the estimates $\hat{\beta}$ and $\hat{\sigma}_{55}$ obtained from the first stage. Let $\hat{\mu}^\pi_i$ be the estimate of μ^π_i constructed using $\hat{\beta}$, and let $\hat{\lambda}^+_i$ and $\hat{\lambda}^-_i$ be the estimates of λ^+_i and λ^-_i constructed using $\hat{\mu}^\pi_i$ and $\hat{\sigma}_{55}$. Before proceeding to fit the regression model, we need to take account of the cross-equation coefficient restrictions implicit in (32) and (33). Observe that (32) and (33) contain the same explanatory variables and differ only in their constant term and in the coefficient of $\ln(\omega/p)$. We can deal with the latter problem (but not the former) by replacing the dependent variable in (40c, d) with $\ln \tilde{y}_i \equiv \ln y_i + \ln (p_i/\omega_i)$. Then, for $i = 1, \ldots, N_1$, we run the following regression

$$\begin{pmatrix} \ln x_i \\ \\ \ln y_i \end{pmatrix} = \begin{pmatrix} 1, 0, s_{1i}, \ldots, s_{Li}, z_{1i}, \ldots, z_{Mi}, \ln\left(\frac{p_i}{\omega_i}\right), \hat{\lambda}_i^+, 0 \\ \\ 0, 1, s_{1i}, \ldots, s_{Li}, z_{1i}, \ldots, z_{Mi}, \ln\left(\frac{p_i}{\omega_i}\right), 0, \hat{\lambda}_i^+ \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{L+1} \\ c_{L+2} \\ \vdots \\ c_{L+M+1} \\ c_{L+M+2} \\ c_{L+M+3} \\ c_{L+M+4} \end{pmatrix} \quad (41)$$

$$+ \begin{pmatrix} \omega_{1i} \\ \\ \omega_{2i} \end{pmatrix} .$$

The mapping from the coefficients $c_0, \ldots, c_{L+M+4}$ to the underlying coefficients in (32) and (33) is as follows

$$c_0 = \left(\frac{1}{1 - \alpha_1}\right) [\ln \alpha_1 + \ln \theta_1] \qquad c_1 = \left(\frac{\alpha_1}{1 - \alpha_1}\right) \ln \alpha_1 + \frac{1}{1 - \alpha_1} \ln \theta_1 \quad (42a)$$

$$c_2 = \frac{\delta_{11}}{(1 - \alpha_1)}, \ldots, c_{L+1} = \frac{\delta_{1L}}{(1 - \alpha_1)} \quad (42b)$$

$$c_{L+2} = \frac{\gamma_{11}}{(1 - \alpha_1)}, \ldots, c_{L+M+1} = \frac{\gamma_{1M}}{(1 - \alpha_1)} \quad (42c)$$

$$c_{L+M+2} = \frac{1}{1 - \alpha_1} \quad (42d)$$

$$c_{L+M+3} = \sigma_{15} \qquad c_{L+M+4} = \sigma_{35} . \quad (42e)$$

For $i = N_1 + 1, \ldots, N$ we run a regression similar to (41) except that we substitute $-\hat{\lambda}_i^-$ for $\hat{\lambda}_i^+$ as the last regressor variable. The coefficients for this regression will be denoted $d_0, d_1, \ldots, d_{L+M+4}$; they are related to the underlying coefficients in (32) and (33) in a manner analogous to (42a-e)--for example, $d_{L+M+2} = 1/(1 - \alpha_2)$, $d_{L+M+3} = \sigma_{25}$, and $d_{L+M+4} = \sigma_{45}$. Since both of these regressions are linear in the coefficients, we can employ ordinary least squares (OLS). Denote the resulting coefficient estimates by $c_0', \ldots, c_{L+M+4}'$ and $d_0', \ldots, d_{L+M+4}'$. The relations in (42a-e) can be used to obtain estimates of the underlying coefficients in the model. For example,

$$\alpha_1' = \frac{c_{L+M+2}' - 1}{c_{L+M+2}'} \tag{43a}$$

$$\delta_{11}' = \frac{c_2'}{c_{L+M+2}'} \qquad \gamma_{11}' = \frac{c_{L+2}'}{c_{L+M+2}'}, \text{ etc.,} \tag{43b}$$

and, similarly, for α_2', δ_{21}', γ_{21}', etc. However, the coefficients θ_1 and θ_2 are overidentified since, from (42a), we obtain two separate estimates of them; for θ_1, we have

$$\ln \theta_1' = \frac{c_0'}{c_{L+M+2}'} - \ln\left(\frac{c_{L+M+2}' - 1}{c_{L+M+2}'}\right) \tag{43c}$$

and

$$\ln \theta_1' = \frac{c_1'}{c_{L+M+2}'} - \left[\frac{c_{L+M+2}' - 1}{c_{L+M+2}'}\right] \ln\left(\frac{c_{L+M+2}' - 1}{c_{L+M+2}'}\right) \tag{43d}$$

and, similarly, for θ_2'. From c_{L+M+3}' and c_{L+M+4}', we obtain σ_{15}' and σ_{35}' and, similarly, for σ_{25}' and σ_{45}'. Estimates of the other identifiable elements of Σ--σ_{11}', σ_{13}', σ_{33}', σ_{22}', σ_{24}', and σ_{44}'--can be obtained from the regression residuals of (41) along the lines indicated by Lee and Trost (1978, p. 361 and 362).

Lee and Trost prove that these estimates are consistent but not efficient. Moreover, the variance-covariance matrix for the estimates c_0', ..., c_{L+M+4}' and d_0', ..., d_{L+M+4}' generated by the OLS regressions is incorrect because the regressors included estimated variables $(\hat{\lambda}_i^+, \hat{\lambda}_i^-)$. Following the suggestions of Greene (1983), we can employ White's (1980) heteroscedasticity-consistent estimator of this variance-covariance matrix which is readily computed from the regression residuals. However, this does not give us standard errors for the estimates of σ_{11}, σ_{13}, σ_{33}, σ_{22}, σ_{24}, and σ_{44} which may be needed testing hypotheses on Σ. Accordingly, following the suggestion of Lee and Trost, we can take the estimates β' and Σ' and use them as starting values for a direct maximization of the likelihood function (36). Since they are consistent, a single Newton-Raphson iteration will provide estimates of β and Σ which have the same asymptotic distribution as the global MLE. Thus, these so-called two-step maximum likelihood estimates are consistent and asymptotically normal and efficient, and their variance-covariance matrix is consistently estimated by the information matrix.

A third approach to estimating the switching regression model (35) involves the application of the EM algorithm of Dempster, Laird, and Rubin (1977). The extension of the EM estimator to switching regressions is described in detail in Tsur (1983a); here we provide only a brief summary. For this purpose, it is convenient to introduce some new notation and rewrite (35) as, for i = 1, ..., N,

$$Y_{1i}^{*} = \mu_{1i}^{X} + \upsilon_{1i} \tag{44a}$$

$$Y_{2i}^{*} = \mu_{2i}^{X} + \upsilon_{2i} \tag{44b}$$

$$Y_{3i}^{*} = \mu_{1i}^{Y} + \upsilon_{3i} \tag{44c}$$

$$Y_{4i}^{*} = \mu_{2i}^{Y} + \upsilon_{4i} \tag{44d}$$

$$Y_{5i}^{*} = \mu_{i}^{\pi} + \upsilon_{5i} \tag{44e}$$

$$Y_{1i} = \begin{cases} Y_{1i}^{*} \text{ if } Y_{5i}^{*} \geq 0 \\ \text{not observed otherwise} \end{cases} \tag{44f}$$

$$Y_{2i} = \begin{cases} \text{not observed if } Y_{5i}^{*} \geq 0 \\ Y_{2i}^{*}, \text{ otherwise} \end{cases} \tag{44g}$$

$$Y_{3i} = \begin{cases} Y_{3i}^{*} \text{ if } Y_{5i}^{*} \geq 0 \\ \text{not observed otherwise} \end{cases} \tag{44h}$$

$$Y_{4i} = \begin{cases} \text{not observed if } Y_{5i}^{*} \geq 0 \\ Y_{4i}^{*}, \text{ otherwise} \end{cases} \tag{44i}$$

$$Y_{5i} = \begin{cases} 1 \text{ if } Y_{5i}^{*} > 0 \\ 0, \text{ otherwise.} \end{cases} \tag{44j}$$

In effect, $Y_1^* \equiv \ln x_1$, $Y_2^* \equiv \ln x_2$, $Y_3^* \equiv \ln y_1$, $Y_4^* \equiv \ln y_2$ and $Y_5^* \equiv \Delta\pi$. In the terminology of Dempster, Laird, and Rubin, the variables $Y_i^* \equiv (Y_{1i}^*, Y_{2i}^*, Y_{3i}^*, Y_{4i}^*, Y_{5i}^*)$ are the "complete" data; they are not observed directly, but only indirectly through the observed or "incomplete" data $Y_i \equiv (Y_{1i}, Y_{2i}, Y_{3i}, Y_{4i}, Y_{5i})$. The logarithm of the joint density of the <u>complete</u> data is

$$\mathcal{L}^*(\beta, \Sigma) = \sum_{i=1}^{N} \log \ell^*(Y_i|\beta, \Sigma) \tag{45}$$

$$= \frac{N}{2} \log|\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^{N} \operatorname{tr} \left\{ \Sigma^{-1} \left(Y_i^* - \mu_i \right) \left(Y_i^* - \mu_i \right)' \right\}$$

where $\mu_i \equiv (\mu_{1i}^x, \mu_{2i}^x, \mu_{1i}^y, \mu_{2i}^y, \mu_i^\pi)$. The EM algorithm involves a sequence of iterations, each iteration consisting of two steps. At the K + 1st iteration, given parameter estimates $\hat{\beta}^K$ and $\hat{\Sigma}^K$, in the E-step one computes the <u>expectation</u> of the log-likelihood function for the complete data <u>conditional</u> on the observed data

$$Q(\beta, \Sigma|\hat{\beta}^K, \hat{\Sigma}^K) \equiv \mathcal{E}\left\{\mathcal{L}^*(\beta, \Sigma)|Y_1, \ldots, Y_N, \hat{\beta}^K, \hat{\Sigma}^K\right\} \tag{46}$$

and in the M-step one finds $\hat{\beta}^{K+1}$, $\hat{\Sigma}^{K+1}$ which solve

$$\max_{\beta,\Sigma} Q(\beta, \Sigma|\hat{\beta}^K, \hat{\Sigma}^K).$$

The algorithm is started with some initial set of parameter estimates, and the iteration of E- and M-steps is continued until a convergence criterion is satisfied. Dempster, Laird, and Rubin prove that these iterations converge to a root of the normal equations for maximizing the log-likelihood function

(36). Thus, the EM algorithm provides an alternative procedure for obtaining the MLE which avoids some of the computational difficulties associated with direct maximization of (36).

In performing the E-step, we observe that

$$\mathcal{E}\left\{\left(Y_i^* - \mu_i\right)\left(Y_i^* - \mu_i\right)' \middle| \; Y_i;\, \hat{\beta}^K,\, \hat{\Sigma}^K\right\} = \left(\bar{Y}_i^K - \mu_i\right)\left(\bar{Y}_i^K - \mu_i\right)' + \tau_i^K \qquad (47)$$

where $\bar{Y}_i^K \equiv \mathcal{E}\left\{Y_i^* \middle| Y_i;\, \hat{\beta}^K,\, \hat{\Sigma}^K\right\}$ and $\tau_i^K \equiv \mathcal{E}\left\{\upsilon_i\, \upsilon_i' \middle| Y_i;\, \hat{\beta}^K,\, \hat{\Sigma}^K\right\}$. Thus, for example,

$$\bar{Y}_{1i}^K = \begin{cases} Y_{1i} & \text{if } Y_{5i} = 1 \\ \mu_{1i}^{xK} + \mathcal{E}\left\{\upsilon_{1i} \middle| Y_{5i} = 0;\, \beta^K,\, \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 0 \end{cases}$$

$$= \begin{cases} Y_{1i} & \text{if } Y_{5i} = 1 \\ \hat{\mu}_{1i}^{xK} + \mathcal{E}\left\{\upsilon_{1i} \middle| \upsilon_{5i} < -\hat{\mu}_i^{\pi k};\, \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 0 \end{cases}$$

$$= \begin{cases} Y_{1i} & \text{if } Y_{5i} = 1 \qquad (48a) \\ \hat{\mu}_{1i}^{xK} - \hat{\rho}_{15}^K \left(\hat{\sigma}_{11}^K\right)^{1/2} \lambda^-\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{-1/2}}\right) & \text{if } Y_{5i} = 0 \qquad (48b) \end{cases}$$

$$\bar{Y}_{2i}^K = \begin{cases} \hat{\mu}_{2i}^{xK} + \hat{\rho}_{25}^K \left(\hat{\sigma}_{22}^K\right)^{1/2} \lambda^+\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{1/2}}\right) & \text{if } Y_{5i} = 1 \qquad (49a) \\ Y_{2i} & \text{if } Y_{5i} = 0 \qquad (49b) \end{cases}$$

where $\hat{\mu}_{1i}^{xK}$ is (32) evaluated using the coefficient estimates $\hat{\beta}^K$, and similarly for $\hat{\mu}_{2i}^{xK}$ and $\hat{\mu}_i^{\pi K}$, where $\lambda^+(a) \equiv \phi(a)/\Phi(-a)$, and $\lambda^-(a) \equiv \phi(a)/\Phi(a)$. The formulas for $\bar{Y}_{3i}^K$ and $\bar{Y}_{4i}^K$ are similar to (48) and (49). The formula for $\bar{Y}_{5i}^K$ is

$$\bar{Y}_{5i}^K = \begin{cases} \hat{\mu}_i^\pi + \mathcal{E}\left\{\upsilon_5 \middle| Y_{5i} = 1;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 1 \\ \hat{\mu}_i^\pi + \mathcal{E}\left\{\upsilon_5 \middle| Y_{5i} = 0;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 0 \end{cases}$$

$$= \begin{cases} \hat{\mu}_i^\pi + \left(\hat{\sigma}_{55}^K\right)^{1/2} \lambda^+\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{1/2}}\right) & \text{if } Y_{5i} = 1 \\ \hat{\mu}_i^\pi - \left(\hat{\sigma}_{55}^K\right)^{1/2} \lambda^-\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{1/2}}\right) & \text{if } Y_{5i} = 0 \end{cases} \tag{50}$$

Similarly,

$$\tau_{11i}^K = \begin{cases} 0 & \text{if } Y_{5i} = 1 \\ \mathcal{E}\left\{\upsilon_1^2\ Y_{5i} = 0;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 0 \end{cases}$$

$$= \begin{cases} 0 & \text{if } Y_{5i} = 1 & \text{(51a)} \\ \hat{\sigma}_{11}^K\left[1 + \left(\hat{\rho}_{15}^K\right)^2 \Lambda^-\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 0 & \text{(51b)} \end{cases}$$

$$\tau_{13i}^K = \begin{cases} 0 & \text{if } Y_{5i} = 1 & \text{(51c)} \\ \hat{\sigma}_{13}^K\left[1 + \dfrac{\hat{\sigma}_{15}^K\ \hat{\sigma}_{35}^K}{\hat{\sigma}_{55}^K\ \hat{\sigma}_{13}^K} \Lambda^-\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}_{55}^K\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 0 & \text{(51d)} \end{cases}$$

$$\tau^K_{15i} = \begin{cases} 0 & \text{if } Y_{5i} = 1 \quad (51e) \\ \hat{\sigma}^K_{15}\left[1 + \Lambda^{-1}\left(\dfrac{-\hat{\mu}^{\pi K}_i}{\hat{\sigma}^{K\,1/2}_{55}}\right)\right] & \text{if } Y_{5i} = 0 \quad (51f) \end{cases}$$

and

$$\tau^K_{22i} = \begin{cases} \xi\left\{\upsilon^2_2 \middle| Y_{5i} = 1;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 1 \\ 0 & \text{if } Y_{5i} = 0 \end{cases}$$

$$= \begin{cases} \hat{\sigma}^K_{22}\left[1 + \left(\hat{\rho}^K_{25}\right)^2 \Lambda^+\left(\dfrac{-\hat{\mu}^{\pi K}_i}{\left(\hat{\sigma}^K_{55}\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 1 \quad (52a) \\ 0 & \text{if } Y_{5i} = 0 \quad (52b) \end{cases}$$

$$\tau^K_{24} = \begin{cases} \hat{\sigma}^K_{24}\left(1 + \dfrac{\hat{\sigma}^K_{25}\,\hat{\sigma}^K_{45}}{\hat{\sigma}^K_{55}\,\hat{\sigma}^K_{24}}\Lambda^+\left(\dfrac{\hat{\mu}^{\pi K}_i}{\left(\hat{\sigma}^K_{55}\right)^{1/2}}\right)\right) & \text{if } Y_{5i} = 1 \quad (52c) \\ 0 & \text{if } Y_{5i} = 0 \quad (52d) \end{cases}$$

$$\tau^K_{25} = \begin{cases} \hat{\sigma}^K_{25}\left[1 + \Lambda^+\left(\dfrac{-\hat{\mu}^{\pi K}_i}{\left(\hat{\sigma}^K_{55}\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 1 \quad (52e) \\ 0 & \text{if } Y_{5i} = 0 \quad (52f) \end{cases}$$

where $\Lambda^+(a) \equiv a\,\lambda^+(a) - \lambda^+(a)^2$, and $\Lambda^-(a) \equiv -\,a\,\lambda^-(a) - \lambda^-(a)^2$. The formulas for τ^K_{33i} and τ^K_{35i} are similar to (51a, b, e, f) and those for τ^K_{44i} and τ^K_{45i} are similar to (52 a, b, e, f). By virtue of the non-identifiability of these terms, we set $\tau^K_{12i} = \tau^K_{14i} = \tau^K_{23i} = \tau^K_{34i} = 0$. Finally,

$$\tau^K_{55i} = \begin{cases} \mathcal{E}\left\{\upsilon_5^2 \mid Y_{5i} = 1;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 1 \\ \mathcal{E}\left\{\upsilon_5^2 \mid Y_{5i} = 0;\ \hat{\beta}^K,\ \hat{\Sigma}^K\right\} & \text{if } Y_{5i} = 0 \end{cases} \tag{53}$$

$$= \begin{cases} \hat{\sigma}^K_{55}\left[1 + \Lambda^+\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}^K_{55}\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 1 \\ \hat{\sigma}^K_{55}\left[1 + \Lambda^-\left(\dfrac{-\hat{\mu}_i^{\pi K}}{\left(\hat{\sigma}^K_{55}\right)^{1/2}}\right)\right] & \text{if } Y_{5i} = 0. \end{cases}$$

The interpretation of (48) and (51) is as follows. For an observation i for which $Y_{5i} = 1$, we actually observe Y^*_{1i}, since, in this case, $Y_{1i} = Y^*_{1i}$. Therefore, our expectation of Y^*_{1i} given the observed Y_{1i} is simply Y_{1i} and all the terms τ^K_{11i}, τ^K_{13i}, etc., become zero. However, if $Y_{5i} = 0$, we do not observe Y^*_{1i}. In this case our expectation of Y^*_{1i} conditional on the fact that it is not observed is given by (48b), and our expectation of υ^2_{1i}, $\upsilon_{1i}\,\upsilon_{3i}$, etc., is given by (51b, d, etc.). This is how we "fill in" the missing values of Y^*_{1i}, and similarly with Y^*_{2i}, Y^*_{3i}, etc. We end up with a full set of N observations or filled-in values for all five variables $Y^*_{1i}, \ldots, Y^*_{5i}$.

Next, we plug (47) into (46) to obtain

$$Q\ \beta, \Sigma\quad \hat{\beta}^K, \hat{\Sigma}^K\ = \frac{N}{2}\log\quad \Sigma^{-1}\quad -\frac{1}{2}\sum_{i=1}^{N} \mathrm{tr}\quad \Sigma^{-1}\quad \tilde{Y}_i^K - \mu_i\quad \tilde{Y}_i^K - \mu_i\ ^1 \qquad (54)$$
$$+\ \tau_i^K$$

In the M-step, we maximize (54) with respect to β and Σ to obtain β^{K+1}, Σ^{K+1}. Setting the derivative of (54) with respect to Σ equal to zero yields

$$\hat{\Sigma}^{K+1} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\tilde{Y}_i^K - \mu_i^{K+1}\right)\left(\tilde{Y}_i^K - \mu_i^{K+1}\right)' + \tau_i^K\right]. \qquad (55)$$

However, the maximization of (54) with respect to β is more complex because, while μ_{1i}, μ_{2i}, μ_{3i}, and μ_{4i} are linear in (transforms of) the elements of β, μ_{5i} is a nonlinear function of β. Hence, at each stage K, an iterative solution procedure would be required to perform the M-step with respect to β since all the elements of β already appear in μ_{1i}, ..., μ_{4i}. We avoid this problem by omitting the terms $(\tilde{Y}_{5i}^K - \mu_{5i})$ from (55) and maximizing μ_{1i}, ..., μ_{4i} with respect to β. In effect, we are estimating filled-in versions of equations (44f, g, h, i) and omitting the filled-in version of (44j). In order to allow for the cross-equation restrictions on coefficients, we proceed in a manner similar to that used in the regression model (41). We replace $\tilde{Y}_{2i}^K$ with $\hat{Y}_{2i}^K \equiv \tilde{Y}_{2i}^K + \ln\,(p_i/\omega_i)$ and $\tilde{\mu}_{2i}$ with $\mu_{2i} \equiv \mu_{2i} +$ $\ln\,(p_i/\omega_i)$, which is legitimate since $(\hat{Y}_{2i}^K - \tilde{\mu}_{2i}) = (\tilde{Y}_{2i}^K - \mu_{2i})$, and similarly for $\tilde{Y}_{4i}^K$ and μ_{4i}. Then, for i = 1, ..., N, we have

$$\begin{pmatrix} \mu_{1i} \\ \tilde{\mu}_{2i} \end{pmatrix} = \begin{bmatrix} 1, 0, s_{1i}, \ldots, s_{Li}, z_{1i}, \ldots, z_{Mi}, \ln\left(\frac{p_i}{w_i}\right) \\ 1, 0. s_{1i}, \ldots, s_{Li}, z_{1i}, \ldots, z_{Mi}, \ln\left(\frac{p_i}{w_i}\right) \end{bmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{L+1} \\ c_{L+2} \\ \vdots \\ c_{L+M+1} \\ c_{L+M+2} \end{pmatrix} \qquad (56a)$$

$$\equiv X_i \, c$$

where the elements of c are related to the underlying coefficients in β by (42a) through (42d). Similarly, for i = 1, ..., N, we have

$$\begin{pmatrix} \mu_{3i} \\ \tilde{\mu}_{4i} \end{pmatrix} = X_i \; d. \qquad (56b)$$

Thus,

$$\tilde{\mu}_i \equiv \begin{pmatrix} \mu_{1i} \\ \tilde{\mu}_{2i} \\ \mu_{3i} \\ \tilde{\mu}_{4i} \end{pmatrix} = \begin{pmatrix} X_i & 0 \\ 0 & X_i \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} \equiv \tilde{X}_i \begin{pmatrix} c \\ d \end{pmatrix}. \qquad (57)$$

Finally, define

$$Y^K_{ci} \equiv \begin{pmatrix} Y^K_{ic} \\ \tilde{Y}^K_{2i} \end{pmatrix}, \quad Y^K_{di} \equiv \begin{pmatrix} Y^K_{3i} \\ \tilde{Y}^K_{4i} \end{pmatrix}, \quad \tilde{Y}^K_i \equiv \begin{pmatrix} Y^K_{ci} \\ Y^K_{di} \end{pmatrix}$$

$$\tilde{Y}^K \equiv \begin{pmatrix} \tilde{Y}^K_1 \\ \vdots \\ \tilde{Y}^K_N \end{pmatrix}, \quad Y^K_c \equiv \begin{pmatrix} Y^K_{c1} \\ \vdots \\ Y^K_{cN} \end{pmatrix}, \quad Y^K_d \equiv \begin{pmatrix} Y^K_{d1} \\ \vdots \\ Y^K_{dN} \end{pmatrix}, \quad \tilde{X} \equiv \begin{pmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_N \end{pmatrix}, \text{ and } X \equiv \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$$

and partition Σ (and Σ^K) into

$$\Sigma = \begin{bmatrix} & & \vdots & \sigma_{15} \\ & \tilde{\Sigma} & \vdots & \vdots \\ \cdots & \cdots & \cdots & \sigma_{45} \\ \sigma_{51} & \cdots & \sigma_{54} & \sigma_{55} \end{bmatrix}.$$

Then, maximization of (54) with respect to $(c, d)'$ when (44_j) is omitted yields

$$\begin{pmatrix} \hat{c}^{K+1} \\ \hat{d}^{K+1} \end{pmatrix} = \left[\tilde{X}' \left(\tilde{\Sigma}^{K+1} \right)^{-1} \tilde{X} \right]^{-1} \tilde{X}' \left(\tilde{\Sigma}^{K+1} \right)^{-1} \tilde{Y}^K. \tag{58}$$

Moreover, because of the block-diagonal structure of $\tilde{X}$ and $\tilde{\Sigma}$ (recall that $\sigma_{12} = \sigma_{14} = \sigma_{23} = \sigma_{34} = 0$), (58) reduces to

$$\hat{c}^{K+1} = (X' X)^{-1} X' Y^K_c \tag{59a}$$

$$\hat{d}^{K+1} = (X' X)^{-1} X' Y^K_d. \tag{59b}$$

Thus, the M-step (59) reduces to a pair of ordinary least squares regressions of Y_{ci}^K and Y_{di}^K on X_i. Given $\hat{c}^{K+1}$ and $\hat{d}^{K+1}$ the elements of β^{K+1} can be obtained by making use of (43a, b), etc. Note that we obtain two separate estimates of $\hat{\theta}_1^{K+1}$ and similarly for θ_2^{K+1}--see (43c, d).

To summarize, at the K + 1st iteration of the EM algorithm given estimates $\hat{\beta}^K$ and $\hat{\Sigma}^K$ we first fill in the missing values of $\tilde{Y}_{1i}^K, \ldots,$ $\tilde{Y}_{5i}^K$, and τ_i^K using the formulas in (48) through (53). Then we compute $\hat{\Sigma}^{K+1}$ from (55), and we run the ordinary least squares regressions of Y_{ci}^K and Y_{di}^K on X_i to obtain the coefficient estimates $\hat{c}^{K+1}$ and $\hat{d}^{K+1}$, from which we obtain β^{K+1} via (43a-d). Upon convergence, we obtain estimates of Σ and β which correspond to the MLE. However, this procedure does not yield a variance-covariance matrix for the coefficent estimates. This can be obtained by direct evaluation of the information matrix based on the Hessian of the log-likelihood function (36) evaluated at the final EM coefficients.

IV. Application

The switching regression model (35) was applied to data on 45 cotton growers in the San Joaquin Valley of California in 1974; for a detailed description of the data set, see Hall (   ) and Farnsworth (1980). Of the growers, 28 employed IPM and 17 employed CPM. The variables and their units of measurement are as follows:

y = output of cotton lint (pounds per acre)

x = pesticide input (dollars per acre)

p = output price (dollars per pound)

$w \equiv 1$ = pesticide price

z_1 = labor input (dollars per acre)

z_2 = machinery input (dollars per acre)

z_3 = irrigation input (acre feet per acre)

z_4 = fertilizer input (dollars per acre)

s_1 = education

s_2 = years of farming experience by growers

A = acres managed by growers

$F_1 - F_2$ = fixed cost of IPM consultant (dollars)

The coefficients of the model were estimated using the EM procedure described above; the resulting estimates are shown on the following two tables. Standard errors of the coefficients (from which t statistics were calculated) were obtained by evaluating the Hessian of the log-likelihood function.

TABLE

Coefficients Estimated by EM Algorithm

Coefficient (explanatory variable)	Coefficient estimate	t statistic
$c_0 \equiv [\ln \alpha_1 + \ln \theta_1] (1 - \alpha_1)^{-1}$	- 4.7489	(-1.42)
$c_1 \equiv \frac{\alpha_1 \ln \alpha_1}{1 - \alpha_1} + \frac{\ln \theta_1}{1 - \alpha_1}$	-11.4121	(-1.88)
$\delta_{11} (1 - \alpha_1)^{-1}$ (education)	- 0.4372	(1.44)
$\delta_{12} (1 - \alpha_1)^{-1}$ (experience)	0.8744	(2.17)
$\gamma_{11} (1 - \alpha_1)^{-1}$ (labor)	1.2198	(1.90)
$\gamma_{12} (1 - \alpha_1)^{-1}$ (machinery)	0.2521	(10.39)
$\gamma_{13} (1 - \alpha_1)^{-1}$ (irrigation)	4.0702	(10.18)
$\gamma_{14} (1 - \alpha_1)^{-1}$ (fertilizer)	0.6862	(1.37)
$(1 - \alpha_1)^{-1}$ (pesticides)	4.1784	(5.51)
$d_0 \equiv [\ln \alpha_2 + \ln \theta_2] (1 - \alpha_2)^{-1}$	- 9.4944	(-0.87)
$d_1 \equiv \frac{\alpha_2 \ln \alpha_2}{1 - \alpha_2} + \frac{\ln \theta_2}{1 - \alpha_2}$	-14.8284	(-2.03)
$\delta_{21} (1 - \alpha_2)^{-1}$ (education)	- 0.4372	(-1.44)
$\delta_{22} (1 - \alpha_2)^{-1}$ (experience)	0.8744	(2.17)

(Continued on next page.)

TABLE 1--continued.

Coefficient (explanatory variable)	Coefficient estimate	t statistic
$\gamma_{21}\ (1 - \alpha_2)^{-1}$ (labor)	2.6795	(1.00)
$\gamma_{22}\ (1 - \alpha_2)^{-1}$ (machinery)	2.8122	(0.14)
$\gamma_{23}\ (1 - \alpha_2)^{-1}$ (irrigation)	- 0.2045	(-0.35)
$\gamma_{24}\ (1 - \alpha_2)^{-1}$ (fertilizer)	- 0.7962	(-0.87)
$(1 - \alpha_2)^{-1}$ (pesticides)	3.4667	(3.70)

TABLE

Transformed Coefficients

Coefficient (explanatory variable)	Coefficient estimate
θ_1	0.4219 / 0.0802
δ_{11} (education)	-0.1046
δ_{12} (experience)	0.2093
γ_{11} (labor)	0.2919
γ_{12} (machinery)	0.0603
γ_{13} (irrigation)	0.9741
γ_{14} (fertilizer)	0.1642
α_1 (pesticides)	0.7607
θ_2	0.0909 / 0.0177
δ_{21} (education)	-0.1261
δ_{22} (experience)	0.2522
γ_{21} (labor)	0.7729
γ_{22} (machinery)	0.8112
γ_{23} (irrigation)	-0.059
γ_{24} (fertilizer)	-0.2297

(Continued on next page.)

(Table --continued.)

Coefficient (explanatory variable)	Coefficient estimate
α_2 (pesticides)	0.7115
σ_{11}	30.8878
σ_{13}	- 24.6836
σ_{15}	284,314.
σ_{33}	135.25
σ_{35}	25,771.9
σ_{22}	16.4918
σ_{24}	3.5416
σ_{25}	-308,100.
σ_{44}	187.819
σ_{45}	-1.818×10^{6}
σ_{55}	1.3254×10^{11}