ABSTRACT

Title of Dissertation: ON THE EVALUATION OF CONSERVATION COST-SHARING PROGRAMS. AN APPLICATION OF A MONTE CARLO EM ALGORITHM

Ricardo Smith-Ramírez, Doctor of Philosophy, 2005

Dissertation directed by: Professor Erik Lichtenberg
Department of Agricultural and Resource Economics

Abstract.

The 2002 Farm Bill has placed a great emphasis on programs cost-sharing the adoption of conservation practices on working land. Empirical evaluation of these programs, however, has received little attention. Reasons explaining this gap in the literature may be the voluntary nature of participation and the multi-objective quality of these programs, which complicate econometric analyses. An adequate assessment of cost-sharing programs requires the modeling of multivariate responses that frequently involve limited-dependent variables and other types of unobserved information. In this study I formulate an algorithm that solves such a problem and then I use this tool to evaluate two cost-sharing programs.

I begin this Dissertation by formulating a Monte-Carlo Expectation-Maximization (MCEM) algorithm that solves a variety of models involving unobserved information in systems of linear-in-parameter equations.

Subsequently, I use the MCEM algorithm to solve a multiple adoption model and evaluate the extent at which cost-sharing payments have influenced cropping operations in Maryland. I find that soil conservation practices expand cropping both at the extensive and the intensive margin. I also prove that farmers implement practices that provide on-farm benefits preferentially. Finally, I show that cost sharing has a perverse effect: since farmers prefer to implement practices that provide private benefits, the expansion in cropping induced by cost-sharing those practices reduces the extent at which practices that provide public goods are used.

In a second empirical analysis, I analyze policy implications of ignoring nutrient dynamics in the targeting strategy of programs cost-sharing soil fertility recovery. The analysis focuses on the phosphorus fixation problem in soils derived from volcanic ash. Using an optimal control framework, I conclude that, conditional on individual characteristics, the optimal fertilization path leads either to follow a low-yield fertilization strategy or to maintain a high-yield phosphorus level in the soil. Empirical estimations on Chilean data show that program impact differs among the two regimes and that program efficiency can be improved by targeting preferentially those farms financially or technologically constrained on the low-yield fertilization path instead of allocating the funds conditional on whether or not phosphorus stock is below or above an exogenously determined target level.

ON THE EVALUATION OF CONSERVATION COST-SHARING PROGRAMS
AN APPLICATION OF A MONTE CARLO EM ALGORITHM

by

RICARDO SMITH-RAMIREZ

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY
[2005]

Advisory Committee:

Professor Erik Lichtenberg, Chair
Professor Marc Nerlove
Professor Richard Just

Dedication

To:

Leoncio and Sofia, Juan and Celmira, Ernesto, and Juan and Teresa,

because every one of my little steps is also theirs.

Acknowledgements

This work benefited greatly from the assistance provided by the members of my Committee. I am deeply indebted to all of them.

Professor Erik Lichtenberg´s permanent encouragement, accessibility and advice were fundamental to progress and finish this Dissertation.

I am particularly obliged to Professor Marc Nerlove. His constant assistance, his guidance through the methodological difficulties I encountered, and his almost telepathic ability to provide me with the exact material at the right moment made my work a lot easier.

Professor Richard Just provided numerous and useful comments that helped to improve the manuscript, particularly Chapter 3.

Dr. Wolfgang Jank from Robert H. Smith School of Business kindly assisted me with some tough issues during the formulation of the Monte Carlo EM algorithm that I use here.

I also want to thank to the Department of Agricultural and Resource Economics of the University of Maryland at College Park for its financial support, which made my doctoral studies possible.

Finally, I must thank to my wife, Marcela. Her strength, everlasting support, and sweet companionship made our adventure both easier and enjoyable.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This dissertation deals with the evaluation of agri-environmental programs and with some econometric methodology to carry out that evaluation. Agri-environmental programs (also called green programs) aim to reduce nonpoint source pollution generated by farming activities, protect wildlife habitat, improve water quality, and sustain farm productivity. Some program goals may also include curbing production of surplus commodity and providing income support for farmers. Agri-environmental programs can be broadly classified in two groups: land retirement programs, and working-land conservation programs. Well-known examples include the Conservation Reserve Program (CRP) and the Environmental Quality Incentives Program (EQIP), respectively. Land retirement programs work by taking land out from production. As compensation they offer long-term contracts consisting in annual rental payments and cost-share assistance to farmers to establish permanent vegetative covers (e.g. grass, trees) on land enrolled. Programs targeting working land attempt to improve environment management on land under production. They cost-share adoption of structural practices that require high initial investment, or provide incentive payments for the adoption of managerial conservation practices whose profitability may be perceived as uncertain by the farmer. These so-called best management practices (BMPs) vary widely in nature and they range from planting riparian buffers and cover crops to the construction of manure and waste storage structures and the use of integrated pest management.

Much has been written about the design and performance of land-retirement conservation programs, particularly after the first signup of the Conservation Reserve Program in the last eighties (Reichelderfer and Bogges, 1998; Ribaudo, 1989; Babcock et al., 1997; Feather et al., 1999; Wu, 2000; Vukina, 2004). Same effort, however, has not been undertaken for working-land conservation programs. Almost all of the existing studies about these programs focus their attention on farmer's minimum willingness to accept in order to induce participation (Lohr and Park, 1995; Cooper and Keim, 1996; Cooper, 1997). More specifically, they use contingent valuation techniques to estimate the minimum payment (and its determinants) necessary to induce a farmer to adopt a given technology or a bundle of them. Although these studies are useful for program design, they do not shed light about how well programs have been doing so far, and consequently, they do not provide much information about how to improve program effectiveness.

In the present study I attempt to address this gap in the literature. I do this by proceeding in two steps. First, I begin by developing an algorithm to estimate systems of structural equations with latent variables. The need for such an econometric tool is due to the fact that the farmer conservation decisions are integral to the production system of the farm and, consequently, can hardly be considered independent of production decisions. Thus, the analysis of farmer choices about whether to apply to a conservation program, what practices to implement and how much land these practices should cover conditional on receiving funding, how much land to crop conditional on the new technology implemented, and how much land must be allocated to grassland and permanent vegetative cover must be considered in a joint framework.

Additionally, data available for these types of studies is frequently rich in dichotomous endogenous variables (e.g. only dichotomous indicators of participation or technology adoption may be observed), which make necessary the introduction of latent variables in the model. The researcher may also decide to use latent variables in order to add flexibility to the model or to represent some underlying unobserved process. As commercial software capable of solving general forms of these types of models is not currently available, I have found it necessary to begin this study by formulating a suitable algorithm by my own.

Then I use the algorithm introduced above and two data sets, one from Maryland and other from Chile, to analyze different issues related to agri-environmental programs.

The remaining of this dissertation is organized in three chapters. In Chapter 2, I implement a flexible and robust algorithm for estimating systems of linear-in-parameters equations involving unobserved information by maximum likelihood. The procedure uses a Monte Carlo Expectation Maximization (MCEM) algorithm that exploits the underlying latent continuum to obtain efficient estimators for a wide variety of models involving systems of structural equations with latent variables. I begin the chapter by discussing the issues that make the estimation of these models by traditional methods too cumbersome or even unfeasible. Then I formulate an algorithm for systems of equations where the latent variables enter only as dependent variables, i.e. on the left-hand side of the equations. The performance of the MCEM algorithm is compared against a numerical integration approach by solving a 3-equation system where the observed counterparts of the dependent variables are either dichotomous or limited-dependent. Finally, I generalize the algorithm to estimate systems of structural equations, i.e. models in which the latent variables appear not only on the left-hand side of the equations but also as endogenous regressors.

In Chapter 3 I use data from Maryland farmers to study the interaction between conservation and production decisions and how this interdependence may generate unintended effects of the program. I show that, by subsidizing the adoption of land-quality augmenting technology, agri-environmental programs may induce changes in land profitability and affect cropping patterns of participant farms. I find that adoption of soil conservation practices expands cropping operations at both the intensive and the extensive margins. Additionally, programs that cost-share the adoption of these practices may have the perverse effect of reducing the share of land under practices that provide public goods. These findings raise questions as to the appropriateness of treating conservation cost-sharing payments as "green-box" under WTO's regulations.

In Chapter 4 I use data from Chile to analyze the targeting policy of a cost sharing program aiming to replenish soil phosphorus fertility. I begin the analysis by solving an optimal-control problem to determine the farmer's optimal soil fertilization strategy. I show that two optimal paths exist. Depending on farmer characteristics and initial level of phosphorus in the soil, the farmer may choose either to follow a subsistence fertilization strategy (low yield path) or remediate soil fertility until a higher level of phosphorus is achieved (high yield path). In order to evaluate the impact of the program on each of the two possible regimes, I formulate an endogenous-switching regression framework with unobserved switching points. A switching equation determines the probability of being in each regime conditional on farm characteristics; then specific fertilization-path equations are used to determine the effect of cost sharing on each regime. Estimation results indicate that program effects are greater on financially and/or technologically constrained farms being on the low yield path, while cost sharing might have only short run effects or be equivalent to a pure transfer on farms in the high yield path.

Chapter 2: A Monte Carlo EM Algorithm to Estimate Structural Equation Systems with Unobserved Information

1 Introduction

Models involving equation systems with latent structures are abundant in applied economics. Examples include sample selectivity models, switching regression models, multivariate and nested tobit models, multivariate and multinomial probit models, and panel data models with random effects. The use of latent variables gives some level of independence from the limitations of the observed data to the applied econometrician. For instance, completely or partially unobserved variables can be added to the empirical model in order to get a better representation of the phenomenon under study. However, the use of latent variables does not come without costs. An important issue is that both the number and the dimensionality of the integral terms in the likelihood function increase with the number of latent variables considered. High dimensional integration slows the estimation down and it can even make the estimation unfeasible in presence of integrals of dimension greater than three. A second issue arises from the inability of many conventional optimization algorithms to identify the parameters of these models even though conditions for formal identification are satisfied. "Fragile" identification, as it called by Keane (1992), tends to happen when the objective function shows little variation in a wide range of parameter values around the maximum, which prevents convergence of gradient-based algorithms. Finally, the selection of starting values in order to initiate the optimization routine is a frequent problem in maximum likelihood estimation of models with latent structures. Consequently, the development of algorithms with low sensitivity to the selection of starting values (i.e. with a larger approximation area) is always welcome.

The traditional approaches for estimating 2-equation systems that involve latent variables have been maximum likelihood and 2-step estimation methods (Heckman, 1979; Maddala, 1983). The first one is more desirable because it produces consistent and efficient estimates; however, it is prone to “fragile” identification and starting value problems. The second one is robust, but it is not efficient. For equation systems involving three or more latent variables, maximum likelihood estimation by numerical integration is often too costly computationally or even unfeasible since quadrature methods for high dimensional integrals are still in development. This so-called “curse of dimensionality” has, however, been partially overcome in the last years by the use of probability simulators (Börsch-Supan and Hajivassiliou, 1993; Geweke et al., 1994) and Monte Carlo and Quasi-Monte Carlo integration methods (Sobol, 1998). Still, the focus of these approaches is only on making the integration of the likelihood function feasible, which implies that the problems of “fragile” identification and starting values remain.

Instead of placing the attention on calculating the integrals in the likelihood function, the approach presented in this study focuses on the latent continuum generating the observed information. By combining a Monte Carlo Expectation-Maximization

(MCEM) algorithm with a sequential conditional maximization procedure, I show that the estimation of any system of linear (in parameters) equations with latent variables can be seen as equivalent to estimating a system of linear (in parameters) equations with fully observed information recursively. The use of a MCEM circumvents the integration problem by imputing the unobserved information using Gibbs sampling (Casella and George, 1992). Since the use of the Gibbs sampler in the Expectation step permits "restoring" the continuum, the Maximization step does not differ much from maximizing the likelihood function of a standard linear equation system. Complementarily, the use of sequential maximization steps permits concentration of the optimization effort on those parameters that are frequently the hardest to identify, i.e. the elements of the disturbance covariance matrix. Finally, the MCEM framework confines the estimates to the parameter space at every iteration of the algorithm and reduces dependency on starting values. This study generalizes the procedure developed by Natarajan et al. (2000) for multinomial probit models. Its main contribution is the implementation of a robust algorithm that exploits explicitly the structural similarity between models that have been traditionally estimated by rather ad-hoc methods.

The remaining of this chapter is organized in the following way. The next section discusses the meaning of unobserved information in the context of this study. The third section presents the MCEM algorithm and exemplifies how it works by estimating a 3-equation problem. The fourth section solves the same problem as in the previous section by numerical integration. The outputs of both approaches are compared. The fifth section generalizes the algorithm to cover system of structural equations with latent variables. The sixth and last section gives final remarks.

2 Unobserved information

For purposes of this study I consider two ways that unobserved information might enter in econometric estimations: by the existence of missing data, and by the presence of latent variables. Missing data are observations that the researcher failed to collect for all the individuals in the sample. Missing information can originate by multiple ways such as inability to sample the same unit along different years when constructing a panel data set, or unwillingness of the respondent to answer specific questions in a survey.

By latent variable I mean a continuous variable that is not observed fully; nonetheless, part of the information contained in the variable is available to the econometrician. This observed counterpart originates a new variable, whose type (e.g. dichotomous, polytomous, limited-dependent) will depend on how it relates to the underlying latent variable.

It must be kept in mind that those variables containing missing data, although less structured, can also be seen and understood as another kind of latent variable. As a consequence, the methods that I am about to present in order to estimate models involving latent variables can be used to deal with missing information as well.

3 The Monte Carlo Expectation- Maximization (MCEM) algorithm

In their presentation before the Royal Statistical Society, Dempster et al. (1977) introduced the Expectation-Maximization (EM) algorithm as an iterative procedure to compute maximum likelihood estimates "... when the observations can be viewed as incomplete data."

The way the notion of "incomplete data" is introduced above is indeed very general and it is this flexibility in the idea of incomplete data what is responsible of a good deal of EM algorithm's broad applicability. To give a flavor of how the algorithm works consider the following many-to-one mapping

$z \in Z \to y = y(z) \in Y$

The information z in Z is not observed directly but through its observed realization y in Y. In words, z is only known to lie in $Z(y)$, the subset of Z determined by the equation $y = y(z)$, where y is the observed (measurable) data.

Let the complete data be written as $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z})$, where $\boldsymbol{z}$ is the unobserved information. Then the log-likelihood function of the observed information can be written as

$$\ell(\theta \mid \boldsymbol{y}) = \ln L(\theta \mid \boldsymbol{y}) = \ln \int_{Z(y)} L(\theta \mid \boldsymbol{x}) d\boldsymbol{z} \qquad (2.1)$$

As previously discussed, the integrals present in (2.1) can make the maximization of $\ell(\theta \mid \boldsymbol{y})$ cumbersome or even impossible to solve by standard optimization methods. Instead of trying to solve (2.1) directly, the EM algorithm focuses on the complete-information log-likelihood $\ell^c(\theta \mid \boldsymbol{x})$ and maximizes $E\left[\ell^c(\theta \mid \boldsymbol{x})\right]$ by executing iteratively two steps. The first one is the so-called Expectation step or E-step, which at iteration $m+1$ computes $Q\left(\theta \mid \theta^{(m)}, \boldsymbol{y}\right) = E\left[\ell^c(\theta \mid \boldsymbol{x})\right]$, where $E\left[\ell^c(\theta \mid \boldsymbol{x})\right]$ is the expectation of the complete-information log-likelihood conditional on the observed information and provided that the conditional density $f\left(\boldsymbol{x} \mid \boldsymbol{y}, \theta^{(m)}\right)$ is known. The E-step is followed by the Maximization step or M-step, which maximizes $Q\left(\theta \mid \theta^{(m)}, \boldsymbol{y}\right)$ to find $\theta^{(m+1)}$. Then the procedure is repeated until convergence is attained. Often, however, this deterministic version of the EM algorithm has also to deal with hefty integrals in the calculation of the expectations in the E-step.

The stochastic version of the EM algorithm presented here avoids troublesome computations in the E-step by imputing the unobserved information conditional on what is observed and on distribution assumptions. In this approach the term $Q\left(\theta \mid \theta^{(m)}, \boldsymbol{y}\right)$ is approximated by the mean $\frac{1}{K}\sum_{k=1}^{K} Q\left(\theta, \boldsymbol{z}^{(k)} \mid \boldsymbol{y}\right)$, where the $\boldsymbol{z}^{(k)}$ are random samples from $f\left(\boldsymbol{x} \mid \theta^{(m)}, \boldsymbol{y}\right)$ (Wei and Tanner, 1990). No integrals need to be estimated in this procedure. Once the unobserved information is imputed, the latent continuum is made “visible” and the estimation can be carried out as we were solving a standard system of linear equations.

3.1 Implementing the Monte Carlo EM algorithm.

I illustrate the use of the MCEM algorithm by solving the following 3-equation system

$$\begin{aligned} y_{1i}^* &= x_{1i}\beta_1 + \varepsilon_{1i} \\ y_{2i}^* &= \gamma_2 y_{1i} + x_{2i}\beta_2 + \varepsilon_{2i} \\ y_{3i}^* &= \gamma_3 y_{1i} + x_{3i}\beta_3 + \varepsilon_{3i} \end{aligned} \qquad (2.2)$$

where y_{1i} is dichotomous, and y_{2i} and y_{3i} are censored from below at zero, i.e.

$$y_{1i} = \begin{cases} 1 \text{ if } y_{1i}^* > 0 \\ 0 \text{ if } y_{1i}^* \leq 0 \end{cases} \qquad y_{2i} = \begin{cases} y_{2i}^* \text{ if } y_{2i}^* > 0 \\ 0 \text{ if } y_{2i}^* \leq 0 \end{cases} \qquad y_{3i} = \begin{cases} y_{3i}^* \text{ if } y_{3i}^* > 0 \\ 0 \text{ if } y_{3i}^* \leq 0 \end{cases}$$

Equation system (2.2) contains only the observed counterparts of the latent variables on the right-hand side of the equations. It is clear that more general cases should consider both latent and observed endogenous regressors. The discussion of those cases will be delayed until Section 5. In the meantime the use of simpler models like (2.2) is more suitable to introduce the Monte Carlo EM algorithm. This will allow us to concentrate on methodological aspects and not get distracted by complications in the model structure.

The disturbance terms in (2.2) are assumed to have a trivariate normal distribution $N(0, \Sigma)$ with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_1\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma^2_{\varepsilon_2} & \sigma_{\varepsilon_2\varepsilon_3} \\ \sigma_{\varepsilon_1\varepsilon_3} & \sigma_{\varepsilon_2\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix} = \begin{pmatrix} 1 & \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_2} & \rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3} \\ \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_2} & \sigma^2_{\varepsilon_2} & \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3} \\ \rho_{\varepsilon_1\varepsilon_3}\sigma_{\varepsilon_3} & \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3} & \sigma^2_{\varepsilon_3} \end{pmatrix} \quad (2.3)$$

where $\sigma^2_{\varepsilon_1} = 1$ is the usual normalization to ensure identification of the coefficients in an equation with a dichotomous dependent variable and $\rho_{\varepsilon_k\varepsilon_l}$ is the correlation coefficient between ε_k and ε_l $\left(k,l = 1,2,3\right)$.

There are two forms of the structural model for the system in (2.2) depending on the value of the observed counterpart of y^*_{1i}, i.e.

$$\begin{array}{ll} y_{1i} = 1 & y_{1i} = 0 \\ \\ y^*_{1i} = x_{1i}\beta_1 + \varepsilon_{1i} & y^*_{1i} = x_{1i}\beta_1 + \varepsilon_{1i} \\ y^*_{2i} = \gamma_2 + x_{2i}\beta_2 + \varepsilon_{2i} & y^*_{2i} = x_{2i}\beta_2 + \varepsilon_{2i} \\ y^*_{3i} = \gamma_3 + x_{3i}\beta_3 + \varepsilon_{3i} & y^*_{3i} = x_{3i}\beta_3 + \varepsilon_{3i} \end{array} \quad (2.4)$$

According to (2.4) the parameters γ_2 and γ_3 only represent shifts in the intercepts of the second and third equations when $y_{1i} = 1$. Thus, under the normality assumption, the complete data likelihood function can be written as

$$L(\theta, \Sigma \mid \boldsymbol{x}) = \prod_i f\left(y_{1i}, y_{2i}, y_{3i}\right) = \prod_i \left[\frac{1}{\left(2\pi\right)^{3/2} \left|\Sigma\right|^{1/2}} \exp\left(-\frac{\boldsymbol{\varepsilon}_i' \Sigma^{-1} \boldsymbol{\varepsilon}_i}{2} \right) \right]$$

Where $\boldsymbol{\theta} = \left(\beta_1, \gamma_2, \beta_2, \gamma_3, \beta_3\right)$, $\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix} = \begin{pmatrix} y_{1i}^* - X_{1i}\beta_1 \\ y_{2i}^* - \gamma_2 y_{1i} - X_{2i}\beta_2 \\ y_{3i}^* - \gamma_3 y_{1i} - X_{3i}\beta_3 \end{pmatrix}$

Correspondingly, the complete information log-likelihood function and its expectation are

$$\ell^c\left(\boldsymbol{\theta}, \Sigma \mid \boldsymbol{x}\right) = -\frac{3N}{2}\ln\left(2\pi\right) - \frac{N}{2}\ln\left|\Sigma\right| - \frac{1}{2}\sum_i \operatorname{tr}\left(\Sigma^{-1}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right)$$

$$E\left[\ell^c\left(\boldsymbol{\theta}, \Sigma \mid \boldsymbol{x}\right)\right] = -\frac{3N}{2}\ln\left(2\pi\right) - \frac{N}{2}\ln\left|\Sigma\right| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_i E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right]\right) \quad (2.5)$$

Where N is the total number of observations and the expectation operator indicates expectation conditional on observed information and distributional assumptions. The E-step is straightforward from equation (2.5) and, at iteration $m+1$, requires the calculation of

$$Q_i\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right) = E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right] = E\left[\begin{pmatrix} y_{1i}^* - X_{1i}\beta_1 \\ y_{2i}^* - \gamma_2 y_{1i} - X_{2j}\beta_2 \\ y_{3i}^* - \gamma_3 y_{1i} - X_{3j}\beta_3 \end{pmatrix}\begin{pmatrix} y_{1i}^* - X_{1i}\beta_1 \\ y_{2i}^* - \gamma_2 y_{1i} - X_{2j}\beta_2 \\ y_{3i}^* - \gamma_3 y_{1i} - X_{3j}\beta_3 \end{pmatrix}^{'} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right] \quad (2.6)$$

$$= \sigma_i^{2(m)} + \begin{pmatrix} \mu_{y_{1i}^*}^{(m)} - X_{1i}\beta_1 \\ \mu_{y_{2i}^*}^{(m)} - \gamma_2 y_{1i} - X_{2i}\beta_2 \\ \mu_{y_{3i}^*}^{(m)} - \gamma_3 y_{1i} - X_{3i}\beta_3 \end{pmatrix}\begin{pmatrix} \mu_{y_{1i}^*}^{(m)} - X_{1i}\beta_1 \\ \mu_{y_{2i}^*}^{(m)} - \gamma_2 y_{1i} - X_{2i}\beta_2 \\ \mu_{y_{3i}^*}^{(m)} - \gamma_3 y_{1i} - X_{3i}\beta_3 \end{pmatrix}^{'}$$

where $\sigma_i^{2(m)} = \operatorname{Cov}\left(y_{1i}^*, y_{2i}^*, y_{3i}^* \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right) = \begin{pmatrix} \sigma_{y_{1i}^*}^{2(m)} & \sigma_{y_{1i}^* y_{2i}^*}^{(m)} & \sigma_{y_{1i}^* y_{3i}^*}^{(m)} \\ \sigma_{y_{1i}^* y_{2i}^*}^{(m)} & \sigma_{y_{2i}^*}^{2(m)} & \sigma_{y_{2i}^* y_{3i}^*}^{(m)} \\ \sigma_{y_{1i}^* y_{3i}^*}^{(m)} & \sigma_{y_{2i}^* y_{3i}^*}^{(m)} & \sigma_{y_{3i}^*}^{2(m)} \end{pmatrix}$ (2.7)

$$\text{and} \begin{pmatrix} \mu^{(m)}_{y^*_{1i}} \\ \mu^{(m)}_{y^*_{2i}} \\ \mu^{(m)}_{y^*_{3i}} \end{pmatrix} = \begin{pmatrix} E\left[y^*_{1i} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y} \right] \\ E\left[y^*_{21i} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y} \right] \\ E\left[y^*_{3i} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y} \right] \end{pmatrix} \tag{2.8}$$

The covariance matrix $\sigma_i^{2(m)}$ in (2.7) and the vector of means in (2.8) can be estimated by Gibbs sampling (Casella and George, 1992) from the joint distribution of $\left(y^*_{1i}, y^*_{2i}, y^*_{3i}\right)$ conditional on parameters $\left(\theta^{(m)}, \Sigma^{(m)}\right)$ and the observed information $\boldsymbol{y}$. It is useful determining first the distribution of $\left(y^*_{1i}, y^*_{2i}, y^*_{3i}\right)$. After recalling that γ_2 and γ_3 are only structural shifts in the second and third equations of (2.4) and given the distribution of the disturbances in (2.3), the distribution of $\left(y^*_{1i}, y^*_{2i}, y^*_{3i}\right)$ at iteration m is $N\left(\mu_i^{(m)}, \Sigma^{(m)}\right)$, where

$$\mu_i^{(m)} = \begin{pmatrix} X_{1i}\beta_1^{(m)} \\ \gamma_2^{(m)} y_{1i} + X_{2i}\beta_2^{(m)} \\ \gamma_3^{(m)} y_{1i} + X_{3i}\beta_3^{(m)} \end{pmatrix} \quad \text{and} \quad \Sigma^{(m)} = \begin{pmatrix} 1 & \sigma^{(m)}_{\varepsilon_1\varepsilon_2} & \sigma^{(m)}_{\varepsilon_1\varepsilon_3} \\ \sigma^{(m)}_{\varepsilon_1\varepsilon_2} & \sigma^{(m)}_{\varepsilon_2} & \sigma^{(m)}_{\varepsilon_2\varepsilon_3} \\ \sigma^{(m)}_{\varepsilon_1\varepsilon_3} & \sigma^{(m)}_{\varepsilon_2\varepsilon_3} & \sigma^{(m)}_{\varepsilon_3} \end{pmatrix} \tag{2.9}$$

3.2 The Gibbs sampler

The moments in (2.7) and (2.8) could be easily calculated if the marginal densities (conditional on parameters and observed information) of y^*_{1i}, y^*_{2i}, and y^*_{3i} were known. However, obtaining those marginal densities may require solving high dimensional integrals. Instead of tackling the problem by integration, the Gibbs sampler provides a way to generate samples from the marginal distributions without requiring analytical expressions for the densities. The moments of interest can then be estimated from the simulated samples.

The implementation of the Gibbs sampler is straightforward using the definitions in (2.9). Before proceeding, let consider the following notation

$$\boldsymbol{y}^*_{i|-j} = \begin{pmatrix} y^*_{1i} \\ \vdots \\ y^*_{j-1i} \\ y^*_{j+1i} \\ \vdots \\ y^*_{ki} \end{pmatrix} \qquad \boldsymbol{X}_{i|-j} = \begin{pmatrix} X_{1i} \\ \vdots \\ X_{j-1i} \\ X_{j+1i} \\ \vdots \\ X_{ki} \end{pmatrix} \qquad \boldsymbol{\gamma}^{(m)}_{-j} = \begin{pmatrix} \gamma^{(m)}_1 \\ \vdots \\ \gamma^{(m)}_{j-1} \\ \gamma^{(m)}_{j+1} \\ \vdots \\ \gamma^{(m)}_k \end{pmatrix} \qquad \boldsymbol{\beta}^{(m)}_{-j} = \begin{pmatrix} \beta^{(m)}_1 \\ \vdots \\ \beta^{(m)}_{j-1} \\ \beta^{(m)}_{j+1} \\ \vdots \\ \beta^{(m)}_k \end{pmatrix}$$

where $j = 1,...,k$ and k is the number of equations in the system to estimate (equal to 3 in our example).

The implementation of the sampler begins with determining the distribution of each y^*_{ji} conditional on the value of the rest of the dependent variables $\boldsymbol{y}^*_{i|-j}$. It is well known that, under the normality assumption, this conditional distribution is univariate normal. Thus, means $\mu_{ji|i(-j)}$ and variances $\sigma^2_{j|-j}$ at the $m+1$ iteration can be estimated by

$$\mu_{ji|i(-j)}^{(m)} = E\left(y_{ji}^{*} \middle| \boldsymbol{y}_{i|-j}^{*}, \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}\right)$$
$$= X_{ji}\beta_{j}^{(m)} + \operatorname{cov}\left(y_{ji}^{*} \middle| \boldsymbol{y}_{i|-j}^{*}, \Sigma^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^{*} \middle| \Sigma^{(m)}\right)\right]^{-1}\left(\boldsymbol{y}_{i|-j}^{*} - \gamma_{-j}^{(m)} - \boldsymbol{X}_{i|-j}\boldsymbol{\beta}_{-j}^{(m)}\right) \tag{2.10}$$

$$\sigma_{j|-j}^{2(m)} = \operatorname{var}\left(y_{ji}^{*} \middle| \boldsymbol{y}_{i|-j}^{*}, \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}\right)$$
$$= \operatorname{var}\left(y_{ji}^{*} \middle| \Sigma^{(m)}\right) - \operatorname{cov}\left(y_{ji}^{*} \middle| \boldsymbol{y}_{i|-j}^{*}, \Sigma^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^{*} \middle| \Sigma^{(m)}\right)\right]^{-1}\operatorname{cov}\left(y_{ji}^{*} \middle| \boldsymbol{y}_{i|-j}^{*}, \Sigma^{(m)}\right)' \tag{2.11}$$

The next step is to sample iteratively from these conditional distributions in order to simulate a sample for the unobserved values of each y_{ji}^{*}. These samples will in turn allow estimating the values in (2.7) and (2.8). Since the simulations for $\boldsymbol{y}_{i}^{*}$ must be done conditional on its corresponding observed information $\boldsymbol{y}_{i}$, the implementation procedure depends on the structure imposed by $\boldsymbol{y}_{i}$ on $\boldsymbol{y}_{i}^{*}$.

The observed counterpart of y_{1i}^{*} in the first equation in (2.2) is dichotomous with y_{1i}^{*} being positive if y_{1i} equals one and non-positive if y_{1i} equals zero. Accordingly, it is necessary to simulate y_{1i}^{*} from a normal distribution with mean $\mu_{1i|i(-1)}^{(m)}$ and variance $\sigma_{1i|-1}^{2(m)}$ truncated below at zero if y_{1i} equals one and truncated above at zero if y_{1i} equals zero.

Variables y_{2i}^{*} and y_{3i}^{*} are both observed when having positive values. Consequently, it is only necessary to simulate them when $y_{2i} = 0$ and $y_{3i} = 0$, respectively. Thus, these variables must be simulated from normal distributions with means $\mu_{ji|i(-j)}^{(m)}$ and variances $\sigma_{ji|-j}^{2(m)}$ truncated above at zero when y_{ji} $(j = 2,3)$ equals zero. When $y_{ji} > 0$ set $y_{ji}^{*} = y_{ji}$.

Sampling from a truncated normal distribution can be easily accomplished by using the inverse distribution method. As an example, assume $y^{*} \sim N(\mu, \sigma^{2})$ and y is limited to be in the interval $[l,u]$. Then according to Devroye (1986, p39), a random draw from the truncated normal distribution of y is given by

$$y = \mu + \sigma\Phi^{-1}\left(P_l + U\left(P_u - P_l\right)\right) \qquad (2.12)$$

Where $P_l = \Phi\left(\frac{l-\mu}{\sigma}\right)$, $P_u = \Phi\left(\frac{u-\mu}{\sigma}\right)$ and U is a random draw from the standard uniform distribution. MATLAB's pseudo random generator (Moler, 1995) was used to simulate the sampling from U in this study. As an alternatively to the use of pseudo random numbers in the sampling, the use of randomized low-discrepancy sequences or quasi-random numbers has been proposed in order to reduce Monte Carlo noise and speed up convergence (Liao, 1998; Jank, 2004).

A complete set of starting vectors $\boldsymbol{y}_i^{*}$ is necessary to begin the Gibbs sampler. In this study y_{ji}^{*} was set equal to zero $\forall i, j$ when the observed variable was dichotomous and equal to y_{ji} when censored. The simulation was then repeated iteratively until completing a sequence $\boldsymbol{y}_i^{*(1)},...,\boldsymbol{y}_i^{*(K^{(m)})}$, where $K^{(m)}$ is a number large enough to ensure convergence. Following Wei and Tanner (1990), it is more efficient to begin with a small $K^{(1)}$ and progressively increase $K^{(m)}$ as m increases. A simple linear rate of increment was used here. Then eliminate a number k_{burn} of simulations from the beginning of the sequence. The remaining observations in the sequence are used to estimate $\sigma_{ji}^{2(m)}$ and $\mu_{ji}^{(m)}$ in (2.7) and (2.8) according to

$$\sigma_i^{2(m)} = \operatorname{cov}\left(y_{1i}^*, y_{2i}^*, y_{3i}^* \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right) \approx \begin{pmatrix} \hat{\sigma}_{y_{1i}^*}^{2(m)} & \hat{\sigma}_{y_{1i}^* y_{2i}^*}^{(m)} & \hat{\sigma}_{y_{1i}^* y_{3i}^*}^{(m)} \\ \hat{\sigma}_{y_{1i}^* y_{2i}^*}^{(m)} & \hat{\sigma}_{y_{2i}^*}^{2(m)} & \hat{\sigma}_{y_{2i}^* y_{3i}^*}^{(m)} \\ \hat{\sigma}_{y_{1i}^* y_{3i}^*}^{(m)} & \hat{\sigma}_{y_{2i}^* y_{3i}^*}^{(m)} & \hat{\sigma}_{y_{3i}^*}^{2(m)} \end{pmatrix}$$

Where $\hat{\sigma}_{y_{ri}^* y_{si}^*}^{(m)} = \dfrac{1}{K^{(m)} - k_{burn} - 1} \sum_{k=k_{burn}+1}^{K^{(m)}} \left(y_{ri}^{*(k)} - \overline{y}_{ri}^*\right)\left(y_{si}^{*(k)} - \overline{y}_{si}^*\right)$

$$\mu_{y_{ji}^*}^{(m)} = E\left[y_{ji}^* \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right] \approx \overline{y}_{ji}^* = \frac{1}{K^{(m)} - k_{burn}} \sum_{k=k_{burn}+1}^{K^{(m)}} y_{ji}^{*(k)}$$

Notice that when y_{ji}^* is fully observed (i.e. when $y_{ji}^* = y_{ji}$) then $\hat{\sigma}_{y_{ri}^* y_{si}^*}^{(m)} = 0$ and $\mu_{y_{ji}^*}^{(m)} = y_{ij} \quad \forall\ m, r, s$.

3.3 Maximization Step

After obtaining $\sigma_{ji}^{2(m)}$ and $\mu_{ji}^{(m)}$ we are ready to move to the Maximization step. From (2.5) and (2.6) we maximize

$$E\left[\ell^{c}\left(\boldsymbol{\theta},\Sigma\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right]=-\frac{3N}{2}\ln\left(2\pi\right)-\frac{N}{2}\ln\left|\Sigma\right|-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_{i}Q_{i}\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right)\quad(2.13)$$

Note that the use of the Gibbs sampler has permitted us to circumvent the estimation of the high dimensional integrals present in (2.21). Also notice that, except for the covariance matrices $\sigma_i^{2(m)}$ present in the terms, the expression in (2.13) is the log-likelihood function of a system of linear equations, where the unobserved information has been replaced by its expected values. Thus, in a certain sense the latent continuum has been restored. Similarly to Meg and Rubin (1993) and Natarajan et al. (2000), I use two conditional maximization steps in order to maximize the expression in (2.13) with respect to $\boldsymbol{\theta}$ and the elements in Σ. The first maximization step maximizes (2.13) with respect to $\boldsymbol{\theta}$ conditional on $\Sigma^{(m)}$ to produce $\boldsymbol{\theta}^{(m+1)}$. This is followed by a maximization on the elements of Σ conditional on the recently updated $\boldsymbol{\theta}^{(m+1)}$ in order to obtain $\Sigma^{(m+1)}$.

It is clear from (2.6) that the maximizer in the first conditional maximization is the generalized least square estimator

$$\boldsymbol{\theta}^{(m+1)}=\left[\tilde{X}_d'\left(\Sigma^{(m)}\otimes I_N\right)^{-1}\tilde{X}_d\right]^{-1}\tilde{X}_d'\left(\Sigma^{(m)}\otimes I_N\right)^{-1}\mu_{y^*}^{(m)}\quad(2.14)$$

where

$\tilde{X}_d = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & \tilde{X}_2 & 0 \\ 0 & 0 & \tilde{X}_3 \end{bmatrix}$, $\tilde{X}_2$ and $\tilde{X}_3$ are the matrices $\left[y_{1i} \vdots X_2\right]$ and $\left[y_{1i} \vdots X_3\right]$ respectively, I_N is the identity matrix of dimension N and $\mu_{y^*}^{(m)}$ is a column vector of dimension Nk constructed by stacking the elements $\mu_{y_{ji}^*}^{(m)}$ from (2.8) in the following way

$$\mu_{y^*}^{(m)} = \left(\mu_{y_{11}^*}^{(m)}, \mu_{y_{12}^*}^{(m)}, \ldots, \mu_{y_{1N}^*}^{(m)}, \mu_{y_{21}^*}^{(m)}, \ldots, \mu_{y_{2N}^*}^{(m)}, \mu_{y_{31}^*}^{(m)}, \ldots, \mu_{y_{3N}^*}^{(m)}\right)' \quad (2.15)$$

After plugging (2.14) into (2.13), $\Sigma^{(m+1)}$ is obtained by maximizing

$$E\left[\ell^c\left(\Sigma \mid \theta^{(m+1)}, \Sigma^{(m)}, \boldsymbol{y}\right)\right] = -\frac{3N}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_i Q_i\left(\theta^{(m+1)} \mid \theta^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right)\right) \quad (2.16)$$

with respect to the $3(3-1)/2+2=5$ different elements in Σ. The maximization of (2.16) can be accomplished with the routine FMINUNC in Matlab. Unlike the log-likelihood function in (2.21), the function in (2.16) is simple enough to obtain an analytical expression for its gradient. This is useful since no time need to be spent in a numerical estimation of the gradient by the optimization routine.

3.4 Convergence issues and stopping rules

Literature discussing convergence of the MCEM is scarce and it suggests that MCEM convergence relies mainly on properties of the deterministic EM algorithm and the Gibbs sampler. Convergence of the EM algorithm has been discussed by Dempster et al. (1977), Boyles (1983) and Wu (1983) and convergence properties of the Gibbs sampler are studied in Geman and Geman (1984) and Casella and George (1992). In one of these studies, Wu (1983) clarifies a common misconception about the superior properties of the EM algorithm in converging to a global maximum. He shows that, like other maximization methods, the EM algorithm converges monotonically to some stationary point of a bounded log-likelihood function; however there is no guarantee that point is the global maximum. Consequently, the EM algorithm is susceptible to starting value problems and may converge to a local maximum, a saddle point, or even may not converge to a unique optimum, getting trapped in a connected set of local maxima instead (e.g. a plateau in the objective function).

In the same study cited above, Wu shows that if the log-likelihood function is well behaved and it has only one stationary point (a maximum) then the EM sequence will converge to the unique maximizer. This property has two immediate implications for unimodal and differentiable log-likelihood functions. First, the EM algorithm has a greater approximation area to the global maximum or, in other words, its estimates are less sensitive to starting values than other optimization techniques. Second and directly related to the first implication, the EM estimates are confined to lie in the parameter space at each iteration. Consequently, problems with estimates of the disturbance covariance matrix like negative variances or correlation coefficients with absolute values greater than one (which are frequent under Newton and Quasi-Newton techniques) do not happen when using the EM algorithm.

Some closely related issues must be discussed before finishing the implementation of the MCEM algorithm. They are the criteria to use in order to determine the size of the Gibbs sample $K^{(m)}$ and to determine when convergence has been attained. Wei and Tanner (1990) have indicated that it is inefficient to begin with large Gibbs samples since MCEM estimates are likely to be far from the true maximizer during the first iterations. Rather it is more reasonable to begin with small samples and make $K^{(m)}$ an increasing function of m in order to reduce the Monte Carlo error as the algorithm approaches the maximizer. However, there is not a single criterion about the way $K^{(m)}$ must be increased at every iteration.

Some approaches consider separately the issues of determining the optimal size of the Gibbs simulation and monitoring convergence. Thus, McCulloch (1997) considers rather abrupt increments in the size of the Gibbs sample every time that m had achieved certain arbitrary values, while McCulloch (1994) uses a linear rate of increment. Convergence monitoring in these works is accomplished by plotting the expected log-likelihood versus iteration number and the algorithm is stopped manually when the process is observed to stabilize (Wei and Tanner, 1990; Natarajan et al., 2000).

More elaborate approaches consider evaluating the Monte Carlo error at iteration m and use that estimation both to determine $K^{(m+1)}$ and to evaluate convergence. These methods can be classified either as likelihood-distance-based or as parameter-distance-based depending on whether they focus on likelihood differences $\left|E\left[\ell^{c}\left(\vartheta^{(j)}\right)\right]-E\left[\ell^{c}\left(\vartheta^{(j-1)}\right)\right]\right|$ or parameter differences $\left|\vartheta^{(j)}-\vartheta^{(j-1)}\right|$, where $\vartheta^{(j)}$ is the estimation of the parameter vector at iteration j. The idea is that if parameter or likelihood differences show variation no greater than the Monte Carlo error then the estimation has been saturated by random variation and the simulation size must be increased in the next iteration. In a complementarily way, a stopping rule can be implemented by establishing a target level for the Monte Carlo error. Examples of the likelihood-distance-based approach can be found in Chan and Ledolter (1995) and Eickhoff et al. (2004). Perhaps the sounder parameter-distance-based approach belongs to Booth and Hobert (1999). They use a Taylor series approximation to construct a confidence ellipsoid around $\vartheta^{(m+1)}$, where the length of the ellipsoid axis on every dimension of the parameter space is a measure of the MC error on the respective dimension. Thus, if the estimate $\vartheta^{(m)}$ is contained in the ellipsoid, the current estimate $\vartheta^{(m+1)}$ is swamped in MC error and $K^{(m+1)}$ must be increased.

This study uses a linear rate of increment for the size of the Gibbs sample and a stopping ruled based both on likelihood and parameter distances. The idea is simply to automate the plotting method of Wei and Tanner (1990) by introducing the following criteria:

$$\sum_{j=M-J}^{M}\left|\frac{E\left[\ell^{c}\left(\vartheta^{(j)}\right)\right]-E\left[\ell^{c}\left(\vartheta^{(j-1)}\right)\right]}{E\left[\ell^{c}\left(\vartheta^{(j-1)}\right)\right]}\right|<10^{-3} \qquad (2.17)$$

$$\max_{k}\left[\sum_{j=M-J}^{M}\left|\frac{\vartheta_{k}^{(j)}-\vartheta_{k}^{(j-1)}}{\vartheta_{k}^{(j-1)}}\right|\right]<10^{-2}$$

where $\vartheta_k^{(j)}$ is the estimate of the k component of the parameter vector at iteration j, M is the current number of iterations, and J is a researcher choice. In this example J was set equal to $0.25\times M$. The algorithm was stopped only when both criteria were satisfied for at least ten consecutive iterations. This last requirement was introduced to avoid false convergence due to the tendency of the MCEM algorithm to stall temporarily before reaching the maximizer. The criteria in (2.17) are simple to implement and, somewhat, they are more stringent than any of those presented in the articles mentioned above and may increase unnecessarily the number of iterations required for convergence. However, given the speed of today's computer, the computational cost is not very high.

3.5 Estimation of the Information matrix

The asymptotic standard errors of the estimates are not among the outputs of the EM algorithm and, typically, additional code needs to be appended to the algorithm in order to estimate them. Louis's identity (Louis, 1982) was used in this study to obtain a Monte Carlo estimation of the information matrix (Guo and Thompson, 1992, Ibrahim et al., 2001). A description of how the approach works is as follows. Let the complete information likelihood function be $L^c(\boldsymbol{\theta};\boldsymbol{x})$, where $\boldsymbol{\theta}$ is the full set of parameters to estimate. Then, the observed log-likelihood can be written as

$$\ell(\boldsymbol{\theta};\boldsymbol{y}) = \ln L(\boldsymbol{\theta};\boldsymbol{y}) = \ln L^c(\boldsymbol{\theta};\boldsymbol{x}) - \ln \frac{L^c(\boldsymbol{\theta};\boldsymbol{x})}{L(\boldsymbol{\theta};\boldsymbol{y})} = \ell^c(\boldsymbol{\theta};\boldsymbol{x}) - \ell^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y}) \quad (2.18)$$

where $\ell^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y}) = \ln \dfrac{L^c(\boldsymbol{\theta};\boldsymbol{x})}{L(\boldsymbol{\theta};\boldsymbol{y})}$ is the logarithm of the complete information likelihood function conditional on the observed information. After taking second derivatives on both sides of (2.18) we have

$$\frac{\partial^2 \ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = \frac{\partial^2 \ell^c(\boldsymbol{\theta};\boldsymbol{x})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 \ell^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

which can be written in terms of information matrices in order to apply the "missing information principle" (Orchard and Woodbury, 1972)

$$I(\boldsymbol{\theta};\boldsymbol{y}) = I^c(\boldsymbol{\theta};\boldsymbol{x}) - I^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y}) \quad (2.19)$$

where $I^c(\boldsymbol{\theta};\boldsymbol{x}) = -E\left[H^c(\boldsymbol{\theta};\boldsymbol{x})\right]$ is the complete information matrix, $H^c(\boldsymbol{\theta};\boldsymbol{x}) = \dfrac{\partial^2 \ell^c(\boldsymbol{\theta};\boldsymbol{x})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ is the complete information Hessian, and $I^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y})$ can be viewed as the missing information matrix. Louis (1982) showed that this last matrix could be written as

$$I^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y}) = -E\left[\frac{\partial^2\ell^m(\boldsymbol{\theta};\boldsymbol{x}\mid\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \mathrm{Var}\left[S^c(\boldsymbol{\theta};\boldsymbol{x})\right] = \\ E\left[S^c(\boldsymbol{\theta};\boldsymbol{x})S^c(\boldsymbol{\theta};\boldsymbol{x})'\right] - E\left[S^c(\boldsymbol{\theta};\boldsymbol{x})\right]E\left[S^c(\boldsymbol{\theta};\boldsymbol{x})'\right] \tag{2.20}$$

where $S^c(\boldsymbol{\theta};\boldsymbol{x}) = \dfrac{\partial\ell^c(\boldsymbol{\theta};\boldsymbol{x})}{\partial\boldsymbol{\theta}}$ is the complete information score vector. All the expectations are taken with respect to the distribution $f\left(\boldsymbol{x}\mid\boldsymbol{y},\boldsymbol{\theta}^{EM}\right)$, where $\boldsymbol{\theta}^{EM}$ is the final MCEM estimation of $\boldsymbol{\theta}$. The evaluation of all the expectations involved commonly prevents the estimation of the observed information matrix in (2.20) by direct calculation. Monte Carlo estimates of the expected complete information Hessian and score can be used to circumvent the problem and estimate the terms in the right hand side of (2.20). To sum up, the procedure implemented in this study is:

Step 1. Use the Gibbs sampler described above to simulate a sequence $\boldsymbol{y}_i^{*(1)},\ldots,\boldsymbol{y}_i^{*(R+r_{burn})}$ while holding $\boldsymbol{\theta}=\boldsymbol{\theta}^{EM}$. Eliminate a number r_{burn} of simulations from the beginning of the sequence.

Step 2. Use the remaining simulations to estimate the expectation of the complete and missing information matrices by using

$$I^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}\right) = -\sum_{i=1}^{N} E\left[H_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}_i\right)\right] \cong -\sum_{i=1}^{N}\frac{1}{R}\sum_{r=1}^{R} H_i^{c(r)}\left(\boldsymbol{\theta}^{EM};\boldsymbol{y}_i^{*(r)}\mid\boldsymbol{y}_i\right)$$

$$I^m\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}\mid\boldsymbol{y}\right)=\sum_{i=1}^{N}\left\{E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}_i\right)S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}_i\right)^{'}\right]-E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}_i\right)\right]E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}_i\right)^{'}\right]\right\}$$
$$\cong\sum_{i=1}^{N}\left\{\frac{1}{R}\sum_{r=1}^{R}S_i^{c(r)}\left(\boldsymbol{\theta}^{EM},\boldsymbol{y}_i^{*(r)}\mid\boldsymbol{y}_i\right)S_i^{c(r)}\left(\boldsymbol{\theta}^{EM},\boldsymbol{y}_i^{*(r)}\mid\boldsymbol{y}_i\right)^{'}-\frac{1}{R}\sum_{r=1}^{R}S_i^{c(r)}\left(\boldsymbol{\theta}^{EM};\boldsymbol{y}_i^{*(r)}\mid\boldsymbol{y}_i\right)\frac{1}{R}\sum_{r=1}^{R}S_i^{c(r)}\left(\boldsymbol{\theta}^{EM};\boldsymbol{y}_i^{*(r)}\mid\boldsymbol{y}_i\right)^{'}\right\}$$

Expressions for the contributions from each observation to the Hessian and score are standard results from the theory of the multivariate normal distribution. Finally, plug the Monte Carlo estimates of $I^c\left(\theta;\boldsymbol{x}\right)$ and $I^m\left(\theta;\boldsymbol{x}\mid\boldsymbol{y}\right)$ in (2.20) and take the inverse of the resulting estimate of $I\left(\theta;\boldsymbol{y}\right)$ to get the asymptotic covariance matrix of θ^{EM}.

4 The numerical integration approach and comparison with the MCEM algorithm.

In this section I use both numerical integration and the MCEM algorithm to estimate the parameters of the same model using the same data. The performances of the two approaches are then compared. I applied the two methods on data from a survey administered to Maryland farmers in 1998 in order to evaluate a conservation cost-sharing program. For a detailed description of the survey see Lichtenberg and Smith-Ramírez (2004).

The econometric setting attempts to estimate the effect of cost-sharing payments on farmers' conservation efforts. The model to solve is the equation system in (2.2), where the dependent variables are defined as follows. Variable y_{1i}^* is the amount of cost-sharing money awarded to the farmer, y_{2i}^* is the number of conservation practices used, and y_{3i}^* is the share of land on which these practices are implemented.

None of these variables is observed fully. We observe only a dichotomous indicator, y_{1i}, of whether cost-share funding was awarded, i.e.

$$y_{1i} = \begin{cases} 1 & \text{if } y_{1i}^* > 0 \\ 0 & \text{if } y_{1i}^* \leq 0 \end{cases}$$

while the observed counterparts of y_{2i}^* and y_{3i}^* are assumed censored from below at zero, i.e.

$$y_{ji} = \begin{cases} y_{ji}^* & \text{if } y_{ji}^* > 0 \\ 0 & \text{if } y_{ji}^* \leq 0 \end{cases} \qquad j = 2,3$$

4.1 Estimation by numerical integration

The general form of the observed-information likelihood function for the equation system (2.2) is

$$\begin{aligned} L = & \prod_{\substack{y_{1i}=0 \\ y_{2i}=0 \\ y_{3i}=0}} \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{-\infty}^{0} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{2i}^* dy_{3i}^* \cdot \prod_{\substack{y_{1i}=1 \\ y_{2i}=0 \\ y_{3i}=0}} \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{2i}^* dy_{3i}^* \cdot \prod_{j=2,3} \prod_{\substack{y_{1i}=0 \\ y_{ji}=0}} \int_{-\infty}^{0}\int_{-\infty}^{0} f\left(y_{1i}, y_{2i}, y_{3i}\right) dy_{1i} dy_{ji} \cdot \\ & \prod_{j=2,3} \prod_{\substack{y_{1i}=1 \\ y_{ji}=0}} \int_{-\infty}^{0}\int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}, y_{3i}\right) dy_{1i}^* dy_{ji} \cdot \prod_{y_{1i}=0} \int_{-\infty}^{0} f\left(y_{1i}^*, y_{2i}, y_{3i}\right) dy_{1i}^* \cdot \prod_{y_{1i}=1} \int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}, y_{3i}\right) dy_{1i}^* \end{aligned}$$

However, since all possible combinations of values for the dependent variables do not exist in the data set used, the observed information likelihood can be reduced to (Lichtenberg and Smith-Ramírez, 2004)

$$L(\boldsymbol{\theta},\Sigma \mid \boldsymbol{y},\boldsymbol{X}) = \prod_{\substack{y_{1i}=0\\y_{2i}=0\\y_{3i}=0}} \int_{-\infty}^{-X_{3i}\beta_3}\int_{-\infty}^{-X_{2i}\beta_2}\int_{-\infty}^{-X_{1i}\beta_1} \phi_3(\varepsilon_{1i},\varepsilon_{2i},\varepsilon_{3i})\, d\varepsilon_{1i} d\varepsilon_{2i} d\varepsilon_{3i} \times$$

$$\prod_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}>0}} \int_{-\infty}^{-X_{1i}\beta_1} \phi_3(\varepsilon_{1i},\varepsilon_{2i},\varepsilon_{3i})\, d\varepsilon_i \times \prod_{\substack{y_{1i}=1\\y_{2i}>0\\y_{3i}>0}} \int_{-X_{1i}\beta_1}^{\infty} \phi_3(\varepsilon_{1i},\varepsilon_{2i},\varepsilon_{3i})\, d\eta_i =$$

$$\prod_{\substack{y_{1i}=0\\y_{2i}=0\\y_{3i}=0}} \int_{-\infty}^{-X_{3i}\beta_3}\int_{-\infty}^{-X_{2i}\beta_2}\int_{-\infty}^{-X_{1i}\beta_1} \phi_3(\varepsilon_{1i},\varepsilon_{2i},\varepsilon_{3i})\, d\varepsilon_{1i} d\varepsilon_{2i} d\varepsilon_{3i} \times$$

$$\prod_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}>0}} \phi_2(\varepsilon_{2i},\varepsilon_{3i}) \int_{-\infty}^{-X_{1i}\beta_1} \phi_{1|2,3}(\varepsilon_{1i} \mid \varepsilon_{2i},\varepsilon_{3i})\, d\varepsilon_i \times \prod_{\substack{y_{1i}=1\\y_{2i}>0\\y_{3i}>0}} \phi_2(\varepsilon_{2i},\varepsilon_{3i}) \int_{-X_{1i}\beta_1}^{\infty} \phi_{1|2,3}(\varepsilon_{1i} \mid \varepsilon_{2i},\varepsilon_{3i})\, d\varepsilon_i$$

where $\boldsymbol{\theta} = (\beta_1,\gamma_2,\beta_2,\gamma_3,\beta_3)$, $\phi_m(\boldsymbol{\varepsilon}) = (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\left(-\frac{\boldsymbol{\varepsilon}'\Sigma^{-1}\boldsymbol{\varepsilon}}{2}\right)$ is the m-dimensional normal pdf and $\phi_{j|k,l}(\varepsilon_{ji} \mid \varepsilon_{ki},\varepsilon_{li})$ is the normal pdf of ε_{ji} conditional on $(\varepsilon_{ki},\varepsilon_{li})$. After a little algebra the log-likelihood of the observed data can be written as

$$\ell(\boldsymbol{\theta},\Sigma \mid \boldsymbol{y}) = \sum_{\substack{y_{2i}>0\\y_{3i}>0}} \ln \phi_2\left(\frac{\varepsilon_{2i}}{\sigma_{\varepsilon_2}},\frac{\varepsilon_{3i}}{\sigma_{\varepsilon_3}},\rho_{\varepsilon_2\varepsilon_3}\right) +$$

$$\sum_{\substack{y_{1i}=0\\y_{2i}=0\\y_{3i}=0}} \ln \Phi_3\left(-\frac{X_{1i}\beta_1}{1},-\frac{X_{2i}\beta_1}{\sigma_{\varepsilon_2}},-\frac{X_{3i}\beta_2}{\sigma_{\varepsilon_3}},\rho_{\varepsilon_1\varepsilon_2},\rho_{\varepsilon_1\varepsilon_3},\rho_{\varepsilon_2\varepsilon_3}\right) +$$

$$\sum_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}>0}} \ln \Phi_1\left(-\frac{X_{1i}\beta_1 + \frac{\rho_{\varepsilon_1\varepsilon_2}-\rho_{\varepsilon_1\varepsilon_3}\rho_{\varepsilon_2\varepsilon_3}}{1-\rho^2_{\varepsilon_2\varepsilon_3}}\frac{\varepsilon_{2i}}{\sigma_{\varepsilon_2}} + \frac{\rho_{\varepsilon_1\varepsilon_3}-\rho_{\varepsilon_1\varepsilon_2}\rho_{\varepsilon_2\varepsilon_3}}{1-\rho^2_{\varepsilon_2\varepsilon_3}}\frac{\varepsilon_{3i}}{\sigma_{\varepsilon_3}}}{\sqrt{(1-\rho^2_{\varepsilon_1\varepsilon_2}-\rho^2_{\varepsilon_1\varepsilon_3}-\rho^2_{\varepsilon_2\varepsilon_3}+2\rho_{\varepsilon_1\varepsilon_2}\rho_{\varepsilon_1\varepsilon_3}\rho_{\varepsilon_2\varepsilon_3})/(1-\rho^2_{\varepsilon_2\varepsilon_3})}}\right) + \quad (2.21)$$

$$\sum_{\substack{y_{1i}=1\\y_{2i}>0\\y_{3i}>0}} \ln \Phi_1\left(\frac{X_{1i}\beta_1 + \frac{\rho_{\varepsilon_1\varepsilon_2}-\rho_{\varepsilon_1\varepsilon_3}\rho_{\varepsilon_2\varepsilon_3}}{1-\rho^2_{\varepsilon_2\varepsilon_3}}\frac{\varepsilon_{2i}}{\sigma_{\varepsilon_2}} + \frac{\rho_{\varepsilon_1\varepsilon_3}-\rho_{\varepsilon_1\varepsilon_2}\rho_{\varepsilon_2\varepsilon_3}}{1-\rho^2_{\varepsilon_2\varepsilon_3}}\frac{\varepsilon_{3i}}{\sigma_{\varepsilon_3}}}{\sqrt{(1-\rho^2_{\varepsilon_1\varepsilon_2}-\rho^2_{\varepsilon_1\varepsilon_3}-\rho^2_{\varepsilon_2\varepsilon_3}+2\rho_{\varepsilon_1\varepsilon_2}\rho_{\varepsilon_1\varepsilon_3}\rho_{\varepsilon_2\varepsilon_3})/(1-\rho^2_{\varepsilon_2\varepsilon_3})}}\right)$$

where $\Phi_m(\cdot)$ is the m-dimensional standard normal cdf.

In this study the log-likelihood function in (2.21) was maximized using the routine FMINUNC in Matlab. I programmed the 3-dimensional standard normal cdf according to the methodology proposed by Steck (1958), which allows reducing the 3-dimensional integral to functions involving only 1-dimensional integrals of exponential functions and the univariate normal cdf. The information matrix was calculated from a finite-difference estimation of the Hessian of the objective function.

Starting values in the approximation area of the maximum were very hard to find. A first attempt using OLS estimates failed to converge. A second approach attempted to estimate the equations in (2.2) by pairs and then use a combination of the resulting estimates as starting values for the 3-equation system. Matlab routines using OLS estimates as starting values were written to estimate these smaller equation systems. However, although convergence in the system constituted by the second and third equations was easily accomplished, neither the routine for the system constituted by the first and second equations nor the one for the system constituted by the first and third equations converged. The routines either ceased to improve in the search for the optimum or the correlation between the disturbances escaped the parameter space.

Finally, a grid search was implemented. To decide the dimension of the grid, estimation attempts were made by fixing one, two, three and four parameters in Σ. Convergence was attained only after one of the two variances and all the correlation coefficients were fixed. Thus, a four-dimensional grid search was implemented on those parameters. Ten equally spaced points from –0.9 to 0.9 were chosen for the correlation coefficients and four equally spaced points from 2 to 8 were taken for σ_{ε_2}, which generated a 4,000-point grid. The time required solving a grid of this size easily becomes unaffordable when the objective function involves high dimensional integrals as in (2.21); however, the actual number of grid points that need to be used can be drastically reduced by two ways. First, it must be noticed that Σ must be kept positive definite at every moment during the estimation. Many points in the grid described above do not satisfy that requirement and they must be eliminated from the search. Second, since we are interested in finding a neighborhood of the global maximum only, by careful monitoring of the search it is possible to exclude large sets of grid points surrounding low values of the objective function. Notice that this last approach is advisable only when the objective function behaves smoothly. A real risk of missing the global maximum exists otherwise. By proceeding this way the grid search used in this study required less than 350 points to locate a point in the approximation area of the global maximum. However, despite of the significant reduction in the size of the grid, it took about 60 hours[1] to solve the grid search and make the final estimation to obtain the maximum maximorum.. Results[2] of the final estimation are presented in Table 2.1.

4.2 Estimation by the MCEM algorithm

The general MCEM method has been already discussed in Section 3. Different sets of starting values were used in this work to reduce the possibility of missing the global

[1] On an AMD Athlon XP-M 2000^{+}, 512 MB RAM, Windows XP, Matlab 6.5.

maximum. For a quite broad array of starting values, the MCEM algorithm always converged to the same maximizer and always kept the estimates in the interior of the parameter space.

The iteration paths for the likelihood function and selected parameters are presented in Figure 2.1. OLS estimates were used as starting values for the parameters in θ and an identity matrix was used for the covariance matrix of the disturbances. The routine converged after 393 iterations when using the stopping criteria described above. The Gibbs sampler was started with 300 simulations and increased by 15 simulations at every iteration of the EM algorithm, i.e. $K^{(m)} = 300 + 15(m-1)$. The number of dismissed simulations, k_{burn}, was kept constant at 150. The algorithm converged after 2.5 hours[3].

A Gibbs sample with $R = 3300$ and $r_{burn} = 300$ was used to estimate the information matrix. Results of the Monte Carlo EM estimation are presented in Table 2.1.

---

[2] I have not included the variable names since these results are presented only for comparison purposes. Details can be found in Lichtenberg and Smith-Ramírez (2004).

[3] On an AMD Athlon XP-M 2000$^+$, 512 MB RAM, Windows XP, Matlab 6.5.

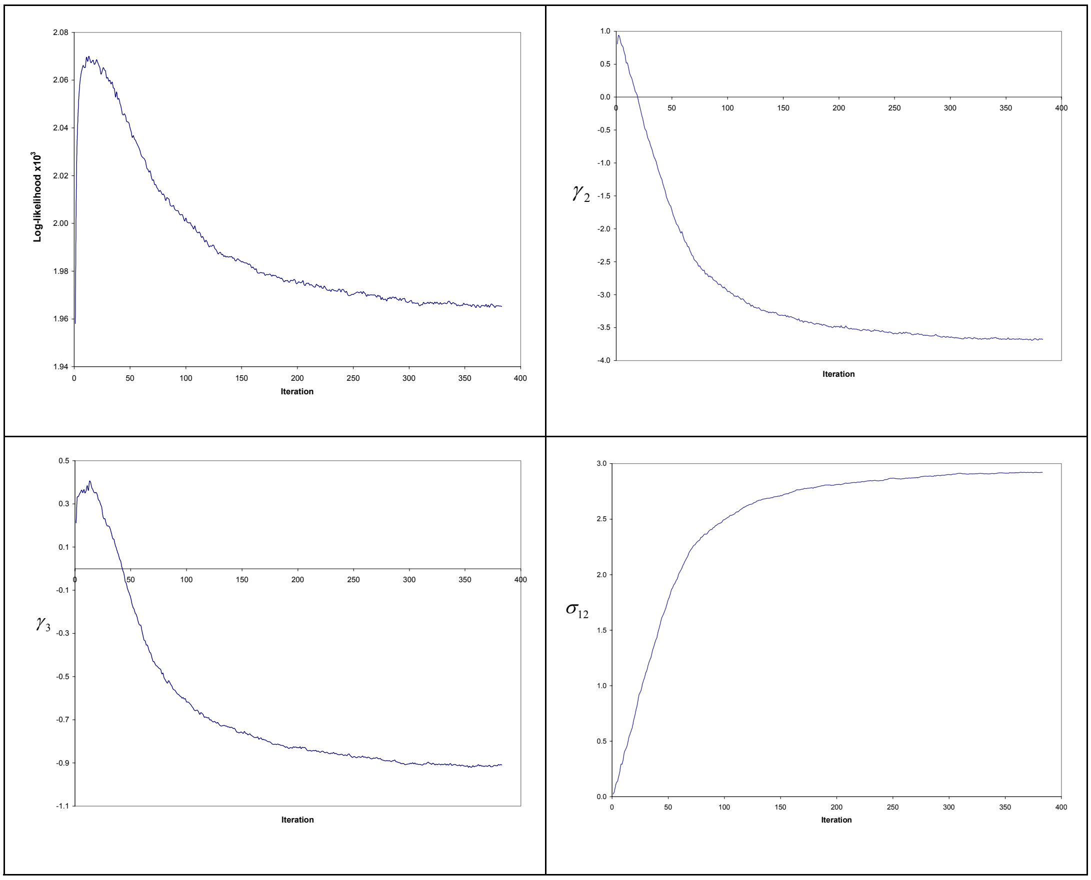


Figure 2.1. Iteration paths of the expected log-likelihood and selected parameter estimates.

Table 2.1. ML estimates obtained by Numerical Integration and Monte Carlo EM algorithm

Equation 1					Equation 2					Equation 3				
Parameter	Numerical Integration		MCEM		Parameter	Numerical Integration		MCEM		Parameter	Numerical Integration		MCEM	
	Estim.	St. err.	Estim.	St. err.		Estim.	St. err.	Estim.	St. err.		Estim.	St. err.	Estimate	St. err.
					γ_2	-3.7074	1.0037	-3.6860	0.6633	γ_3	-0.9229	0.5486	-0.9163	0.3701
$\beta_{1,1}$	-0.7881	0.8856	-0.7987	0.8661	$\beta_{2,1}$	5.6232	1.7064	5.6173	1.6708	$\beta_{3,1}$	2.4189	0.8131	2.4572	0.8024
$\beta_{1,2}$	-0.2282	0.1515	-0.2259	0.1481	$\beta_{2,2}$	-0.8770	0.2572	-0.8756	0.2557	$\beta_{3,2}$	-0.2127	0.1234	-0.2175	0.1237
$\beta_{1,3}$	0.2126	0.1246	0.2108	0.1156	$\beta_{2,3}$	0.5992	0.2212	0.5981	0.2189	$\beta_{3,3}$	0.2409	0.1055	0.2474	0.1055
$\beta_{1,4}$	0.5194	0.2436	0.5245	0.2364	$\beta_{2,4}$	0.7933	0.5164	0.7917	0.5109	$\beta_{3,4}$	0.1260	0.2447	0.0919	0.2452
$\beta_{1,5}$	0.1809	0.0598	0.1807	0.0591	$\beta_{2,5}$	0.1665	0.1404	0.1656	0.1376	$\beta_{3,5}$	-0.0077	0.0670	-0.0045	0.0660
$\beta_{1,6}$	-0.0254	0.0374	-0.0255	0.0349	$\beta_{2,6}$	0.1776	0.0683	0.1777	0.0682	$\beta_{3,6}$	0.1279	0.0322	0.1279	0.0326
$\beta_{1,7}$	-0.6267	0.4196	-0.6193	0.4132	$\beta_{2,7}$	-0.3023	0.9046	-0.2999	0.8983	$\beta_{3,7}$	0.2795	0.4265	0.1697	0.4310
$\beta_{1,8}$	0.6602	0.2999	0.6603	0.3009	$\beta_{2,8}$	2.9621	0.9029	2.9580	0.8907	$\beta_{3,8}$	0.1504	0.4280	0.1656	0.4259
$\beta_{1,9}$	0.2370	0.4209	0.2368	0.4196	$\beta_{2,9}$	0.9667	0.9991	0.9667	0.9861	$\beta_{3,9}$	0.0294	0.4704	0.0008	0.4715
$\beta_{1,10}$	-0.5050	0.4405	-0.5042	0.4364	$\beta_{2,10}$	-0.4288	0.9658	-0.4272	0.9424	$\beta_{3,10}$	-0.9509	0.4544	-0.9809	0.4512
$\beta_{1,11}$	-0.4257	0.4467	-0.4211	0.4372	$\beta_{2,11}$	1.1889	0.9724	1.1912	0.9567	$\beta_{3,11}$	-0.1385	0.4567	-0.1933	0.4570
$\beta_{1,12}$	0.7764	0.5478	0.7765	0.5484	$\beta_{2,12}$	1.9579	1.4739	1.9549	1.4671	$\beta_{3,12}$	1.0929	0.6942	1.1256	0.6993
$\beta_{1,13}$	1.5578	0.6997	1.5608	0.6946	$\beta_{2,13}$	2.4029	2.0006	2.3937	1.9660	$\beta_{3,13}$	0.3721	0.9544	0.3632	0.9447
$\beta_{1,14}$	0.8016	0.3230	0.7989	0.3147	$\beta_{2,14}$	2.7804	0.5673	2.7797	0.5651	$\beta_{3,14}$	0.9317	0.2699	0.9430	0.2720
$\beta_{1,15}$	-0.3057	0.3915	-0.2967	0.3720	$\beta_{2,15}$	0.8671	0.8077	0.8654	0.8013	$\beta_{3,15}$	-0.5440	0.3825	-0.6334	0.3864

$\beta_{1,16}$	-0.5323	0.3793	-0.5337	0.3714	$\beta_{2,16}$	2.2971	0.8371	2.2984	0.8327	$\beta_{3,16}$	0.5500	0.3952	0.5474	0.3974
$\beta_{1,17}$	-0.2323	0.2829	-0.2327	0.2629	$\beta_{2,17}$	0.6061	0.4912	0.6062	0.4891	$\beta_{3,17}$	0.0311	0.2321	0.0364	0.2338
$\beta_{1,18}$	-0.7459	0.2661	-0.7483	0.2534	$\beta_{2,18}$	-0.4733	0.5233	-0.4723	0.5186	$\beta_{3,18}$	-1.0625	0.2484	-1.0542	0.2485
$\beta_{1,19}$	0.3718	0.2607	0.3725	0.2449	$\beta_{2,19}$	0.4493	0.4998	0.4481	0.4973	$\beta_{3,19}$	-0.6189	0.2362	-0.6046	0.2378
$\beta_{1,20}$	0.4153	0.5098	0.4219	0.4890	$\beta_{2,20}$	0.4398	1.2337	0.4407	1.2298	$\beta_{3,20}$	-1.2735	0.6017	-1.2638	0.6035
$\beta_{1,21}$	-2.1581	0.9470	-2.1740	0.8672	$\beta_{2,21}$	0.5753	1.2234	0.5797	1.2053	$\beta_{3,21}$	-0.1927	0.5817	-0.1420	0.5771
$\beta_{1,22}$	-1.0930	1.0749	-1.0957	0.6018	$\beta_{2,22}$	-3.6848	1.0036	-3.6882	1.0091	$\beta_{3,22}$	-1.7546	0.5032	-1.7449	0.5119
$\beta_{1,23}$	-1.9051	0.5381	-1.9013	0.4659										
$\beta_{1,24}$	0.2723	0.3561	0.2740	0.3431										
$\beta_{1,25}$	-0.2717	0.5340	-0.2715	0.4657										
σ_{12}	2.9339	0.4003	2.9217	0.0966										
σ_{13}	0.7756	0.2481	0.7823	0.0609										
σ_{22}	16.0218	1.6406	16.001	0.6411										
σ_{23}	4.9290	0.5513	4.9887	0.2782										
σ_{33}	3.5432	0.3591	3.6201	0.1633										

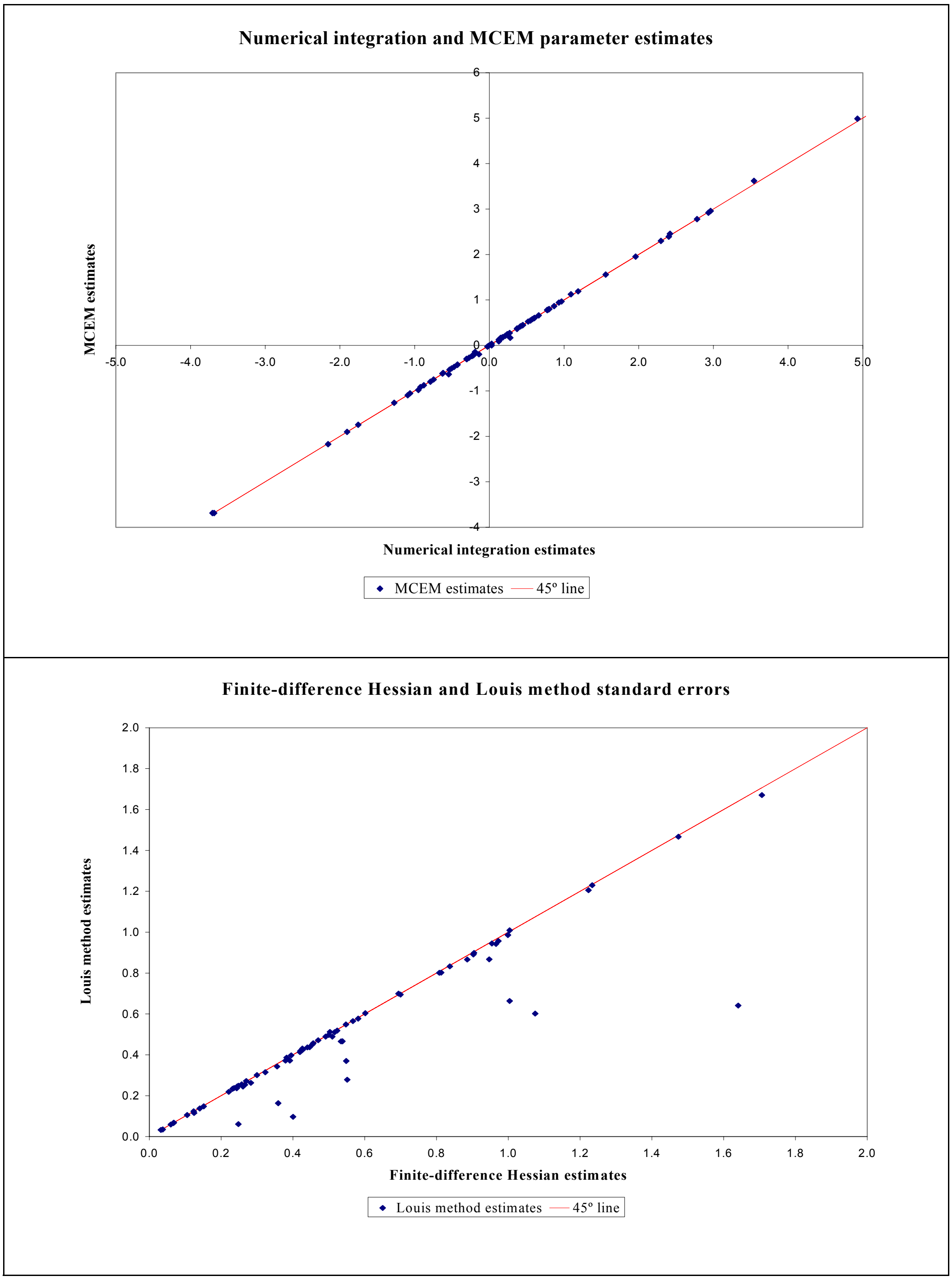


Figure 2.2. Comparison between numerical integration and MCEM estimates.

Graphical comparisons between the estimates obtained by numerical integration and those obtained by the MCEM algorithm are depicted in Figure 2.2. The remarkable match in the parameter estimates proves that both approaches converged to the same maximizer. The main difference, of course, is the robustness of the MCEM to the election of starting values, which allowed achieving the solution in less than one-twentieth of the time needed when using numerical integration. Certainly, the use of QMC integration or the use of probability simulators may help to reduce the estimation time when using the numerical integration approach. Yet, these methods only provide an alternative to estimate the integral terms in the likelihood function. They do not help neither with problems of "practical" identification nor with the starting value problem.

The matching for the standard errors in Table 2.1, however, is not that close. The lower graph in Figure 2.2 shows that, in general, the standard errors obtained by using a finite-difference Hessian are larger than those produced by Louis' method. In an attempt to reduce the disparity, the simulation size used to estimate the information matrix in the Monte Carlo approach was enlarged from $R = 3300$ to $R = 5300$. However, no significant change was observed in the estimates.

In order to determine if the origin of the mismatch was in the Hessian estimated by finite differences, the Hessian estimation was repeated several times reducing iteratively the size of the perturbation size[4]. This approach reduced the mismatch, which suggests that the origin of the problem is in the numerical estimation of the Hessian. The size of the perturbation, however, cannot be reduced arbitrarily. The

[4] By perturbation size I mean the magnitude of the finite difference used to calculate the numerical derivatives

numerical integration approach used to estimate the likelihood function in (2.21) may be free of Monte Carlo error but it has inaccuracies originated in the numerical integration procedure. This fact sets a lower bound for the size of the perturbation that we can use to estimate the Hessian: it cannot be smaller than the estimation error of the likelihood function. The accuracy of the numerical integrals can, certainly, be increased; however, it is well known that the computational costs of proceeding that way increment exponentially. Figure 2.2 presents the standard errors for a percentage perturbation equal to 10^{-4}.

The limitations of the numerical Hessian approach contrast with the advantages of the stochastic version of the Louis' method. The Louis' method is easy to implement and we can use it to obtain standard errors with any needed accuracy simply by increasing the number of simulations. Doing this is relatively inexpensive since the score and the Hessian of the expectation in (2.5) exist in closed forms and their calculation involves only matrix algebra.

5 Implementing the Monte Carlo EM algorithm for models with latent endogenous regressors.

This section extends the MCEM algorithm presented in the last section in order to include systems of structural equations, i.e. cases where latent variables show up on both sides of the equations. Only for the purpose of illustrating the flexibility of the method, consider again the equation system (2.2) but now with y_{1i}^{*} instead of y_{1i} at the right-hand side of the second and third equations, and y_{1i} defined as a polytomous variable

$$y_{1i} = \begin{cases} b_1 & \text{if } a_0 < y_{1i}^* \le a_1 \\ b_2 & \text{if } a_1 < y_{1i}^* \le a_2 \\ \vdots & \\ b_k & \text{if } a_{k-1} < y_{1i}^* \le a_k \end{cases} \tag{2.22}$$

The structural model is now

$$\begin{aligned} y_{1i}^* &= X_{1i}\beta_1 + \varepsilon_{1i} \\ y_{2i}^* &= \gamma_2 y_{1i}^* + X_{2i}\beta_2 + \varepsilon_{2i} \\ y_{3i}^* &= \gamma_3 y_{1i}^* + X_{3i}\beta_3 + \varepsilon_{3i} \end{aligned} \tag{2.23}$$

Thus, under the normality assumption, the complete data likelihood function can be written as

$$L(\boldsymbol{\beta},\Gamma,\Sigma \mid \boldsymbol{x}) = \prod_i f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) = \prod_i \left[\frac{1}{(2\pi)^{3/2} |\Gamma| |\Sigma|^{1/2}} \exp\left(-\frac{\varepsilon_i' \Sigma^{-1} \varepsilon_i}{2} \right) \right]$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$, $\Gamma = \begin{bmatrix} 1 & -\gamma_2 & -\gamma_3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\varepsilon_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{pmatrix} = \begin{pmatrix} y_{1i}^* - X_{1i}\beta_1 \\ y_{2i}^* - \gamma_2 y_{1i}^* - X_{2i}\beta_2 \\ y_{3i}^* - \gamma_3 y_{1i}^* - X_{3i}\beta_3 \end{pmatrix}$, and

Σ is defined as in (2.3).

Correspondingly, the complete information log-likelihood function and its expectation are

$$\ell^c\left(\boldsymbol{\beta},\Gamma,\Sigma \mid \boldsymbol{x}\right) = -\frac{3N}{2}\ln\left(2\pi\right) - N\ln\left|\Gamma\right| - \frac{N}{2}\ln\left|\Sigma\right| - \frac{1}{2}\sum_i \operatorname{tr}\left(\Sigma^{-1}\varepsilon_i\varepsilon_i^{'}\right) \quad (2.24)$$

$$E\left[\ell^c\left(\boldsymbol{\beta},\Gamma,\Sigma \mid \boldsymbol{x}\right)\right] = -\frac{3N}{2}\ln\left(2\pi\right) - N\ln\left|\Gamma\right| - \frac{N}{2}\ln\left|\Sigma\right| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_i E\left[\varepsilon_i\varepsilon_i^{'}\right]\right) \quad (2.25)$$

Where N is the total number of observations and the expectation operator indicates expectation conditional on observed information and distribution assumptions. The E-step at iteration $m+1$ requires the calculation of

$$\begin{aligned} Q_i\left(\boldsymbol{\beta},\Gamma,\Sigma \mid \boldsymbol{\beta}^{(m)},\Gamma^{(m)},\Sigma^{(m)},\boldsymbol{y}\right) &= E\left[\varepsilon_i\varepsilon_i^{'} \mid \boldsymbol{\beta}^{(m)},\Gamma^{(m)},\Sigma^{(m)},\boldsymbol{y}\right] \\ &= E\left[\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-\gamma_2 y_{1i}^{*}-X_{2j}\beta_2 \\ y_{3i}^{*}-\gamma_3 y_{1i}^{*}-X_{3j}\beta_3 \end{pmatrix}\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-\gamma_2 y_{1i}^{*}-X_{2j}\beta_2 \\ y_{3i}^{*}-\gamma_3 y_{1i}^{*}-X_{3j}\beta_3 \end{pmatrix}^{'} \mid \boldsymbol{\beta}^{(m)},\Gamma^{(m)},\Sigma^{(m)},\boldsymbol{y}\right] \quad (2.26) \\ &= \Gamma^{'}\sigma_i^{2(m)}\Gamma \; + \; \begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)}-\gamma_2\mu_{y_{1i}^{*}}^{(m)}-X_{2i}\beta_2 \\ \mu_{y_{3i}^{*}}^{(m)}-\gamma_3\mu_{y_{1i}^{*}}^{(m)}-X_{3i}\beta_3 \end{pmatrix}\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)}-\gamma_2\mu_{y_{1i}^{*}}^{(m)}-X_{2i}\beta_2 \\ \mu_{y_{3i}^{*}}^{(m)}-\gamma_3\mu_{y_{1i}^{*}}^{(m)}-X_{3i}\beta_3 \end{pmatrix}^{'} \end{aligned}$$

Where $\sigma_i^{2(m)}$ and $\mu_{y_{ji}^{*}}^{(m)}$ are defined as in expressions (2.7) and (2.8). Some small modifications must be introduced to the Gibbs sampler implemented in section 3.2 in order to estimate these moments. The conditional mean $\mu_{ji|i(-j)}^{(m)}$ must be now estimated according to

$$\mu_{ji|i(-j)}^{(m)} = X_{ji}\beta_j^{(m)} + \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Sigma^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^* \middle| \Sigma^{(m)}\right)\right]^{-1}\left(\boldsymbol{y}_{i|-j}^* - \boldsymbol{y}_{i|-j}^*\boldsymbol{\gamma}_{-j}^{(m)} - \boldsymbol{X}_{i|-j}\boldsymbol{\beta}_{-j}^{(m)}\right)$$

where $\gamma_{-j}^{(m)} = \begin{pmatrix} \gamma_1^{(m)} \\ \vdots \\ \gamma_{j-1}^{(m)} \\ \gamma_{j+1}^{(m)} \\ \vdots \\ \gamma_k^{(m)} \end{pmatrix}$

To construct a sample conditional on the observed information as defined in (2.22) proceed as follows. For every $a_\kappa < y_{1i} \le a_{\kappa+1}$ simulate y_{1i}^* from a normal distribution with mean $\mu_{1i|i(-1)}^{(m)}$ and variance $\sigma_{1i|-1}^{2(m)}$ truncated below at a_κ and truncated above at $a_{\kappa+1}$. Do the same for every $\kappa = 0,1,2,\ldots k$, where k is the number of intervals defined by the polytomous variable.

Simulations for the unobserved values of y_{2i}^* and y_{3i}^* are obtained in the same way as in Section 3.2.

The maximization step does not differ significantly from the case analyzed previously except by the presence of Γ in the log-likelihood function, which motivates a slight change in the arguments of the conditional maximization steps. The objective function is

$$-\frac{3N}{2}\ln(2\pi) - N\ln|\Gamma| - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_i Q_i\left(\boldsymbol{\beta},\Gamma,\Sigma \mid \boldsymbol{\beta}^{(m)},\Gamma^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right) \quad (2.27)$$

The first conditional maximization updates $\boldsymbol{\beta}$ conditional on the elements in Γ and Σ. From (2.26) the estimate of $\boldsymbol{\beta}$ can still be written as a generalized least squares estimator:

$$\boldsymbol{\beta}^{(m+1)} = \left[X_d' \left(\Sigma^{(m)} \otimes I_N \right)^{-1} X_d \right]^{-1} X_d' \left(\Sigma^{(m)} \otimes I_N \right)^{-1} \hat{\mu}_{y^*}^{(m)} \qquad (2.28)$$

where

$X_d = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix}$, $\hat{\mu}_{y^*}^{(m)} = \left(\Gamma'^{(m)} \otimes I_N \right) \mu_{y^*}^{(m)}$, I_N is the identity matrix of dimension N and $\mu_{y^*}^{(m)}$ is the column vector defined in (2.15).

The second conditional maximization updates Γ and Σ conditional on the updated estimate of $\boldsymbol{\beta}$. Numerical techniques must be used in this step to maximize

$$-\frac{3N}{2} \ln(2\pi) - N \ln|\Gamma| - \frac{N}{2} \ln|\Sigma| - \frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \sum_i Q_i \left(\boldsymbol{\beta}^{(m+1)}, \Gamma \mid \boldsymbol{\beta}^{(m)}, \Gamma^{(m)}, \Sigma^{(m)}, \boldsymbol{y} \right) \right)$$

with respect to the elements in Γ and Σ and obtain estimates for $\Gamma^{(m+1)}$ and $\Sigma^{(m+1)}$. Notice that the second term in the objective function vanishes in this particular example as Γ is triangular.

It is evident that the combination of the Expectation and Maximization steps as described here can be extended to admit systems with a larger number of linear equations involving any type of latent variables. Only small adjustments to the Gibbs sampler in order to take into account the types of latent variables involved are necessary.

6 Final remarks

This chapter has presented a MCEM algorithm suitable for estimating systems of simultaneous equations and structural models that contain latent variables. The applicability of the model is independent of whether the latent variables appear in the model as dependent variables or as endogenous regressors. The general formulation presented in Section 5 permits that the algorithm can be applied to solve a variety of models with latent structures from a one-equation tobit to a n-equation multinomial probit. Only small adjustments in the Gibbs sampler are necessary to shift from one model to another in order to internalize the types of latent variables and the nature of the unobserved information involved.

The MCEM algorithm as formulated here has a number of advantages over more traditional methods. First, it does not require integrating the unobserved information out from the likelihood function. This characteristic reduces the estimation time dramatically as no numerical integration is needed and, similarly to methods based on probability simulators, permits to solve problems involving more than three latent variables. Second, it reduces the estimation of the vector of slopes to the calculation of a GLS estimator and numerical optimization is required only to estimate the elements in the disturbance covariance matrix. Since the GLS estimator and the gradient and Hessian of the objective function for the estimation of the disturbance covariance matrix have closed forms, almost no time is consumed in the Maximization step and it is easier to keep the whole set of parameters in the parameter space. This property of the MCEM reduces substantially the problems of “fragile” identification and selection of starting values, which are serious limitations of traditional approaches. Third, it can accommodate potentially any linear-in-parameters equation system. This is valid not only for the cross-sectional models mentioned above but also for panel data models and stochastic frontier models, where the random effects and efficiency terms can be treated as one more latent variable. Finally, the estimation of standard errors by the Louis method circumvents the limitations associated to the estimation of numerical Hessians by finite-difference methods, which is the standard in traditional procedures. The accuracy of the estimates of the standard errors can be improved easily by increasing the number of simulations of a closed form of the Information matrix, which is much less expensive than reducing the perturbation size in the numerical Hessian approach.

Chapter 3: How Green Are Cost Sharing Payments? A Joint Disaggregated Evaluation of the Interactions between Farmers' Conservation and Production Decisions

1 Introduction

In the last twenty years farm support policies have shown an important shift towards the conservation of farmland natural resources and the reduction of pollution from agricultural activities. In the United States, for example, the Conservation Reserve Program (CRP), a 36-million-acre land retirement program, was established in 1985 with the aim of reducing agricultural non-point source pollution and preserving natural resources for the future by taking land out of production. Eleven years later, the 1996 Farm Bill replaced four existing programs (the Agricultural Conservation program, the Water Quality Control Program, the Great Plains Conservation Program, and the Colorado River Basin Salinity Control Program) by the Environmental Quality Incentives Program (EQIP). Complementarily to CRP, EQIP does not retire land from production but aims to reduce non-point source pollution generated on working land and to sustain farmland productivity. The 2002 Farm Bill has reinforced the emphasis on working-land programs by authorizing a fivefold increment in EQIP's budget and the creation of the Conservation Security Program (CSP).

Underlying land retirement programs like CRP is the presumption that rental payments will stop production activities on land enrolled without altering the use of the land either on non-participant farms or on the non-enrolled tracts of land of participant farms. If this presumption does not hold then these programs may have unintended consequences that undermine their effectiveness. Wu (2000), for example, has shown that farmers adjust to CRP actions by bringing marginal land into production. The expansion is a consequence of the rise in output prices caused by the reduction in crop acreage[5] and substitution effects derived from economies of scale and requirement of fixed inputs in farm production. Wu estimates that these unintended effects have offset 9% and 14% of the CRP benefits from reduction in water pollution and wind erosion, respectively.

Working-land cost-sharing programs like EQIP provide technical assistance and monetary incentives to farmers to implement best management practices (BMPs). The underlying presumption of these programs is that payments will change the technology used in existing operations without altering the scope or intensity of farm operations. Or, in other words, the adoption of BMPs will reduce negative environmental spillovers and resource degradation without changing the amount of land cropped and without increasing the intensity at which fertilizer and pesticides are used. Whether these premises are true or not has not been researched explicitly to date. Following the results presented for CRP, it is logical to ask whether the enrollment in a working-land conservation cost-sharing program affects production decisions and whether the way farmers adjust offsets part of the benefits of the program.

---

[5] To date, land enrolled in CRP amounts 9% of the cropland in the contiguous states, an area roughly the size of Iowa (Roberts and Lubowski, 2004).

One possibility is that adoption of conservation technologies makes it profitable to expand cropping, in which case expansion of cropping onto sensitive areas may reduce or eliminate the environmental gains generated by the adoption of the technology (Wu 2000). Other losses may come as result of cropping intensification. For instance, if adoption of conservation practices increases the use of double cropping, then a higher load of fertilizers and pesticides may be introduced in the ecosystem. Thus, it is clear that determination of the effect of cost sharing on acreage cropped requires analyzing the influence of conservation technology on cropping patterns at both the extensive and the intensive margins.

Most practices that aim to reduce soil erosion are land-quality augmenting since they alter (in fact, they reduce) the impact of land quality on productivity (Caswell and Zilberman, 1986). Reduced tillage, contour farming, strip cropping, grade stabilization, and rock- grass-lined waterways are all land quality augmenting as they prevent soil runoff and, by doing so, they reduce the impact of land steepness on farming activities. Although some studies about the effect that land-quality-augmenting technology may have on cropping exist, they focus on technologies designed primarily to increase productivity (Lichtenberg, 1989). The effect of adoption of conservation technology on acreage cropped has not been studied to date, however. Theoretical analyses (Lichtenberg, 2004; Malik and Shomaker, 1993) suggest that adoption of conservation technology may increase cropping at the extensive margin. Cropping expansion, i.e. the extensive margin effect, would happen as consequence of the higher profitability that conservation technology presents on highly erodible land when compared with traditional cropping techniques.

To understand this point better, let A be the total acreage of the farm, A_s the acreage under single cropping, and A_d the acreage under double cropping. By double cropping I mean the combination of two short-season crops planted on the same field, where one of the crops is planted following immediately the harvest of the other. Thus, double cropping allows obtaining two harvests from a single tract of land in a single season. Total acreage cropped or tilled in the farm is $A_s + 2A_d$, while the share of land cropped, i.e. the acreage used for planting crops, is only $A_s + A_d$. In this framework, an extensive margin effect is a change in the sum $A_s + A_d$, i.e. an increase or decrease in the total area planted regardless of how many times that area is cropped during the season. An intensive margin effect, on the other hand is a change in A_d, the area of the farm cropped more intensively. Notice that if double cropped acreage increases at the expenses of single cropped area only, there will be no extensive margin effect even though there is a positive intensive margin effect.

To the best of the author's knowledge, intensive margin effects of conservation technology have not been investigated empirically to date. Intensification via double cropping may arise from labor and time savings. For instance, no-till and minimum till techniques reduce labor and machinery time in crop production because no plowing is necessary before seeding.

There is another reason to analyze unintended effects of adoption of conservation technology on production decisions. Land retirement programs and conservation cost sharing programs are classified under the green-box exemption of the World Trade Organization (WTO); i.e. they are excluded from payment reduction commitments. EQIP has been considered "...the first green program of any significant magnitude in the U.S that is not a land retirement program" by Batie (1999). Existence of the green-box exemption is the WTO recognition of market failures to provide adequate incentives to farmers in order to supply socially optimum amounts of public goods and to protect natural resources (Klonsky and Jacquet, 2003). Green-box payments, as defined by the WTO (2004), must

> "… have no, or at most minimal, trade-distorting effects or effects on production. "

But do programs that promote the adoption of conservation technology fulfill this requirement? As indicated before, conservation technologies may increase land productivity by enhancing land quality or reducing the impact of land quality on productivity. Other technologies may reduce production costs. For instance, nutrient management and integrated pest management improve efficiency of fertilizer and pesticide applications respectively and, by doing so, reduce variable costs of production. Hence, programs that cost share conservation technology may expand production indirectly and, consequently, distort trade.

Working-land cost-sharing programs have traditionally been funded at more modest levels than land-retirement programs. However, the new emphasis placed by the 2002 Farm Bill on working-land programs suggests that this situation may change in the near future. Evidences of this new tendency are the authorization of a fivefold increase in the EQIP budget and the creation of the Conservation Security Program (CSP). Increased spending, a new program, and improvements in targeting and implementation as result of experience accumulated since creation of EQIP will certainly increase the environmental impact of cost sharing programs. However, it is also likely that associated impacts on farm production become more evident, which may bring these programs under a more stringent scrutiny of the WTO. Comprehension of the interactions between conservation and production decisions is necessary in order to design cost sharing programs having its larger impact on the provision of environmental amenities and minimal effects on trade.

The goal of this paper is threefold. First, I estimate the effect of adoption of conservation technology on cropping at both the extensive and intensive margins, focusing on three of the most popular practices used to control soil erosion in Maryland. Secondly, I estimate the effect of cost-sharing programs on the adoption levels of the three practices mentioned above. Finally, I estimate the extent to which cost-sharing payments have influenced the scope and intensity of cropping operations in Maryland. The remainder of this paper is organized as follows. Section Two presents the data and develops and estimates the econometric model. Section Three discusses the results and policy implications. Section Four concludes.

2. Data, econometric model and estimation

2.1 Data

Data used in the econometric estimation come from a University of Maryland BMP/Cost Share Survey administered in 1998. The Maryland Agricultural Statistics Service (MASS) conducted the survey. Stratified random sampling was used to ensure a sufficient number of responses from commercial operations, especially larger ones. Samples were drawn from the MASS master list of farmers and MASS provided expansion factors for deriving population estimates. The survey was administered using a computer assisted telephone survey instrument. The data set gathers information from 354 farms including: farm operator characteristics, land ownership, cropping patterns, livestock numbers, farm finance, farm topography, BMPs used in the farm, acreage served by each practice, last year at which cost sharing was received, and type and distance to the closest water body.

2.2 The econometric model.

In the empirical part of this study I analyze the impact of cost-share payments on the adoption of three soil conservation practices. Simultaneously, I estimate the effects that adoption of these practices has on the scope and intensity of cropping operations and on grassland and permanent vegetative cover. I consider two soil-erosion reducing practices for the analysis: reduced tillage and grass- or rock-lined waterways. Additionally I include filter strips and riparian buffers, which are treated as a single practice. They are described briefly in Table 3.1.

Table 3.1. Description of conservation practices

Practice	Description
Filter strips/riparian buffers	Filter or buffer strips are land areas of either planted or indigenous vegetation, situated between a potential pollutant-source area and a surface-water body that receives runoff. A properly designed filter strip provides water-quality protection by reducing the amount of sediment, organic matter, and some nutrients and pesticides in the runoff before they enter the surface-water body.
Reduced tillage	I include here no-till and minimum till. In both cases, the surface is covered by crop residue after planting to reduce soil erosion by water, or where soil erosion by wind is the primary concern. In no-till systems, planting is the only operation which disturbs the soil. Minimum till allows some tillage to solve weed problems and to deal with high moisture and heavy clay soil conditions.
Grass- or rock-lined waterways	Grading natural drainage ways to form a smooth channel, then planting with grass or lining with rock in order to protect the drainage way from gully erosion and trap sediment running off a field.

These three practices were chosen because they are among the most frequently cost-shared and thus they provide more representative treatment groups. Program participation is low in Maryland, but it is probably greater than in many other states because of Maryland's state programs aimed at the protection of the Chesapeake Bay (Lichtenberg and Smith-Ramírez, 2004). None of the practices surveyed had been implemented in more than the 45% of the farms at the time of the survey, and, for most practices, fewer than 10% of the farms that had implemented those practices had received some funding (see Table 3.2).

It can be noted from Table 3.2 that use of filter strips and riparian buffers were asked as separate questions in the survey, but appear as a single practice in Table 3.1. Filter strips and riparian buffers are used essentially for the same purpose: filtering pollutants from soil runoff before it reaches a surface water body. Often the terms are used interchangeably although the term filter strips usually refers to strips planted with grass while the term riparian buffers usually refers to strips planted with brush or trees. Given the similarity in their functions, I have aggregated the participation data of both practices in order to increase the size of the treatment group.

Table 3.2. Use of conservation practices and participation in cost sharing programs

Conservation Practice	Proportion of Farmers		
	using the practice with CS	using the practice without CS	not using the practice
Critical area seeding	0.013	0.270	0.717
Filter strips	0.032	0.300	0.668
Riparian buffer(s)	0.009	0.190	0.801
Contour farming	0.014	0.200	0.786
Strip cropping	0.005	0.270	0.725
Minimum till or no till	0.027	0.450	0.523
Grade stabilization	0.002	0.150	0.848
Grass- or rock-lined waterways	0.076	0.220	0.704
Terraces	0.002	0.050	0.948
Diversions	0.019	0.090	0.891
Sediment troughs	0.003	0.060	0.937
Permanent vegetative cover	0.008	0.310	0.682
Wildlife habitat	0.025	0.280	0.695

The farmer decision process is multivariate in nature and the econometric model must acknowledge the interdependency between production and conservation decisions as much as possible. Estimating the effect of conservation technology adoption on acreage cropped and acreage in grassland or under permanent vegetative cover involves the modeling of the following decisions: 1) whether to participate or not in a cost sharing payment program, 2) the size of the cropping operation, 3) acreage allocated to pastures, hayland and permanent vegetative cover, and 4) whether to use each conservation technology and how much land to use it on.

Consistent estimation of the effects that this study attempts to measure requires modeling these decisions jointly. Therefore, the following model is proposed

$$
\begin{array}{ll}
y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} & \text{Cost sharing equation} \\
y_{2i}^{*} = X_{2i}\beta_2 + y_{1i}X_{2i}\eta_2 + \varepsilon_{2i} & \text{Cropping intensity equation} \\
y_{3i}^{*} = X_{3i}\beta_3 + y_{1i}X_{3i}\eta_3 + \varepsilon_{3i} & \text{Vegetative cover equation} \\
y_{4i}^{*} = X_{4i}\beta_4 + y_{1i}X_{4i}\eta_4 + \varepsilon_{4i} & \text{Filter strips equation} \\
y_{5i}^{*} = X_{5i}\beta_5 + y_{1i}X_{5i}\eta_5 + \varepsilon_{5i} & \text{Minimum/no tillage equation} \\
y_{6i}^{*} = X_{6i}\beta_6 + y_{1i}X_{6i}\eta_6 + \varepsilon_{6i} & \text{Grass or rock-lined waterways equation}
\end{array}
\qquad (3.1)
$$

where the left-hand side variables are defined as follows.

y_{1i}^{*} is a latent variable giving the amount of cost sharing received by farm i. It signals farmer's and agency's preferences, and it can be seen as the equilibrium between farmer's willingness to apply for cost sharing (cost-sharing demand) and agency's willingness to provided it (cost-sharing supply). Only a binary variable, y_{1i}, is observed, taking the value 1 if cost sharing has been granted for the implementation of at least one of the practices included in Table 3.1 of for planting or preserving pastures, hayland or permanent vegetative covers, and 0 if not, i.e.

$$
y_{1i} = \begin{cases} 0 & \text{if } y_{1i}^{*} \le 0 \\ y_{1i}^{*} & \text{if } y_{1i}^{*} > 0 \end{cases}
$$

y^*_{2i} is a latent variable measuring the total acreage cropped in the farm as a proportion of total acreage operated. Acreage used to construct this variable includes land used in corn, soybeans, small grains, vegetables, and tobacco. It was also included the acreage under the category "other crops", which were not specified in the survey. In order to represent the actual size of cropping operations and following the discussion about intensive and extensive effects in the introduction, double-cropped acreage is summed twice. Thus, the observed counterpart of y^*_{2i} is given by (I have dropped the subindex i here) $y_2 = \left(A_s + 2A_d\right)/A$, which range from 0 to 2. The minimum value is obtained in farms with no cropping operations, while the maximum is attained when all the land operated is double cropped. The desired level of cropping y^*_{2i} is modeled here as a latent variable whose observed counterpart is censored from below at zero and from above at 2.

$$y_{2i} = \begin{cases} 0 & \text{if } \; y^*_{2i} \le 0 \\ y^*_{2i} & \text{if } \; 0 < y^*_{2i} < 2 \\ 2 & \text{if } \; y^*_{2i} \ge 2 \end{cases}$$

Variables y^*_{3i}, y^*_{4i}, y^*_{5i}, and y^*_{6i} are the levels at which vegetative cover (including pastures, hayland and permanent vegetative cover), filter or buffer strips, reduced tillage, and grass- or rock-lined waterways are used in the farm, respectively. Their observed counterparts y_{3i}, y_{4i}, y_{5i} and y_{6i}, are estimated by the ratio between the acreage served by the respective practice and total acreage operated. Thus, they are censored from below at zero and from above at 1. Summarizing:

$$y_{ji} = \begin{cases} 0 \quad \text{if} \;\; y_{ji}^* \leq 0 \\ y_{ji}^* \quad \text{if} \;\; 0 < y_{ji}^* < 1 \\ 1 \quad \text{if} \;\; y_{ji}^* \geq 1 \end{cases} \qquad\qquad j = 3,4,5,6$$

Vectors X_{kj} $k = 1,...,6$ in (3.1) represent exogenous explanatory variables whose descriptions are presented in Table 3.3. Coefficients β_k are parameter vectors related to exogenous regressors, η_k are vectors of parameters taking into account the effect of possible interactions between cost-share funding and exogenous covariates, and ε_{kj} are normally distributed error terms with zero means and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \sigma_{34} & \sigma_{35} & \sigma_{36} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_4^2 & \sigma_{45} & \sigma_{46} \\ \sigma_{15} & \sigma_{25} & \sigma_{35} & \sigma_{45} & \sigma_5^2 & \sigma_{56} \\ \sigma_{16} & \sigma_{26} & \sigma_{36} & \sigma_{46} & \sigma_{56} & \sigma_{66} \end{bmatrix} \qquad (3.2)$$

where σ_1^2 has been set equal to one following the usual normalization required to identify the parameters in equations that involve dichotomous dependent variables.

A sequential process is implicit in this model. In the first stage farmers apply for cost sharing and the administrative body of the program decides which projects to fund. This process is modeled by the first equation in the equation system (3.1). In the second stage farmers decide about conservation actions and production operations conditional on cost sharing received.

Table 3.3 describes the exogenous control variables X_{ji}. Farmer characteristics like age (both current and at the last year of receiving cost-share funding), farming experience, and education were included. Farmer age is used as a measure of farmer's time horizon and is expected to have a negative impact on the adoption of conservation technology. Experience and education, on the other hand, are expected to influence adoption positively as more experienced and educated farmers should have a better appreciation of private and social benefits of conservation technology.

Farm topography is considered by introducing two variables HIGH and MODERATE. HIGH is the percentage of land operated with a slope equal to or higher than 8%. MODERATE is the percentage of land operated with a slope higher than 2% but lower than 8%. Due to its impact on soil erodability, topography should be an important determinant of both the probability of receiving a cost-share award and farmers' decisions regarding adoption of the three BMPs in Table 3.1.

Tenancy is incorporated as the percentage of land operated that is rented in. Farmers are widely believed to have less incentive to invest in conservation on rented land since long run returns from conservation accrue to the landlord, not the tenant. According to this perception, farmers who rent a larger share of the land they operate should be less willing to apply for cost sharing and should invest in conservation projects that are smaller in size and scope. However, it can be argued that some private benefits from conservation technologies accrue in the short run (Soule et al., 2000). As mentioned in the introduction of this chapter, conservation tillage reduces time and labor cost since no plowing is necessary before seeding. Many conservation practices are land-quality enhancing, which means that they may increase profitability of marginal land. Also, since machinery used in conservation tillage requires high investments and maintenance costs, farmers may rent land in to increase the size of their cropping operations and thus reduce fixed costs per acre. Additionally, practices like contour farming and strip cropping help to improve crop yields by preventing that soil and fertilizer wash away in highly sloped areas that receive intense rain during the crop season. Consequently, if use of the technologies included in Table 3.1 is profitable in the short run, it is not that clear whether the effect of tenancy on adoption is positive or negative.

Total acreage operated was included to control for the effects of farm size. Large farms may have more incentives to apply for cost sharing since they are likely to have a more diverse topography. Program administrators, on the other hand, may target preferentially those farms that are likely to be large pollution sources such as large cropping or cattle operations. Farm size may also influence adoption decisions. For instance, conservation technologies may exhibit economies of scale, as might be the case with reduced tillage since larger farms can save proportionally more time in machinery operations and can spread the fixed costs of machinery over more acreage.

Table 3.3. Dependent and exogenous explanatory variables

Variable Name	Description	Mean	Standard Deviation
Cost share	Binary variable indicating whether the farmer has received cost sharing between 1995 and 1998 inclusively for implementation of at least one of the practices in Table 3.1 or for planting or preserving hayland, pastures or other vegetative cover (yes=1).	0.121	0.327
Acreage cropped proportion	Ratio of land planted to crops to land operated. Double cropped acreage is counted twice.	0.335	0.366
Vegetative cover share	Proportion of land under hayland, pasture or permanent vegetative cover.	0.439	0.351
Filter/buffer share	Proportion of land served by filter strips and/or riparian buffers	0.100	0.298
Reduced tillage share	Proportion of land cropped using minimum or no tillage techniques	0.203	0.306
Grass/rock waterways share	Proportion of land served by grass or rock lined waterways	0.039	0.150
CSAge[1]	Age of the farmer in the most recent year cost share funding was received since 1995	59.907	12.295
CSExperience[1]	Years of experience as farm operator at the most recent year cost share funding was received since 1995	28.297	14.224
Age[2,4,5,6]	Current farmer age	60.032	12.259
Experience[2,4,5,6]	Current years of experience as farm operator	28.422	14.222
College[1,2,4,5,6]	Farmer has college education or higher or has attended to technical school (yes = 1)	0.357	0.480
Income[1,2,4,5,6]	Proportion of total farmer income obtained from farm operations	0.453	0.399
High[1,2,4,5,6]	Proportion of highly sloped land in the total acreage operated (slope > 8%)	0.075	0.166
Moderate[1,2,4,5,6]	Proportion of moderately sloped land in the total acreage operated (slope 2-8%)	0.304	0.343
Rented[1,2,4,5,6]	Proportion of total land operated that is rented in	0.170	0.298
Land[1,2,4,5,6]	Total acreage operated (acres)	182.73	278.24
Cattle[3]	Number of grazing animals in the farm including beef cattle, dairy cattle, sheep, and horses (animal units)	38.643	76.736
Distance[1,4]	Distance to the nearest water body (miles)	0.778	3.119
Dairy[2,5]	The farm has more than 5 animal units of dairy cattle (yes = 1)	0.132	0.115
NDCattle[2,5]	The farm has more than 5 animal units of non-dairy cattle (yes = 1)	0.492	0.250

Poultry[2,5]	The farm has more than 5 animal units of poultry (yes = 1)	0.151	0.128

[1] Included as regressor in equation 1; [2] Included as regressor in equation 2; [3] Included as regressor in equation 3; [4] Included as regressor in equation 4, [5] Included as regressor in equation 5, [6] Included as regressor in equation 6. N=354

A continuous variable giving the distance to the closest water body (stream, lake, pond, wetland, or the Chesapeake Bay) was included. Since protection of water quality in the Bay and its tributaries is the expressed top priority of Maryland's conservation programs, it is expected that proximity to water bodies increases the likelihood to receive funding. Finally, a set of three operation type dummies (Dairy, NDCattle, and Poultry) was included to control for preferences of operations other than cropping. Dairy takes the value one if there is a dairy operation on the farm whose size exceeds five animal units[6], and takes the value zero otherwise. NDCattle takes the value one if the aggregated size of non-dairy cattle operations (beef cattle, swine operations, horses, sheep and lambs) exceeds five animal units, and takes the value zero otherwise. Poultry takes the value one if the there is a poultry operation on the farm whose size exceeds five animal units, and takes the value zero otherwise.

Before proceeding to the model estimation, I discuss the procedure to calculate the effect of program participation on the level of adoption of the different practices involved in this study. I also present the method to estimate the extensive and intensive effects of cost sharing on acreage cropped.

[6] I used the Animal Units Calculation Worksheet of the State of Wisconsin Department of Natural resources to calculate the number of animal units on each farm. The worksheet is available from http://www.dnr.state.wi.us/org/water/wm/nps/pdf/ag/cafo/form340025a.pdf.

2.3 Estimating the effects

2.3.1 Effect of cost sharing on adoption level

First, I present the procedure used to estimate the effects of cost sharing payments on the adoption level of the different practices. I evaluate these effects by estimating the change in the share of land served by each practice $j = 3,4,5,6$. I distinguish three cases:

1) Farmers who have implemented practice j and have received cost sharing. If the current observed share of land covered by practice j in farm i is y_{ji}, then the effect of cost sharing is:

$$\begin{aligned} \Delta S_{ji} &= y_{ji} - E\left[y_{ji} \mid y_{1i} = 0 \right] \\ &= y_{ji} - \Pr\left[0 < y_{ji} < 1 \mid y_{1i} = 0 \right] E\left[y_{ji} \mid 0 < y_{ji} < 1, y_{1i} = 0 \right] - \Pr\left[y_{ji} = 1 \mid y_{1i} = 0 \right] \end{aligned} \quad (3.3)$$

2) Farmers who have implemented practice j and have received no cost sharing. If the current observed share of land covered by practice j in farm i is y_{ji}, then the effect of cost sharing is:

$$\begin{aligned} \Delta S_{ji} &= E\left[y_{ji} \mid y_{1i} = 1 \right] - y_{ji} \\ &= \left(1 - \Pr\left[y_{ji} = 1 \mid y_{1i} = 1 \right] \right) E\left[y_{ji} \mid 0 < y_{ji} < 1, y_{1i} = 1 \right] + \Pr\left[y_{ji} = 1 \mid y_{1i} = 1 \right] - y_{ji} \end{aligned} \quad (3.4)$$

where the first term in (3.4) incorporates the assumption that a farmer that has been using practice j without cost sharing will not stop using the practice as consequence of being awarded funding

3) Farmers who have not implemented practice j. In this case, the effect of cost sharing is:

$$\begin{aligned}\Delta S_{ji} &= E\left[y_{ji} \mid y_{1i}=1\right]-0 \\ &= \Pr\left[0<y_{ji}<1 \mid y_{1i}=1\right]E\left[y_{ji} \mid 0<y_{ji}<1, y_{1i}=1\right]+\Pr\left[y_{ji}=1 \mid y_{1i}=1\right]\end{aligned} \tag{3.5}$$

Probabilities $\Pr\left[y_{ji}=1 \mid y_{1i}=1\right]$ and $\Pr\left[y_{ji}=1 \mid y_{1i}=0\right]$ in (3.4) can be calculated by $\dfrac{\Phi_2\left(X_{1i}\beta_1, \dfrac{X_{ji}\left(\eta_j+\beta_j\right)-1}{\sigma_j}, \rho_{\varepsilon_1\varepsilon_j}\right)}{\Phi_1\left(X_{1i}\beta_1\right)}$ and $\dfrac{\Phi_2\left(-X_{1i}\beta_1, \dfrac{X_{ji}\beta_j-1}{\sigma_j}, -\rho_{\varepsilon_1\varepsilon_j}\right)}{\Phi_1\left(-X_{1i}\beta_1\right)}$ respectively, where $\rho_{\varepsilon_j\varepsilon_k}$ is the correlation between ε_j and ε_k. The calculation of the other probabilities and expectations in (3.4) is more complicated. However, following the work of Rosenbaum (1961) it is possible to write

$$\Pr\left[0<y_{ji}<1 \mid y_{1i}=1\right]E\left[y_{ji} \mid 0<y_{ji}<1, y_{1i}=1\right]=$$

$$=\frac{\Phi_2\left(X_{1i}\beta_1, \dfrac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j}, \rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1, \dfrac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j}, \rho_{1j}\right)}{\Phi\left(X_{1i}\beta_1\right)}\Big\{X_{ji}\left(\eta_j+\beta_j\right)+$$

$$E\left[\varepsilon_{ji} \mid -X_{ji}\left(\eta_j+\beta_j\right)_j<\varepsilon_{ji}<1-X_{ji}\left(\eta_j+\beta_j\right)_j, \varepsilon_{1i}>-X_{1i}\beta_1\right]\Big\}$$

$$
=\frac{\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j},\rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j},\rho_{1j}\right)}{\Phi\left(X_{1i}\beta_1\right)}\Big\{X_{ji}\theta_j+
$$

$$
\frac{\phi\left(\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j}\right)}{\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j},\rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j},\rho_{1j}\right)}\Phi\left(-\frac{X_{1i}\beta_1-\rho_{1j}\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j}}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)+
$$

$$
\rho_{1j}\frac{\phi\left(X_{1i}\beta_1\right)}{\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j},\rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j},\rho_{1j}\right)}\Phi\left(-\frac{\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j}-\rho_{1j}X_{1i}\beta_1}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)-
$$

$$
\frac{\phi\left(\frac{X_{ji}\left(\eta_j+\beta_j\right)-1}{\sigma_j}\right)}{\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j},\rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j},\rho_{1j}\right)}\Phi\left(-\frac{X_{1i}\beta_1-\rho_{1j}\frac{X_{ji}\left(\eta_j+\beta_j\right)-1}{\sigma_j}}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)- \qquad (3.6)
$$

$$
\rho_{1j}\frac{\phi\left(X_{1i}\beta_1\right)}{\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j}{\sigma_j},\rho_{1j}\right)-\Phi_2\left(X_{1i}\beta_1,\frac{X_{ji}\left(\eta_j+\beta_j\right)_j-1}{\sigma_j},\rho_{1j}\right)}\Phi\left(-\frac{\frac{X_{ji}\left(\eta_j+\beta_j\right)-1}{\sigma_j}-\rho_{1j}X_{1i}\beta_1}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)\Bigg\}
$$

and

$$
\Pr\left[y_{ji}>0\mid y_{1i}=0\right]E\left[y_{ji}\mid y_{ji}>0,y_{1i}=0\right]=
$$

$$
=\frac{\Phi_2\left(\frac{X_{ji}\beta_j}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)-\Phi_2\left(\frac{X_{ji}\beta_j-1}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)}{\Phi\left(-X_{1i}\beta_1\right)}\Big\{X_{ji}\beta_j+
$$

$$
\frac{\phi\left(\frac{X_{ji}\beta_j}{\sigma_j}\right)}{\Phi_2\left(\frac{X_{ji}\beta_j}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)-\Phi_2\left(\frac{X_{ji}\beta_j-1}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)}\Phi\left(\frac{X_{1i}\beta_1-\rho_{1j}\frac{X_{ji}\beta_j}{\sigma_j}}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)-
$$

$$\rho_{1j}\frac{\phi(X_{1i}\beta_1)}{\Phi_2\left(\frac{X_{ji}\beta_j}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)-\Phi_2\left(\frac{X_{ji}\beta_j-1}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)}\Phi\left(-\frac{\frac{X_{ji}\beta_j}{\sigma_j}-\rho_{1j}X_{1i}\beta_1}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)-$$

$$\frac{\phi\left(\frac{X_{ji}\beta_j-1}{\sigma_j}\right)}{\Phi_2\left(\frac{X_{ji}\beta_j}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)-\Phi_2\left(\frac{X_{ji}\beta_j-1}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)}\Phi\left(-\frac{X_{1i}\beta_1-\rho_{1j}\frac{X_{ji}\beta_j-1}{\sigma_j}}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)+ \quad (3.7)$$

$$\left.\rho_{1j}\frac{\phi(X_{1i}\beta_1)}{\Phi_2\left(\frac{X_{ji}\beta_j}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)-\Phi_2\left(\frac{X_{ji}\beta_j-1}{\sigma_j},-X_{1i}\beta_1,-\rho_{1j}\right)}\Phi\left(-\frac{\frac{X_{ji}\beta_j-1}{\sigma_j}-\rho_{1j}X_{1i}\beta_1}{\left(1-\rho_{1j}^2\right)^{1/2}}\right)\right\}$$

To estimate the effect of cost-share funding on the share of land in vegetative cover we can use expressions (3.3) through (3.5) with $j = 3$. Whereas the net effect of cost sharing on cropping can be estimated by using the same formulas with $j = 2$.

The bivariate and trivariate normal distribution functions contained in expressions (3.3) through (3.7) were programmed in Matlab. Sample estimates of the effects were calculated by estimating the individual effects and then calculating the (weighted) mean. Finally, standard errors of these effects were estimated by the delta method.

2.3.2 Extensive and intensive margin effects

The estimation of the extensive and intensive margin effects of cost sharing needs some further elaboration. Recalling the notation used in the introduction, let A be total acreage operated, A_s the single-cropped acreage, A_d the double-cropped acreage, A_v the acreage in vegetative cover, and A_o acreage under other use (e.g.

machinery storage, livestock facilities, housing, roads). Thus, we have $A = A_s + A_d + A_v + A_o$, while the proportion of acreage cropped to total acreage, in the way defined in this article, is given by

$$y_2 = \frac{A_s + 2A_d}{A} = \frac{A_s + A_d}{A} + \frac{A_d}{A} = \frac{A_c}{A} + \frac{A_d}{A}$$

where A_c / A is the share of the farm used to plant crops.

Consider now the following assumptions:

Assumption 1: *The change in acreage used in machinery storage, livestock facilities, housing or roads, ΔA_0, due to cost sharing funding is negligible*. This is not a restrictive assumption since cost sharing provided for implementing practices in Table 3.1 or for planting or preserving hayland, pastures or other vegetative covers should not be used in building, housing or roads. It could be argued, on the other hand, that the construction of machinery storage is a possible indirect effect of cost sharing the adoption of minimum tillage. However, this effect (if existing) is likely to be very small compared with the direct effects affects ΔA_s and ΔA_d (a farmer willing to invest in machinery and storage facilities is expected to use that machinery).

Assumption 2: *Total acreage operated, A, does not change as consequence of cost-sharing funding*. This is a more debatable assumption, mainly because practices like minimum tillage may incentive farmers to rent land in to reduce financial and maintenance cost of the machinery and take advantage of reductions in cropping time and labor costs.

If the two assumptions above hold, then changes in A_c / A and y_2 caused by cost sharing funding can be written as follows

$$\frac{\Delta A_c}{A} = \frac{\Delta A}{A} - \frac{\Delta A_{\mathrm{v}}}{A} - \frac{\Delta A_{\mathrm{o}}}{A} \cong -\frac{\Delta A_{\mathrm{v}}}{A} \tag{3.8}$$

$$\begin{aligned} \Delta y_2 = \frac{\Delta A_c}{A} + \frac{\Delta A_d}{A} &= \frac{\Delta\left(A - A_{\mathrm{v}} - A_{\mathrm{o}}\right)}{\mathrm{A}} + \frac{\Delta A_d}{A} \\ &\cong -\frac{\Delta A_{\mathrm{v}}}{A} + \frac{\Delta A_d}{A} = \frac{\Delta A_c}{A} + \frac{\Delta A_d}{A} \end{aligned} \tag{3.9}$$

Thus, the change in the share of land used for cropping resulting from cost sharing (extensive margin effect) can be estimated by the negative of the change in the share of land used in vegetative cover. Complementarily, the change in the share of land under double cropping (intensive margin effect) can be estimated by the sum of the change in the acreage cropped share Δy_2 and the change in the share of land under vegetative cover $\Delta A_{\mathrm{v}} / A$. Alternatively, the intensive margin effect can be estimated by the change Δy_2 minus the extensive margin effect $\Delta A_c / A$.

One of the two possible extensive margin effects of adopting conservation technology is moving farmers from having only grassland (hayland and pastures) or other vegetative covers to start cropping activities. The other extensive margin effect is the expansion or reduction in cropping on farms that already practiced cropping before adopting the technology. While equation (3.8) provides a way to estimate the overall extensive margin effect, the analysis of the first component mentioned above is also desirable because it gives us the possibility of studying how conservation cost-sharing programs might change farming patterns in places were cropping was absent previously.

The all three practices included in this study can be implemented either on grassland or on cropland. According to the survey data, 24% of the farmers using buffer/filter strips do not crop, 8% of the people using minimum tillage do not crop, and 30% of the people using rock- or grass-lined waterways practice no cropping. Since these people do not crop, they must use these practices on grassland (hayland and/or pastures). Grassland is associated frequently to land with limitations (e.g. high erodibility) that prevent a more intensive use of the soil. Minimum tillage and waterways protection, however, makes land less susceptible to erosion, which might incentive farmers to move part of their land from grassland to cropland.

To estimate the effect of conservation technology on the probability of moving from no-cropping to cropping operations, I calculate the difference between the probability of growing crops having adopted technology j and the probability of growing crops without implementing that technology. Mathematically

$$\Delta \Pr_{cropping} = \Pr\left[y_{2i} > 0 \mid y_{1i} = 0, y_{ji} > 0 \right] - \Pr\left[y_{2i} > 0 \mid y_{1i} = 0, y_{ji} = 0 \right]$$

$$= \frac{\Phi_3\left(-X_{1i}\beta_1, \frac{X_{2i}\beta_j}{\sigma_2}, \frac{X_{ji}\beta_j}{\sigma_j}, -\rho_{\varepsilon_1\varepsilon_2}, -\rho_{\varepsilon_1\varepsilon_j}, \rho_{\varepsilon_2\varepsilon_j} \right)}{\Phi_2\left(-X_{1i}\beta_1, \frac{X_{ji}\beta_j}{\sigma_j}, -\rho_{\varepsilon_1\varepsilon_j} \right)} - \frac{\Phi_3\left(-X_{1i}\beta_1, \frac{X_{2i}\beta_j}{\sigma_2}, -\frac{X_{ji}\beta_j}{\sigma_j}, -\rho_{\varepsilon_1\varepsilon_2}, \rho_{\varepsilon_1\varepsilon_j}, -\rho_{\varepsilon_2\varepsilon_j} \right)}{\Phi_2\left(-X_{1i}\beta_1, -\frac{X_{ji}\beta_j}{\sigma_j}, \rho_{\varepsilon_1\varepsilon_j} \right)} \quad j = 4,5,6 \qquad (3.10)$$

In order to discuss the pure effect of the technology, I estimate this effect clean of cost sharing influence (i.e. conditional on $y_{1i} = 0$). Thus, the effect will be positive if the adoption of conservation technology that makes farming less dependent of land quality creates incentives for practicing cropping. A positive effect could open the possibility for an unintended effect of cost sharing programs: they might be (inadvertently) promoting cropping in farms where cropping was not practiced before.

2.4 Estimation by the Monte Carlo EM (MCEM) algorithm

The number of equations involved and the presence of latent variables complicates the estimation of the equation system (3.1) by maximum likelihood. The observed counterparts of the dependent variables of all the six equations are either dichotomous or censored variables, which introduces 6-dimensional integrals in the likelihood function. In order to avoid the estimation problems associated to traditional methods, I use here the MCEM algorithm formulated in Chapter 2.

The complete information likelihood function for our problem is standard and can be written as follows

$$L^c\left(\boldsymbol{\eta},\boldsymbol{\beta},\Sigma \mid \boldsymbol{y}\right)=\prod_i\left[\frac{1}{\left(2\pi\right)^3\left|\Sigma\right|^{1/2}}\exp\left(-\frac{\boldsymbol{\varepsilon}_i^{'}\Sigma^{-1}\boldsymbol{\varepsilon}_i}{2}\right)\right]$$

where $\boldsymbol{\theta}=\begin{pmatrix}\beta_1 & \beta_2 & \eta_3 & \beta_3 & \eta_4 & \beta_4 & \eta_5 & \beta_5 & \eta_6 & \beta_6\end{pmatrix}^{'}$, and $\boldsymbol{\varepsilon}_i=\begin{pmatrix}\varepsilon_{1i} & \varepsilon_{2i} & \varepsilon_{3i} & \varepsilon_{4i} & \varepsilon_{5i} & \varepsilon_{6i}\end{pmatrix}^{'}$

Thus, the complete data log-likelihood function is

$$\ell^c\left(\boldsymbol{\theta},\Sigma \mid \boldsymbol{y}\right)=-N\left(3\ln\left(2\pi\right)+\frac{1}{2}\ln\left|\Sigma\right|\right)-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_i\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right) \qquad (3.11)$$

E-Step. The expectation of expression (3.11) can be written as

$$E\left[\ell^c\left(\boldsymbol{\theta},\Sigma \mid \boldsymbol{y}\right)\right]=-N\left(3\ln\left(2\pi\right)+\frac{1}{2}\ln\left|\Sigma\right|\right)-\frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\sum_i E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right]\right) \quad (3.12)$$

where the expectation is conditional on observed information and distribution assumptions. The E-step at iteration $m+1$, requires the calculation of

$$Q_i\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)=E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'} \mid \boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]$$

$$= \sigma_i^{2(m)} + \begin{pmatrix} \mu_{y_{1i}^*}^{(m)} - X_{1i}\beta_1 \\ \mu_{y_{2i}^*}^{(m)} - X_{2i}\beta_2 \\ \mu_{y_{3i}^*}^{(m)} - X_{3i}y_{1i}\eta_3 - X_{3i}\beta_3 \\ \mu_{y_{4i}^*}^{(m)} - X_{4i}y_{1i}\eta_4 - X_{4i}\beta_4 \\ \mu_{y_{5i}^*}^{(m)} - X_{5i}y_{1i}\eta_5 - X_{5i}\beta_5 \\ \mu_{y_{6i}^*}^{(m)} - X_{6i}y_{1i}\eta_6 - X_{6i}\beta_6 \end{pmatrix} \begin{pmatrix} \mu_{y_{1i}^*}^{(m)} - X_{1i}\beta_1 \\ \mu_{y_{2i}^*}^{(m)} - X_{2i}\beta_2 \\ \mu_{y_{3i}^*}^{(m)} - X_{3i}y_{1i}\eta_3 - X_{3i}\beta_3 \\ \mu_{y_{4i}^*}^{(m)} - X_{4i}y_{1i}\eta_4 - X_{4i}\beta_4 \\ \mu_{y_{5i}^*}^{(m)} - X_{5i}y_{1i}\eta_5 - X_{5i}\beta_5 \\ \mu_{y_{6i}^*}^{(m)} - X_{6i}y_{1i}\eta_6 - X_{6i}\beta_6 \end{pmatrix}' \qquad (3.13)$$

where $\sigma_i^{2(m)} = \mathrm{Cov}\left(y_{1i}^*, y_{2i}^*, y_{3i}^*, y_{4i}^*, y_{5i}^*, y_{6i}^* \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right)$ (3.14)

and $\mu_{y_{ji}^*}^{(m)} = E\left[y_{ji}^* \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right] \qquad j = 1, \ldots, 6$ (3.15)

M-Step. Following Meg and Rubin (1993), it is advisable to replace the M-step by two conditional M-steps. The first conditional M-step maximizes

$$-N\left(3\ln(2\pi) + \frac{1}{2}\ln|\Sigma|\right) - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_i Q_i\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y}\right)\right) \qquad (3.16)$$

with respect to the elements in $\boldsymbol{\theta}$ conditional on $\boldsymbol{\theta}^{(m)}$ and $\Sigma^{(m)}$. It is easy to see from (3.13) that the maximizer in this first conditional maximization can be written as a generalized least square estimator

$$\boldsymbol{\theta}^{(m+1)} = \left[\tilde{X}_d'\left(\Sigma^{(m)} \otimes I_N\right)^{-1}\tilde{X}_d\right]^{-1}\tilde{X}_d'\left(\Sigma^{(m)} \otimes I_N\right)^{-1}\mu_{y^*}^{(m)} \qquad (3.17)$$

where $\tilde{X}_d = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & \tilde{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \tilde{X}_6 \end{bmatrix}$, and $\tilde{X}_k = \left[X_k y_1 \,\vdots\, X_k \right] \quad k = 2,3,4,5,6$

$\mu_{y^*}^{(m)} = \left(\mu_{y_1^*}^{(m)} \quad \mu_{y_2^*}^{(m)} \quad \mu_{y_3^*}^{(m)} \quad \mu_{y_4^*}^{(m)} \quad \mu_{y_5^*}^{(m)} \quad \mu_{y_6^*}^{(m)} \right)'$, and

$$\mu_{y_j^*}^{(m)} = \left(\mu_{y_{j1}^*}^{(m)} \quad \mu_{y_{j2}^*}^{(m)} \quad \cdots \quad \mu_{y_{jN}^*}^{(m)} \right)'$$

The second conditional M-step estimates $\Sigma^{(m+1)}$ by maximizing the following function

$$-N\left(3\ln(2\pi) + \frac{1}{2}\ln|\Sigma| \right) - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} \sum_i Q_i \left(\boldsymbol{\theta}^{(m+1)} \mid \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}, \boldsymbol{y} \right) \right) \tag{3.18}$$

with respect to the elements of Σ conditional on $\boldsymbol{\theta}^{(m+1)}$.

There remains the implementation of the Gibbs sampler necessary to estimate the matrices Q_i in the objective function. Since the simulations for $\boldsymbol{y}_i^*$ must be done conditional on its corresponding observed information $\boldsymbol{y}_i$, the implementation procedure depends on the structure imposed by $\boldsymbol{y}_i$ on $\boldsymbol{y}_i^*$.

The observed counterpart of y_{1i}^* is dichotomous with y_{1i}^* being positive if y_{1i} equals one and non-positive if y_{1i} equals zero. Accordingly, we simulate y_{1i}^* from a normal distribution with mean $\mu_{1i|i(-1)}^{(m)}$ and variance $\sigma_{1i|-1}^{2(m)}$ truncated below at zero if y_{1i} equals one and truncated above at zero if y_{1i} equals zero.

Variable y_{2i}^{*} is censored from below at zero and from above at 2. Consequently, we simulate y_{2i}^{*} from a normal distribution with mean $\mu_{2i|i(-2)}^{(m)}$ and variance $\sigma_{2i|-2}^{2(m)}$ truncated above at zero when y_{2i} equals zero and truncated below at 2 when y_{2i} equals 2. If $0 < y_{2i} < 2$ then $y_{2i}^{*} = y_{2i}$. Variables y_{ji}^{*} $\left(j = 3,4,5,6\right)$ are all censored from below at 0 and from above at 1. Thus, the unobserved components must be simulated from normal distributions with means $\mu_{ji|i(-j)}^{(m)}$ and variances $\sigma_{ji|-j}^{2(m)}$ truncated above at zero if $y_{ji} = 0$ and from normal distributions truncated from below at 1 if $y_{ji} = 1$. If $0 < y_{ji} < 1$ set $y_{ji}^{*} = y_{ji}$.

In this study the Gibbs sampler was started with 600 simulations and increased by 15 simulations at every iteration of the MCEM algorithm. The number of dismissed simulations, k_{burn}, was kept constant at 150. A simulation of size $R = 3600$ and $r_{burn} = 600$ was used for the estimation of the information matrix and asymptotic standard errors. The full set of interactions between the cost-sharing variable y_{1i} and regressors in Table 3.3 was considered in an initial estimation. Only those interactions that were significant at 5% were kept in the model and a second estimation was carried out. Results of the second estimation are presented in Table 3.4.

3 Estimation results and discussion

Overall, signs of the coefficients in Table 3.4 match expectations. Several of then are not significant statistically, however. None of the interaction coefficients in η_2 was significant and all of them were dropped from the second equation in the last estimation. A Wald test for the null hypothesis of the interactions coefficients simultaneously equal to zero gave a result of 10.1356 (6 df), which does not allow rejecting H_0.

Table 3.4. Estimation results

Equation	Variable	Estimate	As. St. error	As. t-value	p-value
Cost-Sharing	Constant	-0.0872	0.5112	-0.1705	0.8648
	CSAge	-0.0238	0.0099	-2.4078	0.0168
	CSExperience	9.31E-04	8.24E-03	0.1129	0.9102
	Education	-0.1162	0.1923	-0.6043	0.5462
	Income	0.4503	0.2410	1.8683	0.0629
	High	0.8713	0.4357	1.9995	0.0466
	Moderate	0.0442	0.2560	0.1727	0.8630
	Rented	-0.2221	0.2937	-0.7564	0.4501
	Land	6.14E-04	2.28E-04	2.6941	0.0075
	Distance	-0.0563	0.0380	-1.4838	0.1391
Cropping intensity	Constant	0.6269	0.1395	4.4935	0.0000
	Age	-0.0073	0.0025	-2.9568	0.0034
	Experience	0.0022	0.0021	1.0381	0.3002
	Education	-0.0998	0.0494	-2.0225	0.0442
	Income	0.2890	0.0690	4.1892	0.0000
	High	0.1784	0.1355	1.3169	0.1891
	Moderate	-0.0418	0.0684	-0.6109	0.5418
	Rented	0.2930	0.0852	3.4401	0.0007
	Land	1.73E-04	8.15E-05	2.1296	0.0342
	Dairy	-0.0033	0.0698	-0.0479	0.9619
	NDCattle	-0.3260	0.0473	-6.8977	0.0000
	Poultry	0.0155	0.0969	0.1603	0.8728
Vegetative Cover	Constant	-0.2213	0.1335	-1.6581	0.0986
	Age	0.0096	0.0020	4.8346	0.0000
	Education	0.0237	0.0487	0.4861	0.6273
	Income	-0.1444	0.0642	-2.2494	0.0254
	High	-0.0963	0.1407	-0.6844	0.4944
	Moderate	0.4280	0.0670	6.3874	0.0000
	Rented	0.0233	0.0869	0.2681	0.7889
	Land	-4.04E-04	8.86E-05	-4.5563	0.0000
	Cattle	1.16E-03	2.54E-04	4.5632	0.0000
	y_{1i} x Constant	0.3830	0.3723	1.0288	0.3046

	y_{1i} x Age	-9.03E-04	5.65E-03	-0.1600	0.8730
	y_{1i} x Education	0.3722	0.1376	2.7048	0.0073
	y_{1i} x Income	-0.0630	0.1653	-0.3814	0.7032
	y_{1i} x High	0.4563	0.2920	1.5628	0.1194
	y_{1i} x Moderate	-0.6939	0.1898	-3.6566	0.0003
Filter Strips/Riparian Buffers	Constant	0.0292	0.2196	0.1328	0.8945
	Age	-0.0109	0.0038	-2.8918	0.0042
	Experience	0.0132	0.0030	4.4248	0.0000
	Education	0.1730	0.0756	2.2886	0.0229
	Income	-0.2413	0.1017	-2.3735	0.0184
	High	0.2157	0.2036	1.0595	0.2904
	Moderate	-0.1010	0.1077	-0.9378	0.3493
	Rented	0.1723	0.1299	1.3265	0.1859
	Land	-8.68E-05	1.19E-04	-0.7277	0.4675
	Distance	-0.0064	0.0117	-0.5466	0.5852
	y_{1i} x Constant	-1.3761	0.5509	-2.4979	0.0131
	y_{1i} x Age	0.0282	0.0080	3.5422	0.0005
	y_{1i} x Education	-0.1220	0.1907	-0.6399	0.5228
	y_{1i} x Income	0.7821	0.2358	3.3171	0.0010
	y_{1i} x High	-0.6306	0.4773	-1.3212	0.1877
	y_{1i} x Moderate	0.6113	0.2514	2.4313	0.0158
Reduced Tillage	Constant	0.7410	0.2101	3.5270	0.0005
	Age	-0.0200	0.0039	-5.0805	0.0000
	Experience	0.0118	0.0034	3.4559	0.0006
	Education	-0.2202	0.0790	-2.7886	0.0057
	Income	0.1879	0.1063	1.7671	0.0784
	High	-0.1684	0.2437	-0.6910	0.4902
	Moderate	0.3441	0.1038	3.3143	0.0011
	Rented	0.3592	0.1270	2.8289	0.0051
	Land	2.53E-04	1.18E-04	2.1361	0.0337
	Dairy	-0.0521	0.1004	-0.5192	0.6041
	NDCattle	-0.1655	0.0711	-2.3288	0.0207
	Poultry	-0.3225	0.1445	-2.2327	0.0265
	y_{1i} x Constant	-1.0767	0.5725	-1.8805	0.0612
	y_{1i} x Age	0.0125	0.0085	1.4803	0.1401
	y_{1i} x Education	0.2213	0.1928	1.1477	0.2522
	y_{1i} x Income	0.2037	0.2350	0.8667	0.3869
	y_{1i} x High	0.8791	0.4353	2.0194	0.0445
	y_{1i} x Moderate	-0.1982	0.2661	-0.7446	0.4572
Grass- or Rock-lined Waterways	Constant	0.1833	0.1591	1.1519	0.2505
	Age	-0.0142	0.0030	-4.7342	0.0000
	Experience	0.0098	0.0026	3.7601	0.0002
	Education	0.0218	0.0574	0.3795	0.7046
	Income	-0.1731	0.0751	-2.3028	0.0221
	High	0.1232	0.1745	0.7057	0.4811
	Moderate	0.3298	0.0759	4.3476	0.0000
	Rented	0.0166	0.0965	0.1722	0.8634
	Land	5.14E-05	8.15E-05	0.6310	0.5286
	y_{1i} x Constant	-0.8046	0.4606	-1.7470	0.0819
	y_{1i} x Age	0.0191	0.0068	2.7835	0.0058
	y_{1i} x Education	0.0672	0.1492	0.4505	0.6528
	y_{1i} x Income	0.2767	0.1854	1.4919	0.1370
	y_{1i} x High	-0.4409	0.3977	-1.1087	0.2686

	y_{1i} x Moderate	0.0487	0.2039	0.2390	0.8113
	σ_{12}	0.0352	0.0128	2.7615	0.0062
	σ_{13}	-0.2457	0.0107	-22.9442	0.0000
	σ_{14}	-0.4657	0.0104	-44.7536	0.0000
	σ_{15}	0.1948	0.0175	11.1297	0.0000
	σ_{16}	-0.1309	0.0113	-11.5416	0.0000
	σ_{22}	0.1525	0.0071	21.3699	0.0000
	σ_{23}	-0.0418	0.0051	-8.1903	0.0000
	σ_{24}	-0.0080	0.0070	-1.1495	0.2515
	σ_{25}	0.1183	0.0080	14.7478	0.0000
	σ_{26}	0.0147	0.0046	3.2105	0.0015
	σ_{33}	0.1558	0.0068	22.8854	0.0000
	σ_{34}	0.1127	0.0063	18.0363	0.0000
	σ_{35}	-0.0238	0.0067	-3.5440	0.0005
	σ_{36}	0.0463	0.0047	9.9103	0.0000
	σ_{44}	0.3018	0.0105	28.7160	0.0000
	σ_{45}	-0.0025	0.0091	-0.2745	0.7839
	σ_{46}	0.1228	0.0073	16.7785	0.0000
	σ_{55}	0.2920	0.0142	20.5548	0.0000
	σ_{56}	0.0523	0.0065	8.0199	0.0000
	σ_{66}	0.1207	0.0060	20.1927	0.0000

3.1 Determinants of cost sharing allocation

It is remarkable that, although the three practices involved in this study control soil and nutrient runoff and most cost sharing programs in Maryland have water quality protection as a stated priority, distance to water bodies is not a significant determinant of cost sharing allocation. Furthermore, only one of the variables controlling for land slope is statistically significant. More than focusing on environmental characteristics, program administrators seem to target preferentially large operations (the coefficient of LAND is significant at 1% in the cost-sharing equation). This outcome could be evidence of attempts to maximize the acreage served by the practices being cost shared. Alternatively it could be an indication of the existence of transaction costs restricting the participation of small farmers. For instance, time costs for doing the necessary paperwork can be significant for poor farmers or farmers with low managerial skills. In relation to this problem, Lynch and Tjaden (2004) observe that: “A participant usually has to take part in more than one program or to piggyback programs to cover all costs. A number of different agencies run these programs, with limited coordination, and each agency has different rules. Often, cost-share payments are not made at the same time monetary expenditures occur; landowners may have to wait for up to a year for reimbursement”.

The results also indicate that likelihood of receiving cost share declines with farmer age. Since it is unlikely that program administrators discriminate by farmer age, this outcome indicates probably that older farmers are less willing to go through the paperwork required to apply for cost sharing. Additionally, older farmers have shorter time horizons, which reduces their willingness to implement practices whose benefits accrue on the long run and thus they are less willing to apply for cost-sharing.

The common belief that renters are less willing to engage in conservation effort than owners are is not supported by results in Table 3.4. Since it is unlikely that program administrators target renters preferentially, the lack of statistical significance of the coefficient of RENTED indicates that renters apply for cost sharing as much as owners do[7] (for the practices included in this study at least).

3.2 Effects of adoption of conservation technology on cropping decisions

The changes in the probabilities of starting cropping activities as result of adopting conservation technologies are presented in Table 3.5. In order to capture the effects free of cost sharing influence, estimates were calculated conditional on no cost-share funding (see formula (3.10)) and only on the sub-sample of farms that do not participate in cost sharing programs. They are presented for different farm sizes and for the sub-sample.

The results in Table 3.5 are in close agreement with the discussion in Section 1. Adoption of both reduced tillage and grass- and rock-lined waterways increases the likelihood of growing crops. These two practices are land-quality enhancing. They alleviate the negative impact of cropping on steep land. Use of reduced tillage reduces soil runoff, while grass- or rock-lined waterways prevent gully erosion. Additionally, its use decreases time and labor costs. Results indicate that adoption of any of these two practices makes cropping profitable in farms where growing crops was not profitable before.

[7] It is interesting noting that the coefficients of the RENTED variable in the cropping intensity and reduced tillage equations are both positive and statistically significant meaning that renters crop more and use minimum or no tillage on more acreage than owners do. This result could be considered as an indication that renters use reduced tillage to expand cropping on marginal land. The coefficient of RENTED in the permanent vegetative cover equation is, however, neither negative nor statistically different from zero, which suggests that reduced tillage affects cropping on the intensive rather than the extensive margin.

Table 3.5. Estimated effects of conservation technology adoption on cropping

Practice	Farm size	Change in probability of practicing cropping	
		Estimate	**As. z-value[5]**
Filter strip - Riparian buffer	Small[1]	-0.0128	-0.7918
	Medium[2]	-0.0046	-0.4037
	Large[3]	-0.0012	-0.2586
	Sample[4] average	-0.0117	-0.7612
Reduced Tillage	Small	0.3211	18.7129
	Medium	0.2294	10.7719
	Large	0.1128	4.1011
	Sample average	0.3064	18.3026
Grass- or Rock-lined Waterways	Small	0.0693	3.5922
	Medium	0.0491	3.7365
	Large	0.0185	2.7712
	Sample average	0.0660	3.6157

[1] Small: farm size is smaller than or equal to 250 acres. [2] Medium: farm size is larger than 250 acres and smaller than or equal to 750 acres. [3] Large: farm size is larger than 750 acres. [4] It includes only individuals with $y_{1i} = 0$. [5] Asymptotic t-values; standard errors were estimated by the Delta method.

The effect of adoption of reduced tillage and grass- or rock-lined waterways on the decision about whether or not to crop is decreasing in farm size. This may happen because small farms are more land-quality constrained than larger farms. As a consequence of their size, medium and large farms have a wider distribution of land quality, which means they have a greater likelihood of including tracts of land on which raising crops is profitable. Thus, small farms are less likely to have crop operations than medium and large farms, all else equal. The adoption of land-quality enhancing practices should thus have a proportionately larger average effect on smaller farms than on larger ones.

The relation between adoption of filter strips/riparian buffers and cropping presents a different situation. Although the related effects in Table 3.5 are not significant statistically, their signs suggest that adoption of these practices is associated with a reduction in the probability of cropping (a negative extensive margin effect). Reasons explaining a non-positive effect of these practices can be found in the analysis of the benefits and costs that the implementation of a filter strip/riparian buffer reports to a farmer. Although buffers and filter strips have some private benefits such as decreasing soil erosion from adjacent fields, their main benefits are public goods such as improved water quality, fish and wildlife habitat, recreation services, and scenic amenities (Lynch and Tjaden, 2004). In contrast to reduced tillage and grass- and rock-lined waterways, buffers and filter strips are not land-quality enhancing practices because they do not make cropping less dependent on land quality and, consequently, they do not create incentives to expand cropping. Rather, buffers and filter strips may reduce cropping output by removing land from production and by attracting wildlife that may cause significant damage to crops. Trees in buffers may shade the field and reduce crop yield, while filter strips may become reservoirs of weeds and insect pests. Finally, riparian buffers and filter strips alter the configuration of the field making machinery maneuvering more difficult (Lynch and Tjaden, 2004).

This section has shown evidence about the mechanisms linking production and conservation decisions. The analysis indicates that adoption of conservation practices may expand or reduce cropping operations depending on whether or not the use of the practice provides private incentives for cropping. The next step is evaluating the effect of cost sharing payments on conservation decisions. This evaluation is critical to determine the green-box quality of cost sharing programs. Since it has been proved that adoption of certain practices can expand cropping, then a positive effect of cost sharing on adoption of these practices would be an indication that cost sharing payments may affect production levels and, consequently, they might influence trade. The analysis of these issues is carried out in the next sections.

3.3 Effect of cost-sharing payments on adoption of conservation practices and permanent vegetative cover

Differences in the expected shares of land served by different practices as result of cost-share funding are presented in Table 3.6.

Table 3.6. Estimated effects of cost-share payments on adoption of conservation practices

Practice	Farm size	Change in the share of land served	
		Estimate	As. z-value
Filter strip - Riparian buffer	Small	-0.0748	-44.7919
	Medium	0.0239	7.2978
	Large	0.0264	5.5538
	Sample average	-0.0599	-35.0453
Reduced Tillage	Small	0.1925	9.3245
	Medium	0.2327	14.9955
	Large	0.2546	8.7606
	Sample average	0.1993	10.4792
Grass- or Rock-lined Waterways	Small	0.0012	0.4445
	Medium	0.0407	15.8530
	Large	0.0649	7.4245
	Sample average	0.0080	3.0424
Vegetative cover	Small	-0.2227	-18.6897
	Medium	-0.0332	-4.4332
	Large	-0.0294	-0.2399
	Sample average	-0.1940	-16.4602

The empirical results confirm hypotheses proposed by Malik and Shoemaker (1993) and by Lichtenberg (2004), namely, cost sharing (1) increases the use of land quality augmenting conservation practices and (2) promotes cropping on lower quality land, resulting in reduced vegetative cover. These results also admit the following interpretation: cost sharing has a positive effect on the scope at which practices presenting private incentives for expanding cropping operations are used; however, participant farms use permanent vegetative cover, pastures, and hayland (i.e. practices that, besides private goods, have positive off-farm spillovers[8]), less intensively than non-participant farms do.

The evidence for filter strips and riparian buffers is mixed. Participating farms that are small in size implement these practices on a smaller share of land than farms that receive no funding, while medium and large participating farms implement more than non-participants. Following the discussion from the previous section, negative impacts of the implementation of a buffer on cropping should be greater when the area planted is a significant proportion of the total land operated (if only because of the higher opportunity cost of land). Therefore, negative effects on cropping should be more distinct on small than on large farms, which is in accordance with results in Table 3.6. This confirms that the opportunity costs of land are greater on small farms, a fact that should be taken into account when determining size and allocation of cost-sharing awards.

According to Table 3.3, a participant farm is a farm that has received funding for preservation of permanent vegetative cover or the implementation of at least one of the three practices in Table 3.1. Results in Table 3.6 suggest that, regardless of the

[8]Scenery and wildlife habitat, for instance. Additionally, pasture and permanent vegetative cover are biodiversity reservoirs, sequester carbon from the atmosphere (Lewandrowsky, 2004) and filter pollutans before they reach surface or underground water

bundle of practices that farmers have agreed to implement at the moment to apply, they will implement those practices that provide on-farm private benefits preferentially. From the previous analysis, we know that the use of reduced tillage and grass- or rock-lined waterways provides on-farm benefits and expand cropping at the extensive margin. If these are the kind of practices being implemented more frequently when cost share is provided, then the resulting expansion in cropping necessarily implies a reduction in the share of land under grassland, filter trips, riparian buffers, or other permanent vegetative covers. Thus, practices that supply off-farm benefits may not be implemented or, if existing on the farm, they could be "displaced" by practices that provide on-farm benefits. Some previous evidence of this perverse effect of cost sharing is provided by Cattaneo (2003), who reports that one or more practices were not implemented in 17% of early EQIP contracts. She suggests that limited enforcing capabilities incentive farmers to include in their applications a selection of BMP that maximize the probability to be awarded; yet, after cost sharing is provided, they implement only those practices in the contract that are profitable. Additional light about this issue is shed by Vukina et al. (2003), who use data from CRP auctions to study farmers' attitudes towards the environment. By analyzing the way farmers construct their bids, the authors conclude that "farmers value those environmental benefits which directly affect the productivity of their land but do not value those benefits which resemble public goods". Among the services they found as poorly valued by farmers are improvement of air quality and wildlife habitat, while they were unable to conclude about water quality improvements. This suggests that farmers do not consider themselves as providers of off-farm benefits.

3.4 Extensive and intensive margin effects of cost-sharing payments on cropping

The preceding sections have shown that (1) conservation technology can expand cropping and (2) cost sharing payments affect positively the adoption of conservation technology in general and the adoption of land-quality enhancing practices in particular. Here, I discuss the overall effects of cost sharing on acreage cropped.

Table 3.7. Estimated effects of cost-share payments on the share of acreage cropped

Farm size	Change in the proportion of acreage cropped Δy_2	
	Estimate	As. z-value
Small	0.0352	1.7746
Medium	-0.0099	-0.5086
Large	0.2125	5.4862
Sample average	0.0361	1.9192

According to equation (3.8), cost-sharing extensive-margin effect equals the negative of the change in vegetative cover. Thus, from Table 3.6, cost sharing has a positive extensive-margin effect on cropping. The effect on the land share cropped is particularly strong on small farms (an expansion of 22 percentage points), while it is smaller (less than 4 percentage points) on medium and large farms. Comparison between these results and the net effect of cost sharing on the share of land cropped (Table 3.7) sheds some light about the intensive margin effects of cost sharing payments. From Table 3.7 and contrasting with the 22 percentage-point expansion inferred from the reduction in vegetative cover, the net effect of payments on cropping in small farms is only 3.5 percentage points. According to equation (3.9), a necessary implication from these figures is that the land share being double cropped has been reduced in about 18.5 percentage points in participant small farms. This is additional evidence that expansion in scope is more profitable than cropping intensification on small farms, and the former will substitute for the latter when technologically feasible. The net effect of payments on cropping is negligible on medium farms, which suggests that the 3.3-points expansion in scope is canceled by an almost equivalent reduction in double-cropped land. Finally, large farms that participate in cost sharing programs crop a land share 21.2 points larger that non-participant farms. Since the same farms show an expansion on marginal land no greater than 3 percentage points, I conclude that the share of land being double cropped on participant farms is about 18 percentage points larger than on farms receiving no cost-share funding. These results suggest that positive intensive margin effects of cost sharing conservation technology increase with farm size. A plausible explanation for this outcome is that more intensive cropping operations require more capital and greater management time and skill. Thus, the impact of using conservation land-quality augmenting technology to intensify cropping is greater on larger farms because medium and large farmers are (on average) less capital constrained and have more time and managerial skill than small farmers

4. Final remarks

In this study I analyzed interactions between farmers' conservation and production decisions in a disaggregated multivariate framework. Several contributions to literature emerged from the analysis. First, I provided evidence that adoption of soil conservation practices, particularly land-quality augmenting ones, expands cropping both at the extensive and the intensive margins. Second, I showed that farms awarded cost share funding implement land-productivity improving practices on a share of land greater than farms that do not receive cost sharing do. Third, I showed the existence of a perverse effect of cost sharing payments: as farmers choose to implement preferentially those practices providing private benefits, the expansion in cropping induced by cost-sharing this type of practices reduces the share of land covered by practices that provide public goods as byproducts.

The analysis indicated that effects of cost sharing payments on cropping change with farm size. Positive extensive margin effects predominate on small farms, which replace grassland and other vegetative covers by cropland. Positive intensive margin effects are more likely on large farms, which show no significant extensive margin effects.

Cropping expansion at both the extensive and intensive margins resulting from the adoption of land-quality augmenting practices may imply additional nutrient runoff and more pesticide usage. The environmental balance of cost sharing these practices seems to be even more negative as results of the substitution of grassland and wildlife habitat by cropland (particularly on small farms). This study's findings question whether the current design of cost-sharing programs permits to achieve their environmental goals. A more stringent oversight seems to be crucial both during the formulation of cost-sharing programs and all along their implementation in order to minimize unintended effects. Additionally, cost-sharing payments for practices providing off-farm benefits may need to be increased in order to make them comparable to the opportunity costs of allocating land to them. Program administrators should consider more severe restrictions to cropping expansion and reconsider whether the adoption of land-quality augmenting practices needs to be cost-shared. It is likely that making not eligible some of the most profitable practices will harm program participation, and significant increment in the cost-sharing rates may be necessary to sustain or increase enrollment. Nonetheless, paying this price may be a need in order to achieve programs' stated goals.

The results obtained in this study raise questions about the "green-box" quality of payments that cost share adoption of land-quality enhancing practices on large farms. Results indicate that these payments have induced a significant increase in the share of land cropped (double cropping included) on large Maryland farms. Although similar effects showed no statistical significance on medium and small farms, they could become significant under the provisions of the new 2002 Farm Bill, which expands the budget of cost sharing programs dramatically.

Maryland, on the other hand, is a small state that contains a small proportion of US cropland and produces only a small share of US agricultural output. Therefore, similar studies using data from larger agricultural states are needed to validate the findings of this study.

Chapter 4: Incorporating Soil Nutrient Dynamics into the Evaluation of Soil Remediation Programs? Evidence from Chile

1. Introduction

Most of the literature dealing with the economics of soil degradation focuses on the problem of soil depth reduction. Soil depth is modeled as a capital stock whose dynamics are determined by farmers' actions and the natural regeneration rate of the soil (McConnell, 1983). Few studies analyze other aspects of soil degradation, such as the reduction of fertility, which may happen even when soil depth reduction is negligible. Additionally, plant nutrients are often included in economic analyses as a strictly variable input in the production function without consideration of nutrient pools. This approach is probably correct when soil nutrient dynamics have only marginal effects on plant nutrition or the prices of fertilizers are low relative to output prices, but these cases do not cover all the possible situations. At least two cases can be mentioned where the inclusion of the dynamics of the soil nutrient pools in the analysis is justified. The first one corresponds to the presence of significant carry-over effects (Kennedy, 1986; Schnitkey et al., 1996). The second case happens when soils act like sinks by fixing a significant proportion of the fertilizer applied by the farmer and thus preventing plants from using it.

This chapter discusses both theoretically and empirically some policy implications when soil nutrient dynamics are mistakenly ignored in the second case mentioned above. Specifically, the phosphorus fixation problem is used to illustrate how a program that subsidizes fertility replenishment can become inefficient if farmer characteristics and dynamics of the soil nutrient reservoirs are not considered when defining the targeting strategy.

Phosphorus (P) is not a limitation for crop production in most of the developed world. For instance, Sharpley et al. (1999) indicate that many states in the US show P-levels in their soils over the agronomic threshold[9] implying that little or no additional P fertilization is required for commercial production of most crops. On the contrary, P-deficit has become a major problem in developing countries as result of either natural low levels of phosphorus or depletion of once well stocked soil reserves (Buresh et al., 1997). Phosphorus fertility depletion occurs when no phosphate fertilization is provided to compensate for the phosphorus exported with products harvested in the farm (including beef or milk from permanently grazed pastures). The main reasons for soil fertility depletion are lack of financial resources to afford an adequate fertilization program or ineffectual or absent land property rights. In other cases the soil itself contributes to the problem. Depending on parent materials and weathering factors, the capacity of the soil to fix phosphorus can dramatically increase the cost of fertilization. Phosphorus fixation occurs when phosphate molecules react with soil particles leading to the formation of insoluble complexes. Only a small group of plants (see for example Gilbert et al., 1999; Hisinger and Gilkes, 1996) is capable of breaking apart some of those complexes and releasing P back to soil solution. For most crops, however, phosphorus becomes unavailable for plant uptake once it is tied up in these complexes. The capacity to fix phosphorus varies widely among soils. In

[9] Agronomic threshold is the amount of a plant nutrient in the soil at which this nutrient is no longer a limitation to achieve the potential yield of the crop.

many sandy soils fixation is minimal, while the fixation capacity of soils derived from volcanic ashes (Andisols) and rich in allophanic clays can present a serious challenge for farming (Vander Zaag and Kagenzi, 1986; Espinosa, 1992). The P-fixation ability of allophanic soils are normally strong enough to reduce carryover effects to negligible levels and crops are able to utilize no more than 10 to 30% of the phosphorus applied as fertilizer (Ludwick, 2002).

There are not many studies that analyze the economic impact of phosphorus depletion (some of the few cases include Abelson and Rowe, 1987; Buresh et al., 1997), and none, to the best of author knowledge, examine the impact of programs that subsidize soil fertility replenishment. Thus, a main contribution of this work is to fill this gap in the literature. In this chapter, I use data from Chile to analyze the implications of including soil phosphorus dynamics when selecting the targeting strategy for a soil remediation program.

The chapter is organized in the following way. Section Two presents a brief description of the dynamics of soil phosphorus. Section Three discusses the existence of a steady state in the farmer's problem of determining the optimal rate of fertilization. The discussion includes the solution of a continuous optimal control problem where the level of soil phosphorus is treated as a stock variable. The fourth section presents data from a fertility replenishment program from Chile and the econometric framework used to evaluate the theoretical model presented in Section Three. Then I use a Monte Carlo EM algorithm to solve an endogenous switching regression model with unobserved switching points, which permits evaluating the impact of the cost sharing program conditional on the alternative fertilization strategies the farmer can follow. Section Five discusses the results of the econometric estimation. Section Six concludes.

2 Soil phosphorus stocks and supply of plant-available phosphorus

Phosphorus is one of the three necessary macro elements for plant nutrition (the others are nitrogen and potassium). Plant roots are able to absorb this nutrient only from the soil solution and only in inorganic forms. When the fixation capacity of the soil is high, phosphate fertilizer applied is attached to soil particles almost immediately (Potash and Phosphate Institute, 2001). As indicated by Ludwick, (2002), the strongest attachment occurs with oxides of iron (Fe) primarily in soils with pH below 4.0. As soil pH increases, P is fixed preferentially in aluminum (Al) compounds. This binding is not that strong as with iron, but the availability of phosphorus to plants is still reduced dramatically. Finally, binding to relatively weak alkaline (Ca) compounds can occur in soils with pH higher than 7.5. In many soils, binding of P happens at the surface of soil particles, which permits the process be reverted by mass action processes. However, natural release of P back to solution happens at a lower rate than binding does (Barrow, 1983a and 1983b). On the other hand, in volcanic soils with allophanic clays and humus-Al complexes, the fixation process tends to be irreversible (Espinosa, 1992; Nanzyo et al., 1997).

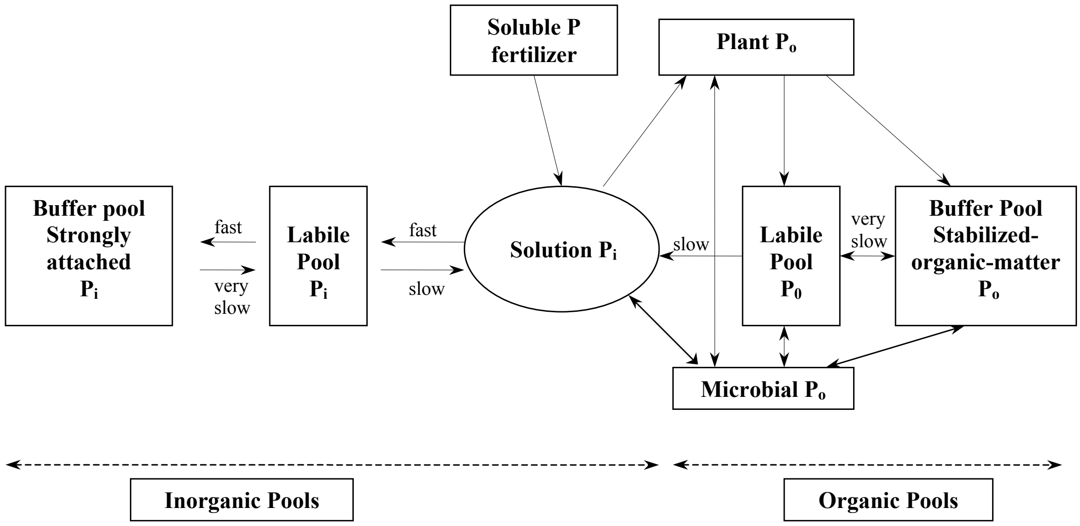


Figure 4.1. Soil-phosphorus reservoirs and phosphorus dynamics.
(Modified from Steward and Sharpley, 1987) P_i and P_o indicate sources of inorganic or organic phosphorus, respectively.

Traditionally, phosphorus in the soil is treated as occurring in three reservoirs or pools (see Figure 4.1): a buffer pool, a labile pool, and phosphorus in the soil solution or solution-P. All of them receive contributions from both organic and inorganic sources. The buffer pool includes inorganic phosphorus contained in soil minerals and strongly attached to Al and Fe compounds, and organic phosphorus occurring in stabilized organic matter. The labile pool includes inorganic phosphorus loosely attached to the surface of clay particles, and a limited amount of organic phosphorus that can be rapidly mineralized and thus made available for plant uptake. Solution-P, on the other hand, is non-attached inorganic phosphorus that can either be immediately taken up by plants, used by soil biota and converted in organic phosphorus, or, alternatively, attached and become part of either the labile pool or the buffer pool. Significant movement of phosphorus between the buffer and labile pools only happens in the long run, while the interaction between the labile pool and the soil solution is much faster. Large contributions to solution-P can be provided through fertilization with a highly soluble source of inorganic phosphorus.

In absence of fertilization and phosphorus exportation, there exists a chemical equilibrium between the three reservoirs mentioned above. If equilibrium is altered, e.g. when concentration of solution-P is reduced by plant uptake, the labile pool releases phosphorus to solution to restore the equilibrium between the labile-pool and the soil solution. This, in turn, reduces the phosphorus level in the labile pool, which alters the equilibrium between labile and buffer pools. In this case, however, given the low rate at which phosphorus can be released from the buffer pool, the speed at which the labile pool recovers is slower than the speed at which phosphorus is released to solution. Pastures used to raise cattle, for instance, feature a vegetative cover during most of the season. Periodic harvest or grazing of this cover permanently alters the equilibrium among phosphorus reservoirs. The labile pool is progressively depleted unless phosphate fertilization is provided. If phosphorus fertilizer is not applied, less and less phosphorus is available from soil solution, which, in turn, reduces forage and crop production.

If farmers wait to fertilize until signals of depletion become apparent, chemical linkages among phosphorus pools often make it difficult to recover the initial fertility levels. Fertilization causes an abrupt increase in the concentrations of phosphorus in the solution, which triggers a change in the kinetics of soil phosphorus. Since the equilibrium among the pools must be restored, most of phosphorus provided by fertilization will go to enrich the labile and buffer pools. The lower the level in the buffer pool the less phosphorus will remain available in solution to plant uptake. This is one of the reasons explaining why, once a certain level of depletion is achieved, soils with high phosphorus-fixation capacity are not remediated in developing countries. Unless cheap sources of phosphate are available, the gains from fertilization do not compensate (at least in the short run) the cost of the massive applications necessary for replenishment.

2.1 State equation of the phosphorus pools.

Modeling of phosphorus dynamics in this chapter is simplified by assuming the existence of only two reservoirs in the soil: the buffer pool and the plant-available pool. The buffer pool is responsible for both P-fixation and long run release of phosphorus to plant-available pool. The plant-available pool includes phosphorus loosely attached in the labile pool and phosphorus free in the soil solution. Considering two stocks instead of three is not a restrictive assumption. Given the fast kinetics of phosphorus between solution and the labile pool, it seems reasonable considering them as a single reservoir for purposes of policy analysis.

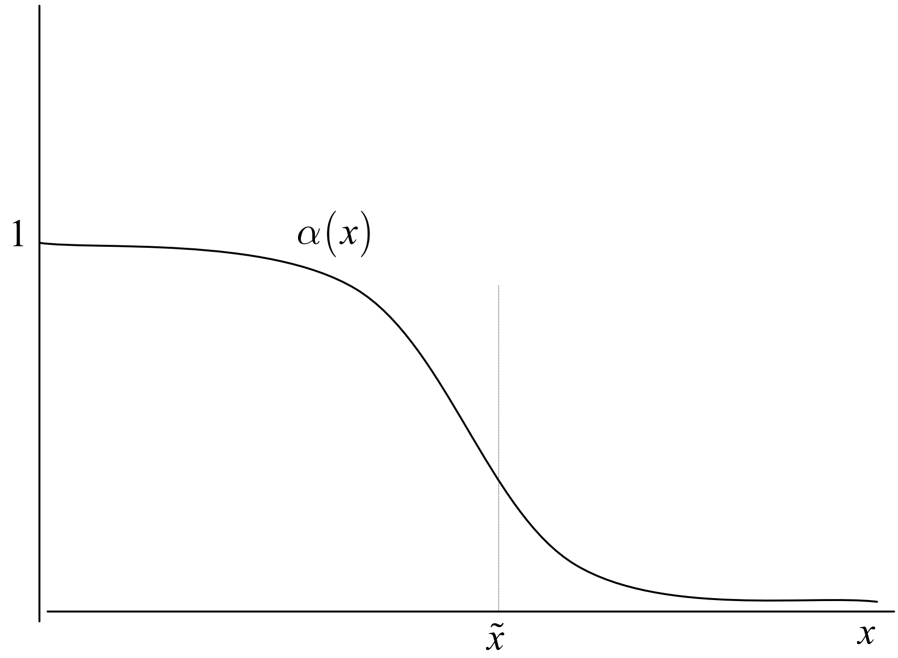


Figure 4.2. Phosphorus fixation function.

Let x be the non-negative level of phosphorus in the buffer pool. I assume that soil fixation power follows a function $\alpha(x)$, which depends only on the level of the buffer pool and represents the proportion of phosphorus fertilizer that is fixed during the season (and, consequently, contributes to the level of phosphorus in the buffer pool x). The properties of $\alpha(x)$ are: $0 < \alpha(x) \leq 1$, $\alpha^{'}(x) \leq 0$, $\alpha^{'}(0) = 0$, and $\alpha^{''}(x)$ is negative for $x < \tilde{x}$ and positive for $x > \tilde{x}$ (see Figure 4.2). The phosphorus binding power of the soil attains its maximum at $x = 0$, i.e. when all possible binding sites in the buffer pool are unoccupied, and approaches zero as x becomes large. The state equation giving the seasonal change in the buffer pool is

$$\dot{x} = \alpha(x)z - \gamma(x) + \kappa \tag{4.1}$$

According to equation (4.1), the level of the buffer pool increases by capturing phosphorus from fertilization z in amount of $\alpha(x)z$ and via natural weathering of parent materials and organic matter at constant rate κ. On the other hand, it is reduced by phosphorus desorption according to a function $\gamma(x)$, which is quasi-convex in x and satisfies $\gamma'(x) > 0$ for $x > 0$, $\gamma'(0) = 0$ and $\gamma(0) = 0$. Desorpted phosphorus goes to the plant-available pool in order to sustain the chemical equilibrium between the two pools.

Accordingly, there are two sources of phosphorus for the plant-available pool: a proportion $1-\alpha(x)$ of the fertilizer applications z, and contributions from the buffer pool in amount $\gamma(x)$.

3. The optimal fertilization strategy.

3.1 The farmer's dilemma as an optimal control problem

I represent the problem in a continuous-time infinite-horizon optimal-control framework with one control and one state variable. The control variable is the rate of fertilization z, while the state variable is the phosphorus level in the buffer pool. Thus, the farmer solves

$$\max \int_0^{\infty} \left\{ f\left(z\cdot\left(1-\alpha\left(x\right)\right)+\gamma\left(x\right)\right)-w\cdot z \right\} e^{-rt}dt$$

$$\begin{aligned} s.t. \quad & \dot{x} = z\cdot\alpha\left(x\right)-\gamma\left(x\right)+\kappa \\ & z\left(t\right)\geq 0\ , \quad x(0) = x_0 \end{aligned} \tag{4.2}$$

where $z(t)$ is the phosphorous fertilization rate at time t, $x(t)$ is the corresponding phosphorus level in the buffer pool, x_0 is the initial buffer pool level, w is the phosphorous price, r is the farmer's discount rate, and $f\left(\cdot\right)$ is a twice-differentiable production function. Prices have been normalized such that output price equals one.

The current-value Hamiltonian is

$$H = f\left(z\left(1-\alpha\right)+\gamma\left(x\right)\right)-wz+\lambda\left(z\alpha-\gamma\left(x\right)+\kappa\right) \tag{4.3}$$

where λ is the shadow value of a marginal increase in the level of the buffer pool. The first order necessary conditions for interior solutions of (4.3) are

$$\frac{\partial H}{\partial z}\leq 0 \quad \Rightarrow \quad f^{'}\left(1-\alpha\right)-w+\lambda\alpha = 0 \tag{4.4}$$

$$\frac{\partial H}{\partial x} = \dot{\lambda}-r\lambda \quad \Rightarrow \quad \dot{\lambda} = \left(z\alpha^{'}-\gamma^{'}\right)\left(f^{'}-\lambda\right)+r\lambda \tag{4.5}$$

$$\frac{\partial H}{\partial \lambda} = \dot{x} \quad \Rightarrow \quad \dot{x} = z\alpha-\gamma+\kappa \tag{4.6}$$

The interpretation of (4.4) follows from the well-known rule of profit maximization: it will be optimal to fertilize until marginal benefits equalize marginal costs. In this case we have two kinds of benefits: on one side there are the revenues from output sale, and on the other side we have an enriched buffer pool, which ensures less phosphorus fixation and fertilization savings in future crop seasons.

Non-convexities make it difficult to find a closed form for the internal solution(s), if any, of problem (4.2). Therefore, I combine graphical and numerical tools in the next section to analyze the existence and nature of potential equilibria.

3.2 Phase-diagram analysis

I begin with the phase diagram in the $z-x$ plane. Differential equations giving the rates of change of x and z are needed in order to characterize the shapes of the curves $\dot{x}=0$ and $\dot{z}=0$. The equation for $\dot{x}$ is simply the state equation (4.6), while a differential equation involving the first derivative of z with respect to time can be obtained by taking the time derivative of equation (4.4). We get

$$(1-\alpha)\left[\left(\gamma^{'}-z\alpha^{'}\right)\dot{x}+\dot{z}(1-\alpha)\right]f^{"}-\alpha^{'}\left(f\,'-\lambda\right)\dot{x}+\dot{\lambda}\alpha=0 \qquad (4.7)$$

Equation (4.7) can be combined with equations (4.4) through (4.6) to obtain

$$\begin{aligned}\dot{z}&=-\frac{1}{(1-\alpha)^{2}f^{"}}\left\{(1-\alpha)\left[\left(\gamma^{'}-z\alpha^{'}\right)\dot{x}\right]f^{"}-\alpha^{'}\left(f\,'-\lambda\right)\dot{x}+\dot{\lambda}\alpha\right\}\\&=-\frac{1}{(1-\alpha)^{2}f^{"}}\left\{(\ \ -\alpha)\left(\gamma^{'}-z\alpha^{'}\right)(z\alpha-\gamma+\kappa)f^{"}-\frac{\alpha^{'}}{\alpha}(f-w)(z\alpha-\gamma+\kappa)\right.\\&\qquad\left.+r\left(w-(1-\alpha)f\,'\right)+\left(f\,'-w\right)\left(z\alpha^{'}-\gamma^{'}\right)\right\}\end{aligned} \qquad (4.8)$$

1

Equations involving $\dot{x}=0$ and $\dot{\lambda}=0$ needed to construct the phase diagram in the $\lambda - x$ plane are provided by the state equation (4.6) and the adjoint equation (4.5).

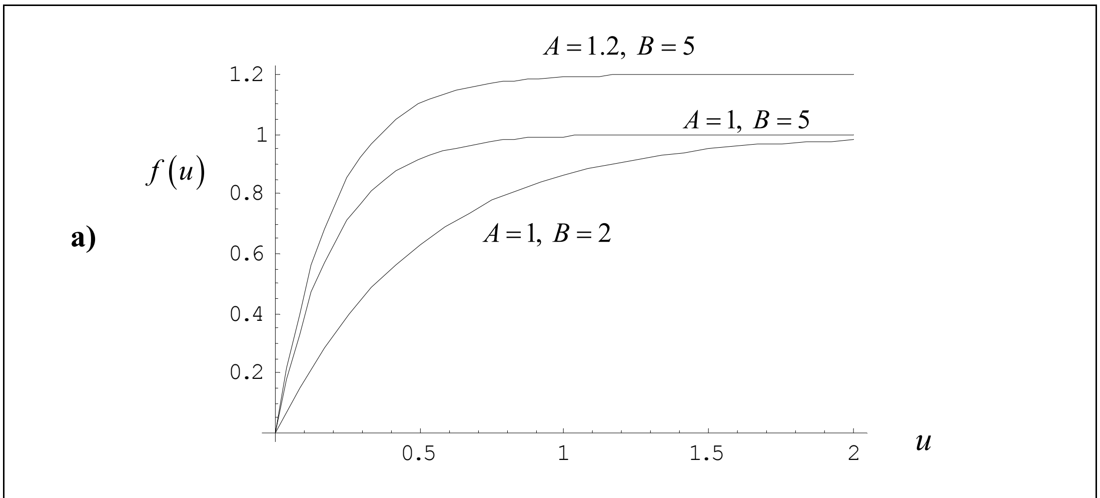
I draw the phase diagrams following four steps: 1) choosing functional forms for $f(\cdot)$, $\alpha(\cdot)$, and $\gamma(\cdot)$, 2) solving numerically the equations $\dot{x}=0$ and $\dot{z}=0$ as functions of x and z, and equations $\dot{x}=0$ and $\dot{\lambda}=0$ as functions of x and λ, 3) drawing the curves $\dot{x}=0$ and $\dot{z}=0$ in the $z-x$ plane and curves $\dot{x}=0$ and $\dot{\lambda}=0$ in the $\lambda - x$, and 4) determining the signs of $\dot{x}$, $\dot{z}$, and $\dot{\lambda}$ in each isosector of the phase diagrams.

3.2.1. Functional forms and baseline parameter values.

I assume the following functional forms for the functions involved.

Production function: $f(u) = A\left(1-e^{-Bu}\right)$, where A represents the maximum yield the farmer can achieve using the technology available and B is a measure of the "speed" at which that potential is attained (see Figure 4.3a). In this case a larger B corresponds to a more efficient use of soil phosphorus.

To choose the values of w, A and B for the baseline simulation, I use the work of Smith-Ramírez et al. (2002), which classifies Chilean dairy farms according to technical, productive and human capital characteristics. I chose to work with dairy farms because dairy and beef farms are the main target of fertility remediation programs in Chile. To simplify the notation I normalize the maximum yield to the unity, i.e. $A = 1$, which transforms w in the price of phosphorus relative to the maximum revenue for acre achievable by the farmer conditional on the technology available.



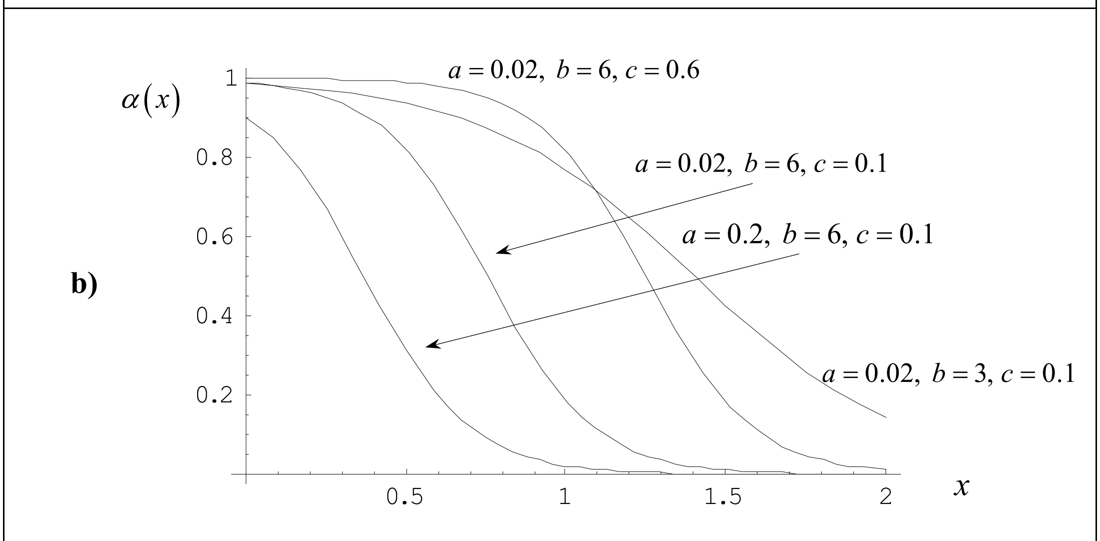


Figure 4.3. Production and fixation functions for different parameter values

In what follow, I use information from Region X in Chile, which produces 70% of the Chilean milk and includes about 6,700 dairy farms. Smith-Ramírez et al. (2002) classify the farms in four groups according to their production levels and technologies used in their production process. Table 4.1 gives some characteristics of the groups.

Table 4.1 Production indicators of Region X dairy farms (Chile)

Group	Indicator	Quartile 1[2]	Median	Quartile 3
S1	production per ha[1] (Kg ha^{-1} $year^{-1}$)	459	701	959
	farm size (ha)	6	10	15
	production per cow (Kg $year^{-1}$)	741	1131	1547
S2	production per ha (Kg ha^{-1} $year^{-1}$)	1345	1906	2588
	farm size (ha)	65	91	154
	production per cow (Kg $year^{-1}$)	1868	2647	3594
S3	production per ha (Kg ha^{-1} $year^{-1}$)	2575	3542	4685
	farm size (ha)	50	83	144
	production per cow (Kg $year^{-1}$)	2220	3053	4038
S4	production per ha (Kg ha^{-1} $year^{-1}$)	3768	4408	5134
	farm size (ha)	120	197	274
	production per cow (Kg $year^{-1}$)	3925	4592	5348

[1] 1 hectare (ha) is equivalent to 2.471 acres.
[2] 25% of the population have a value below quartile 1; 25% of the population have a value above quartile 3.

The Chilean dairy pays different prices to farmers according annual production. Currently, farmers in the S1 group receive an average of 13.6 cents per kilogram of fluid milk, those in the S2 group receive 15.3 cents, those in the S3 group receive 17.0 cents, and those in the S4 group receive an average of 18.7 cents per kilogram. The price of the ton of phosphorus (in its form of P_2O_5) is 645 dollars in Chile. In order to obtain values of w for each group in Table 4.1, I need to know the potential yield per hectare for each of them. Since these data are not available, I used the 95th percentile for the production per hectare for each group. After combining all this information, the values for w are 2.39, 1.30, 0.79, and 0.57 for groups S1, S2, S3, and S4, respectively. The selection of a value for parameter B is discussed below in conjunction with the selection of the parameters for the phosphorus fixation function.

Phosphorus fixation function: $\alpha(x) = \dfrac{1}{1+ae^{b(x-c)}}$, where parameters a, b, and c determine together the location of the inflection point of the curve in Figure 4.3b. The most influential parameter in this function is c, which controls the curvature of the function before it reaches the inflection point. According to Escudey et al. (2001), fertile Chilean Ultisols and Andisols (both volcanic ash soils) contain from 733 mg P kg $^{-1}$ to 3470 mg P kg $^{-1}$. For purposes of the simulation, I take the middle value, 2200 mg P kg $^{-1}$ as a measure of the aggregated capacity of the buffer and plant-available pools. The plant-available pool, however, contains only a small fraction of the phosphorus in the soil, which means that 2200 mg P kg $^{-1}$ is also a good measure of the buffer pool only. To reproduce the high fixation power of Region X Andisols, I choose parameters a, b, and c such that the soil fixes 90% of the fertilizer applied when the buffer pool is at half of its capacity. Soil fixation power declines then rapidly to fix only 20% of the fertilizer when the buffer pool is at two thirds of its capacity. Both conditions, along with the requirement that $\alpha(x)$ goes to zero as x approaches 2200, are satisfied if $a = 0.018$, $b = 6$ and $c = 0.1$ (see Figure 4.3b).

To choose parameter B in the production function, I use two pieces of information: (1) the average dairy farm in Region X produces no more than a third of his potential yield (Smith-Ramírez, 1999) and (2) the phosphorus content of the plant-available pool never exceeds 5% of the phosphorus content of the buffer pool (Rowell, 1994). Therefore, I choose B such as the yield is one third of the potential when the level of the plant-available pool is 0.1 mg P kg $^{-1}$, which corresponds to $B = 5$ (see Figure 4.3a, u is the level of the plant-available pool).

Desorption function: $\gamma(x) = sx$. A linear form is assumed, where the parameter s denotes the proportion of the buffer pool being released to the plant-available pool during the crop season and ranges from zero to one. There is little useful information to guide the selection of a value for s to use in the simulation. However, since the plant available pool rarely contains more than 5% of the phosphorus in the buffer pool, I chose s equals to 0.05. For the contribution from parental material to the buffer pool (the parameter κ in the state equation), I chose a value $\kappa = 0$ to better represent volcanic soil conditions.

3.2.2 Determination of the curves $\dot{x} = 0$, $\dot{z} = 0$ and $\dot{\lambda} = 0$

The $\dot{x} = 0$ and $\dot{z} = 0$ curves were drawn in the $x - z$ plane by writing the right-hand sides of expressions (4.6) and (4.8) as functions of z and x. Both expressions were then set equal to zero and solved as simultaneous equations by numerical methods[10] for various combinations of parameter values. The $\dot{x} = 0$ and $\dot{\lambda} = 0$ curves were drawn in an analogous manner using equations (4.5) and (4.6).

According to the previous discussion, the parameter values used for the baseline simulation were $\kappa = 0,\ r = 0.15,\ w = 1.8,\ A = 1,\ B = 5,\ a = 0.018,\ b = 6,\ c = 0.1$ and $s = 0.05$.

---

[10] I used Mathematica 4.1 from Wolfram research Inc.

A value $w = 1.8$ was chosen in order to represent farmers in the groups S1 and S2 (see Table 4.1), which accounts for about the 80% of the farmers in Region X and have similar characteristics as three quarters of Chilean dairy farmers (Smith-Ramírez et al. 2002). Common values for the farmer's discount rate in the literature for natural resource economics range between 10% to 20% (Kremen et al. 2000; Lu and Stocking 2000). Here, I have chosen $r = 0.15$. Graphical outcomes are presented in Figure 4.4.

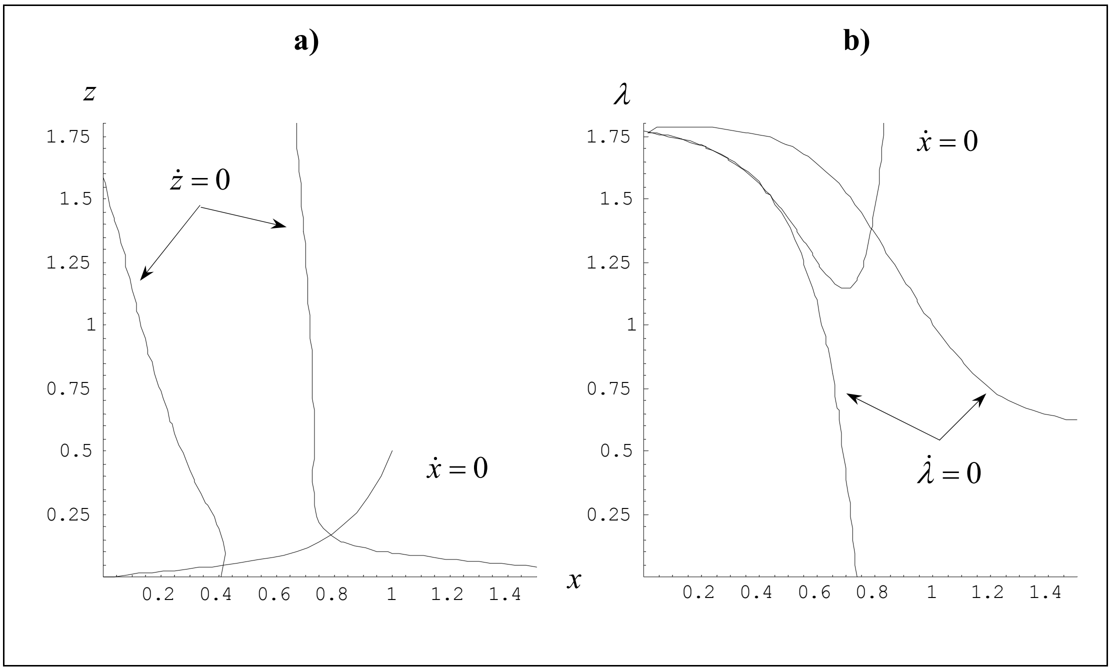


Figure 4.4. a) Curves $\dot{x} = 0$ and $\dot{z} = 0$ in plane $x - z$, b) curves $\dot{x} = 0$ and $\dot{\lambda} = 0$ in plane $x - \lambda$.

A noteworthy characteristic of the solutions depicted in Figure 4.4 is the existence of two candidates for interior solutions, which originate in double branched curves for the loci $\dot{z} = 0$ and $\dot{\lambda} = 0$. The two solutions $\left(x^*, z^*, \lambda^*\right)$ for the parameter values used in the simulation are $(0.467,\ 0.027,\ 1.383)$ and $(0.947,\ 0.185,\ 1.261)$. The only observable amounts in these two sets of values are the fertilization rates: 0.027 and 0.185 tons of phosphorus (in its P_2O_5) form) per hectare. They correspond to 60 and 400 kilograms of Triple Superphosphate (the most used phosphate fertilizer in Chile). Fertilization rates of 60 Kg of Triple Superphosphate are typical in subsistence farming in Chile and characterize most of the farmers in the S1 group. Fertilizer application rates between 300 and 400 Kg of Triple Superphosphate are usual among the most productive farms in groups S3 and S4.

The existence of two branches in the loci $\dot{z}=0$ and $\dot{\lambda}=0$, however, is conditional on the values of function parameters, phosphorus price, and discount rate. Figure 4.5 sketches the phase diagram for different combinations of values on the $x-z$ plane. In order to give some structure to the next discussion, I label the branches of $\dot{z}=0$ as the "low-yield branch" (the left-hand side branch) and "high-yield branch" (the right-hand side branch). The labels follow the crop yields at each solution. For instance, for the parameter values used to construct Figure 4.4, the yield (recall that maximum potential yield equals one) at the low-yield (LY) solution is 0.127, while it is 0.603 at the high-yield (HY) solution. This outcome reproduces closely Chilean conditions: farmers in groups S1 and S3 produce milk on grassland, however, farmers in S3 quintuplicate the production per hectare of those farmers in group S1 (see Table 4.1).

According to Figure 4.5a, under low fixation power and/or high production potential, only one internal solution exists: the high-yield equilibrium. Thus, under high production levels and high milk prices (groups S3 and S4), soil remediation is always optimal and every farmer attains the steady state.

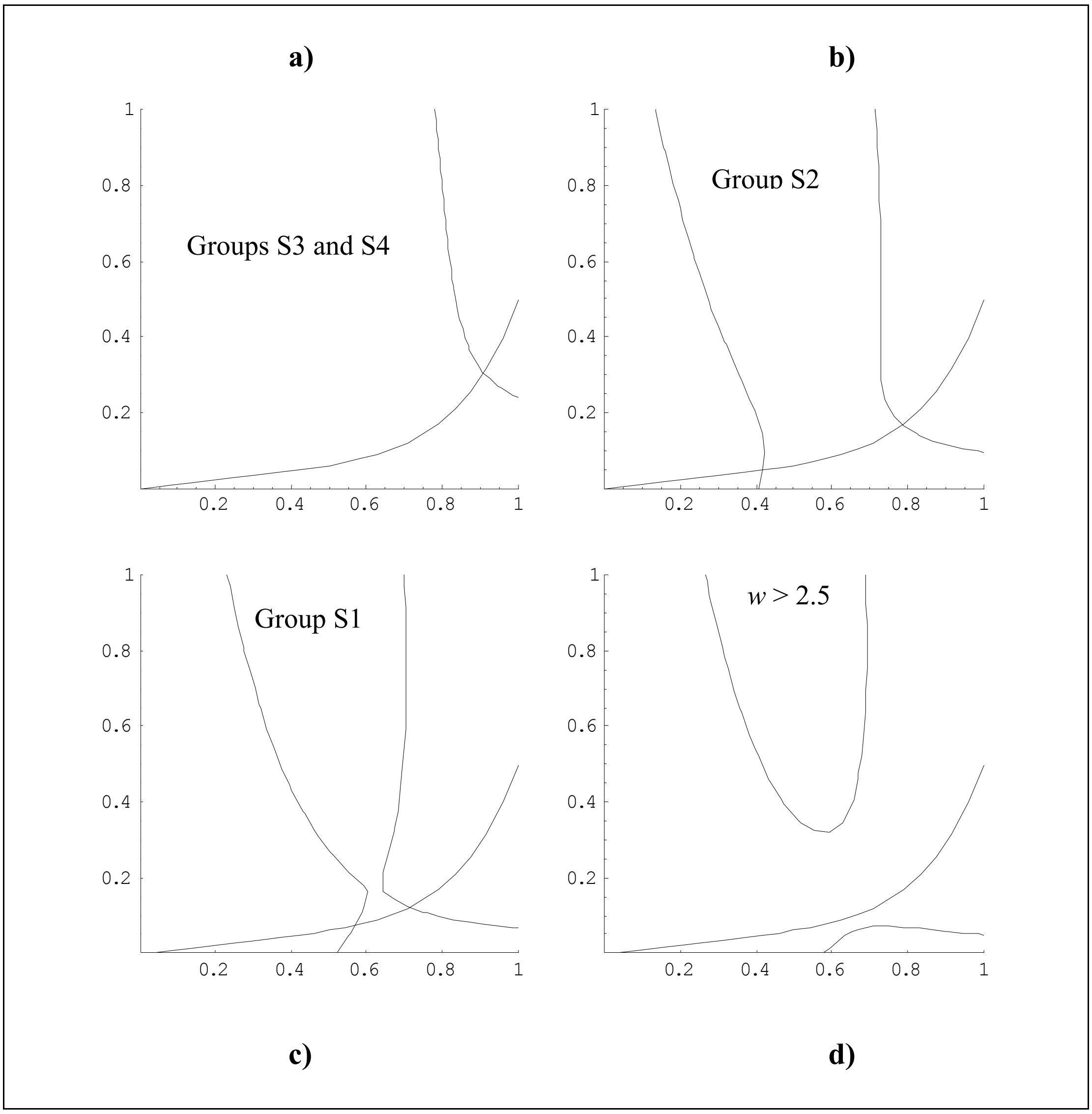


Figure 4.5. Phase diagrams on plane $x-z$ for different parameter values

Figures 4.4b and 4.4c, on the other hand, show two alternative fertilization paths with two corresponding interior solution. Conditions for the existence of low-yield and high-yield solutions include: medium to high phosphorus prices relative to revenue per hectare (groups S1 and S2 in Table 4.1), medium to high phosphorus fixation power, and/or limited technology for achieving high crop yields. I discuss this case extensively below since it characterizes the conditions facing most Chilean dairy farmers.

Finally, Figure 4.5d presents the case in which low milk prices and/or poor technological levels prevent the existence of an interior solution. This is the case in which, after the resource is exhausted, farm operations are abandoned or the farm is sold. This case represents to many small farmers that in the 1990s migrated to cities and sold their farms either to bigger farmers or to people who currently use the land for forestry[11].

3.2.3 Analysis of the phase diagram and optimal paths

Since we are interested in determining optimal fertilization strategies, I analyze here the case depicted in Figure 4.4b on the $x-z$ plane. As mentioned previously, cases depicted in figures 4.3b and 4.3c are the ones that better represent the conditions of Chilean farms. The full diagram is presented in Figure 4.5.

First, I determine the nature of the internal solutions. I follow the procedure in Léonard and Van long (page 96, 1992), which includes taking a first order approximation of the system composed by the differential equations[12] for $\dot{x}$ and $\dot{z}$

---

[11] Chilean government supports actively the planting of fast-growing trees on eroded soil or on soil whose fertility has been mined.

[12] Alternatively, we can use a different pair of equations such as those for $\dot{x}$ and $\dot{\lambda}$ or those for $\dot{z}$ and $\dot{\lambda}$.

(i.e. equations (4.6) and (4.8)) in the neighborhood of each solution and finding the characteristic roots (eigenvalues) of the linearized system. The nature of the solutions is determined then according to the magnitudes and signs of the roots and whether they are real or complex.

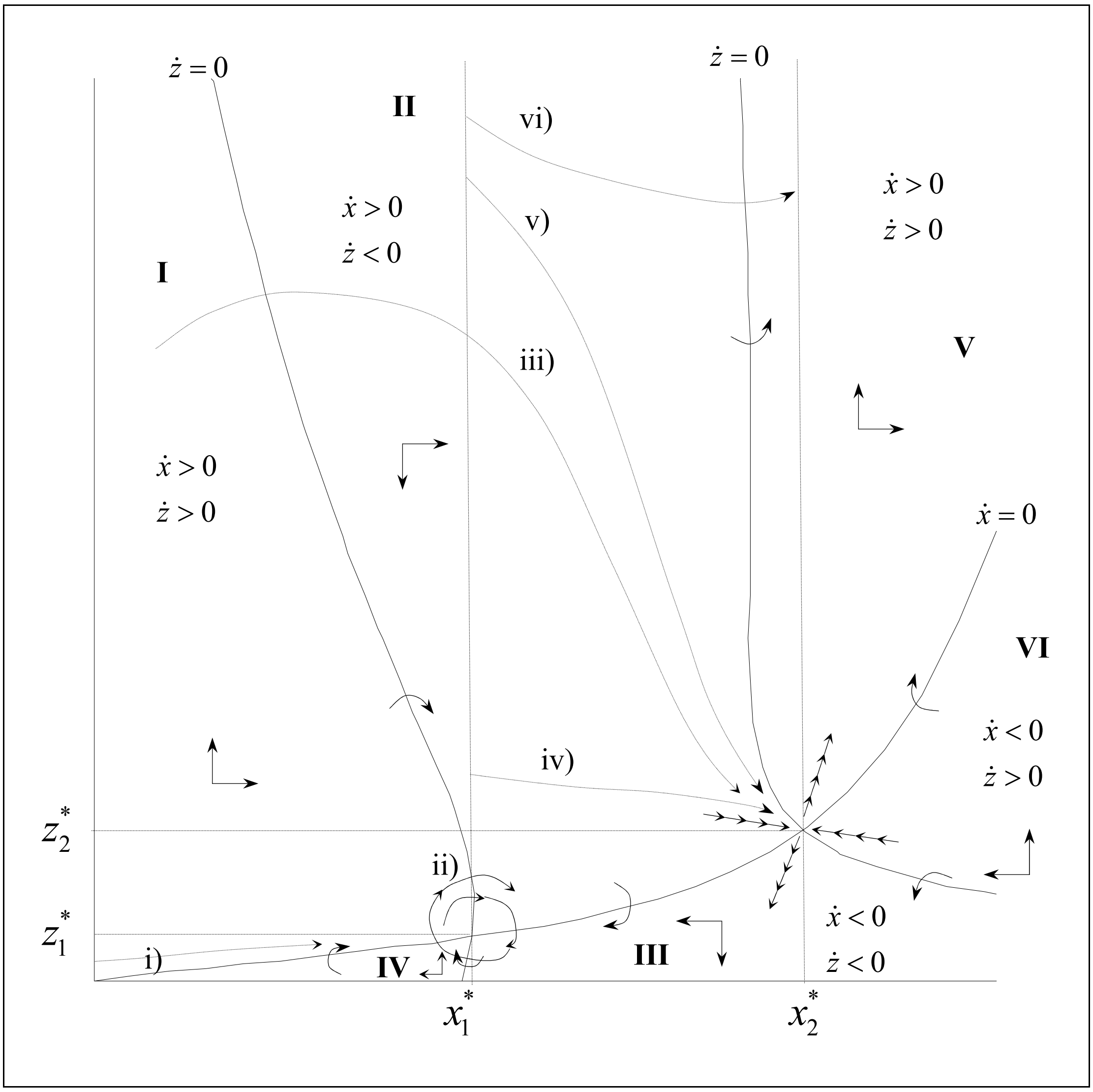


Figure 4.6. Phase diagram on plane $x-z$

For the baseline simulation parameters, the roots at the LY node are $0.075+1.168i$ and $0.075-1.168i$, i.e. one is the complex conjugate of the other. Since the real part of the roots is positive, the LY node is an unstable focus spiraling away from the node. The analysis of the eigenvalues, however, characterizes only the behavior of the system in the closest neighborhood of the LY node. Therefore, the existence of a limit cycle around the LY solution cannot be ruled out (Clark, 1990 pages 185-192). If a limit cycle exists, both the LY fertilization rate and the LY buffer stock level oscillate around the node $\left(x_1^*, z_1^*\right)$. Regarding to the HY node, its roots are 0.476 and -0.326. Since they are real, distinct, and have different signs, the HY solution is a saddle point.

In summary, two fertilization regimes can be postulated: a cyclic regime represented by the LY solution $\left(x_1^*, z_1^*\right)$ and the HY stable regime at $\left(x_2^*, z_2^*\right)$. Intuitively, these two solutions arise from the behavior of the production and fixation power functions, $f(\ )$ and $\alpha(\ )$ respectively. In the LY regime, although the marginal productivity of phosphorus is high, a larger share of phosphorus applied goes into the buffer stock so it is optimal to limit phosphorus application. In the HY regime, the opposite occurs. How high x_1^* can be depends on the relative behaviors of the functions $f(\ )$ and $\alpha(\ )$. In soils with high fixation power (high value of c), since the marginal productivity of phosphorus decreases faster than the marginal reduction in fixation power does, x_1^* cannot correspond to a significant enrichment of the buffer pool.

Therefore, the solution $\left(x_1^*, z_1^*\right)$ represents a regime associated with low crop yields, a "subsistence" fertilization path.

The existence of an oscillatory "equilibrium" for the system has its counterpart on the field. A common practice among farmers in groups S1 and S2 is to carry out periodical "low-scale" fertility remediations. The cycle begins with the planting of grass on a degraded pasture, which includes the application of "high" rates of phosphorus (most commonly 200-250 Kg of Triple Superphosphate per hectare). Unable to sustain this level of fertilization for more than one season, the farmer will apply only a reduced fraction (some farmers apply nothing) of the initial rate during the next seasons. Thus, the initial "remediation" corresponds to investing in improving the content of phosphorus in the buffer pool. During the next seasons, the farmer gets his investment back by harvesting from the enhanced soil and reducing the phosphorus applications. After a certain number of seasons (three years usually), the soil is back to its original condition and it is time to repeat the process.

The relevant issue from a policy standpoint at this stage of the analysis is the identification of the determinants that make a farmer to choose one regime over the other. The phase diagram in Figure 4.6 provides several clues pointing toward which those determinants are. Note that no optimal fertilization path combining rates below z_2^* can reach the HY steady state. Shifting from the LY to the HY regime involves a large short run increase in phosphorus applications since attaining the HY steady state is feasible only by applying $z > z_2^*$. But that short run increase in the fertilization rate may not be affordable for farmers who face a high effective price of phosphorus or interest rate because of credit constraints or who utilize less efficient technology. In the field, this is the case of many small Chilean farmers, who lack of adequate technological and managerial skills (low values for parameters A and B). Many of them have little schooling; many are also old and have short planning horizons (hence high discount rates).

Farms with highly depleted soils in region I can reach the HY equilibrium following an optimal trajectory such as iii). It begins with fertilization rates well above z_2^* and then gradually reduces phosphorus applications as it approaches the HY equilibrium. Farmers already in the LY steady state can move to the HY equilibrium by applying fertilization rates greater than z_2^*. The higher are these fertilization rates the faster is the approaching to the equilibrium. Trajectory iii), for instance, is faster than trajectory iv) at early stages of the fertility remediation, while v) is faster than iii). A more aggressive approach is depicted by trajectory vi), which considers the use of very high fertilization rates to move from the LY regime to x_2^*. Only once the equilibrium buffer pool level is attained, phosphorus applications can be reduced until reaching the HY solution $\left(x_2^*, z_2^*\right)$

From the analysis, there are two issues relevant for policy analysis. First, we have the existence of two alternative long-term regimes, or, in other words, two optimal long-term fertilization strategies. Second, the endogenous nature of the regimes raises questions as to the appropriateness of using exogenous indicators as the core components of the targeting policy of a soil fertility remediation program. As the preceding analysis has shown, the optimal levels of remediation of the buffer pool depend on individual characteristics such as soil properties, discount rate, efficiency in the production process, and technology used in the farm.

According to the analysis of the phase diagram, it seems that the capacity to afford high fertilization rates is crucial in determining the fertilization path a farmer will follow. In order to find worthwhile to remediate the soil, the present value of the HY steady state must exceed the present value of the cost of attaining it. For farmers with very low initial buffer pools, i.e. $x(0) < x_1^*$, this may not be the case, particularly if they face managerial or financial constraints. These farmers may not have a choice except to follow a LY fertilization strategy. On the other hand, we have farmers that are neither technologically nor financially constrained and thus they will consider achieving the HY steady state (x_2^*, z_2^*) as the optimal option.

In order to allocate a limited budget efficiently, a program aiming soil fertility remediation should provide financial support preferentially to those farms considering the LY solution optimal if left to their own, and support should be provided only up to the point at which soil fertility makes the HY equilibrium affordable. It can be argued that providing funding beyond that point will bring benefits from accelerating the transition to the HY steady state. However, unless enough money is available to accomplish the two tasks, program budget will bring more benefits if aimed to move farmers from a LY regime to a path ending at the HY steady state. Making it affordable for farmers who would otherwise choose the LY regime to attain the HY equilibrium brings short and long run benefits, while accelerating the transition brings short run gains only.

Adequate targeting is expected to be difficult in one-time subsidies for remediation, like the one to be discussed in the next sections, which do not consider follow-up subsidies for fertility maintenance. If such a program fails in collecting adequate information from farmers, then it is possible that part of the budget goes to people that, after cost sharing is stopped, find optimal to move back to the LY regime. Consider, for instance, old farmers with short time-horizons or farmers with technological constraints that make unaffordable to sustain a fertilization rate z_2^*. Graphically, this outcome corresponds to a trajectory that begins at some point on the vertical line representing x_2^*, moves through region III in Figure 4.6, and returns to the cyclical path around $\left(x_1^*, z_1^*\right)$.

The second important issue is the endogenous nature of the steady state levels $\left(x_1^*, z_1^*\right)$ and $\left(x_2^*, z_2^*\right)$. According to the simulations in Figure 4.5, the steady state level of the buffer pool depends on soil properties (κ, γ, and α), phosphorus price, and farmer's attributes (such as age, farming experience, and financial condition). Thus, steady state levels are not exogenously determined but rather depend on individual characteristics, a fact that has implications for targeting cost-sharing.

Fertility remediation programs like the one to be discussed in the next section currently use an exogenous target level to allocate the program budget. Thus, only farms having a phosphorus level below the target are eligible to receive funding. These programs provide remediation cost sharing (or, in other words, a subsidized phosphorus price) to enrolled farmers until the target level of phosphorus in the soil is achieved. This could take more than one crop season, but, once the target level is reached, cost sharing is stopped.

The consequences of using an exogenously determined target level to allocate the budget on program efficiency depend on where the target level is located with respect to the alternative equilibria in Figure 4.6. Consider for instance that the exogenous target level x_t is greater than x_2^* for a particular farmer. In this case, the farmer is eligible for receiving cost sharing; however, if the buffer pool level was already at x_2^*, then cost sharing will have only short run effects. After x_t is attained and cost sharing ceased, the farmer will return to the HY equilibrium and cost sharing effects will vanish. On the other hand, if the farmer condition was on some cyclical path around $\left(x_1^*, z_1^*\right)$, then, after attaining the target, the farmer will follow some optimal trajectory back to $\left(x_2^*, z_2^*\right)$. In this case, cost sharing has the desired effect, although some cost sharing money may be spent inefficiently when taking the phosphorus level beyond x_2^*.

Consider now a farmer for who the target level is in between the two equilibriums, i.e. $x_1^* < x_t < x_2^*$. In this case, if the farmer is already at the HY equilibrium, then he is not eligible. If the farmer is at the LY equilibrium or on the way to the HY equilibrium but still below x_t, he is eligible for cost sharing. Awarded farmers will move from region I to region II until reaching $x = x_t$, then cost sharing will be stopped. What happen afterwards will depend on farmer attributes. Some farmers may follow a stable path, such as trajectories iii) and iv), and reach the HY steady state. Others may move from region II to region III and finally finish again in LY regime. In the first case, cost sharing has the desired effect of altering the long run fertility regime. In the second case, there are short term effects only. The effect of cost sharing on farmers on the way to the HY equilibrium will be to accelerate the transition, but the effect will be short run as well. Thus, if $x_1^* < x_t < x_2^*$, cost sharing will have long run effects only on farmers who would otherwise have been at the LY regime and stay at the HY equilibrium once attained it.

Evaluating the performance of a fertility remediation program under the theoretical framework developed above requires dealing with two econometric issues. First, it is necessary to handle the problem of determining which long-term fertilization strategy the farmer would choose in the absence of cost sharing. Note that even we may assume the farmer has attained some equilibrium, we, in general, do not know in which of the two possible regimes he is. Second, both short and long run effects of the program must be estimated in order to identify clearly the two fertilization regimes predicted by the theoretical model. Actually, if farmers that are not financially constrained receive cost sharing, then long run effects should be negligible for them. I discuss these two issues in the following sections, where I describe the program from which data was collected and develop the econometric model.

3.3 A description of the Chilean soil-fertility replenishing program.

Beginning in 1996 after successive measures to open the local economy to foreign markets, the Chilean government implemented a cost-sharing program aiming to replenish soil fertility in those agricultural operations whose existence could be jeopardized in a free-trade environment. The program provides support to farms with phosphorus pools depleted either naturally or by human activities, which are widespread in the country due to the volcanic origin of most of its soils and the fact that managerial and financial constraints prevent many farmers from replenishing phosphorus stocks. So far, farms awarded by the program have been mainly beef and dairy operations on grassland in central and southern Chile

Participation in the program is voluntary and farmers must submit an application to be considered for funding. Applications can be prepared by an authorized agronomist, who can be a private consultant or on the staff in a public agency. Applications are ranked according to a number of factors, the principal one being whether the soil phosphorus level falls below a target level previously established by the program. Applicants who are awarded funding receive enough to ensure that the target level is attained, a process that might take more than one season.

The current phosphorus target level has been kept constant in the last years at 15 mg Kg^{-1} Olsen in the plant-available pool. This level is motivated by agronomic considerations: 15 mg Kg^{-1} Olsen in the plant-available pool is about the minimal level necessary to guarantee survival of most grass species and hence commercial production on most type of grasslands in Chile. The premise of the program is that, once the target is achieved, farmers will maintain a phosphorus level of at least 15 mg Kg^{-1} Olsen without additional funding.

The Chilean program gives us a good opportunity to test the theoretical model presented in Section 3. At least three testable hypotheses emerge from the preceding analysis. First, there exist two farmer sub-populations. One is constrained financially and/or technologically and thus unable to achieve the HY equilibrium level of soil phosphorus. The other sub-population faces no significant constraints. Thus, the level of phosphorous observed in the second group in absence of the program is the HY steady state level (or the farm is on the way to that equilibrium level).

Second, the short run effect of cost sharing on the LY subpopulation exceeds the short run effect on the HY subpopulation. Although program administrators want to grant cost-share funding to the constrained group preferentially and to those farms on the way to the HY equilibrium alternatively, they cannot avoid awarding farmers already at the HY steady state because they do not know neither which sub-population farmers belong to nor the individual HY phosphorous levels. Nonetheless, from the discussion in Section 3.2.3, the short run effect of cost sharing should be greater on farms in the LY regime because those farms are the ones receiving the largest phosphorus applications.

Finally, there is a third hypothesis: cost sharing has no long run impact on the HY subpopulation but does on the LY subpopulation. However, panel data, which is not available in this study, are needed to test this hypothesis.

The following sections present an econometric framework that models the existence of two phosphorus fertilization paths in a farmer population, where the adoption of one or the other depends on whether the farmer faces or not constraints that prevent him from attaining the HY equilibrium. Simultaneously, the impact of a cost-sharing program is evaluated conditional on the fertilization strategy adopted by the farmer. By using an endogenous-switching regression with unobserved switching points, I analyze the suitability of using an exogenous target level as the core of the targeting policy in a cost-sharing program aiming the recovery of a natural-resource stock.

4 Data, Econometric model and Estimation

4.1 Data.

The Chilean Agricultural Policy and Statistics Office provided the data used to evaluate the theoretical model. Data were collected in 2001 by surveying a total of 856 farms from the population targeted by the cost-sharing program. The survey sample was stratified according to geographic location (4 north-to-south strata) and total acreage (2 strata). The sample was distributed proportionally within each stratum; the Agricultural Policy and Statistics Office provided the corresponding expansion factors. After cleaning the data set from observations with contradictory or missing information, the sample contained observations on 505 farms, 177 of which received cost sharing for at least one year and 328 of which were never awarded cost share funding. A short description of the variables used in this study follows. Descriptive statistics of the full set of variables is presented in Table 4.2.

The variable SHARE indicates the share of land enrolled in the program at some point between years 1996 and 2000 inclusively and is censored from below at zero. The variable POLSEN gives the current (year 2001) level of phosphorus in the plant available phosphorus. Twelve soil samples were used to build a composite sample for each farm, which was analyzed to obtain an estimate of the average level of soil phosphorus content. Phosphorus fixation power is proxied by the variable ALSAT, which gives the aluminum saturation of the soil or level of aluminum in the soil solution. In Chilean soils, most of the phosphorus is fixed to aluminum compounds. Thus, following the discussion from the introduction, the higher the aluminum concentration in a soil, the higher is its phosphorus fixation power (i.e. the higher the parameter c in $\alpha(x)$).

Variables related to credit accessibility and cash flow included in the econometric estimation are AACRE, and REV. AACRE gives farm acreage suitable for agricultural operations, and REV provides an estimation of average annual revenue per hectare. It is expected that farms with more utilizable acreage have more access to the credit market, while farms with higher average revenue face lower financial constraints.

An operation-type dummy, CATTLE, was included to control for cost sharing targeting preferences. Program regulations prohibit awarding cropping operations unless crops are being used at the initial stage of a rotation previous to establish pasture land. The variable CATTLE takes the value one if the farm held more than one animal unit of dairy or beef cattle.

Two sets of geographic dummies were included as explanatory variables. Excluding the Metropolitan Area (which includes the capital city, Santiago), Chile is divided administratively in twelve regions that are numbered in ascending order from north to south. The phosphorus cost sharing program has concentrated on Regions Seven through Ten. It is well known among Chilean agronomists that soil parent materials change north to south from alluvial and granite materials to volcanic ash. Most of the volcanic soils are located in Regions Nine and Ten. To control for soil properties other than aluminum saturation, a set of four location dummies for Regions Seven, Eight, Nine, and Ten were included in the empirical models. It is expected that probability of being awarded cost-sharing is higher for southern regions since volcanic soils show a particularly strong phosphorus fixation power. Consequently, the current level of phosphorus is expected to decrease north to south (other things equal).

A second set of three location dummies (ANDES, VALLEY and COAST was included. The purpose of these three dummies was to characterize soil limitations to agricultural activities. In Chile the more productive soils are located in the valley between the coastal range next to the Pacific and the Andes Mountains. Soils in the piedmont of the Andes are young volcanic soils with poor chemical properties that limit crop and grass production. Soils on the coastal range, on the other hand, are highly erodible and sensitive to droughts during the dry season (summer in Chile). Thus, it is expected a higher level of cost-share awarding among farmers close to the Andes or located on the coastal range, because they are likely to be more financially constrained than farmers on the Valley and thus a preferred target for program administrators. Finally, a higher level of phosphorus is expected on farms on the Valley given the less restrictive qualities of their soil.

Table 4.2. Dependent and Explanatory variables

Variable	Description	Mean	St. dev.
SHARE	Binary variable indicating whether or not the farm has received cost sharing between 1996 and 2000 inclusively.	0.3626	0.4812
POLSEN	Logarithm of the phosphorus level (measured by the Olsen method) in the plant-available pool.	2.0123	0.7915
AACRE	Acreage usable for agricultural production (10^3 ha)	0.0321	0.0528
REV	Average annual revenue per hectare (10^6 Ch$ ha^{-1})	0.0223	0.0345
AlSat	Percentage of Al saturation in the soil solution	0.8456	0.8180
CATTLE	Farm holds more than 1 animal unit of beef or dairy cattle (yes=1)	0.8474	0.3600
SEVEN	Farm is located in Region Seven (yes=1)	0.0746	0.2630
EIGHT	Farm is located in Region Eight (yes=1)	0.1215	0.3271
NINE	Farm is located in Region Nine (yes=1)	0.1899	0.3926
TEN	Farm is located in Region Ten (yes=1)	0.6140	0.4873
ANDES	Farm is located on the hills at the feet of Andes range (yes=1)	0.4097	0.4923
VALLEY	Farm is located in the valley between the Coast and Andes ranges (yes=1)	0.3160	0.4654
COAST	Farm is located on the Coastal range or in the hills at the eastern side of that range (yes=1)	0.2743	0.4466

The available data include farms that have received funding during one or more years between 1996 and 2000 inclusively. A positive short-run effect of the program is expected on every farm because the program monitors the application of the fertilizer. Hence, the 2001 survey data used in this study is likely to detect some positive effect even on those farms facing no constraint to achieve the HY equilibrium. These data are cross-sectional and thus do not permit investigation of the long run effects. The theoretical model does, however, indicate that the effect of cost-sharing on fertilizer use during the transition period on farms in the LY regime should exceed that on farms in the HY regime, permitting a test of this hypothesis.

In what follows I develop a framework that allows testing the existence of two farm subpopulations with different fertilization strategies as indicated by the theoretical analysis. Simultaneously, I determine whether or not the effect of the program is conditional on the subpopulation a farmer belongs to. This framework also allows for an examination of how well cost sharing funds have been targeted. If two separate groups of farmers do exist and those groups can be distinguished by observable characteristics, then those characteristics can and should be used to determine how cost share funds are allocated. If those groups cannot be distinguished by observable characteristics, then the current allocation strategy of the program may be adequate and an exogenous target level may be a reasonable criterion for determining funding awards.

4.2 The econometric model.

Let y_{2i}^{*} denote the amount of cost-share money allocated to farm i and y_{3i}^{*} be the phosphorus content in the corresponding plant-available pool. A model that allows evaluating the impact of cost sharing on soil phosphorus level is the following

$$\begin{aligned} y_{2i}^{*} &= X_{2i}\beta_2 + \varepsilon_{2i} \\ y_{3i}^{*} &= \alpha y_{2i} + X_{3i}\beta_3 + \varepsilon_{3i} \end{aligned} \qquad (4.9)$$

where the X_{ji} $\left(j=2,3\right)$ are vectors of exogenous explanatory variables, and β_j are parameter vectors to estimate. Cost sharing funding is not an event that is exogenous to farmer decisions since farmers self-select by deciding whether to apply for funding. Consequently, the equations in (4.9) cannot be estimated independently and the correlation between equation disturbances must be allowed to adjust freely during the estimation.

The theoretical model suggests that farmers' fertilization strategies, hence their responses to receiving cost sharing, depend on farm characteristics. To introduce the process of selecting a fertilization strategy (in simpler words: whether to be in the LY regime or in the HY regime), I include the following equation to the equation system (4.9)

$$y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} \qquad (4.10)$$

The unobserved variable y_{1i}^* gives "farmer's propensity to attain the HY equilibrium". Thus, y_{1i}^* is negative or positive depending on whether farmer i faces constraints that reduce his chances to attain the HY equilibrium. Farmers facing no constraint should be closer to the HY phosphorus level so that the receipt of cost share funds should have only a small short-run effect on the level of their phosphorus stocks. By contrast, cost share funds should allow constrained farmers ($y_{1i}^* \leq 0$) to switch from an LY to a HY regime and should thus have a larger long-run effect on the plant-available buffer phosphorus stocks.

The econometric model is now

$$\begin{aligned}
y_{1i}^* &= X_{1i}\beta_1 + \varepsilon_{1i} \\
y_{2i}^* &= X_{2i}\beta_2 + \varepsilon_{2i} \\
y_{3i}^* &= \alpha_3 y_{2i} + X_{3i}\beta_3 + \varepsilon_{3i} \qquad y_{1i}^* < 0 \\
y_{3i}^* &= \alpha_4 y_{2i} + X_{4i}\beta_4 + \varepsilon_{4i} \qquad y_{1i}^* \geq 0
\end{aligned} \qquad (4.11)$$

Model (4.11) is a switching regression model with endogenous switching and fertilization strategy conditional on cost-share funding. Since farmer "propensities", y_{1i}^*, are unobserved, the model has unobserved switching points.

Note that if fertilization decisions are not conditional on farmer constraints, the switching should generate no differential effects on the parameters of the remaining equations. In other words: if a single equilibrium exists, then elements in the parameter vectors β_3 and β_4 should be equal to each other. However, if alternative regimes do exist, then these parameters should be different. I use a Wald test to check equality between the two sets of estimated parameters.

4.3 Implementing a MCEM for a switching regression with unobserved switching points

If distribution assumptions are made for the disturbances then the parameters in the equation system (4.11) can be estimated by maximum likelihood. Before proceeding to the estimation method it is necessary to establish the relation between the dependent variables and its observed counterparts. Two out of the three dependent variables are latent. As discussed previously, the variable y_{1i}^* is fully unobserved and y_{2i}^* is binary. The variable y_{3i}, on the other hand, is observed fully. Thus

y_{1i}^* is unobserved fully $\qquad y_{2i} = \begin{cases} 1 & \text{if } \; y_{2i}^* > 0 \\ 0 & \text{if } \; y_{2i}^* \leq 0 \end{cases} \qquad y_{3i} = y_{3i}^*$

The model in (4.11) is a system of structural equations combining latent and observed variables and, in consequence, the MCEM algorithm introduced in Chapter 2 can be used for its estimation. For estimation purposes the observed counterpart y_{2i} is estimated by the dichotomous variable SHARE and y_{3i} by the logarithm of the continuous variable POLSEN.

Let proceed now with the implementation of the MCEM algorithm for this problem. First, let assume that the disturbance terms in (4.11) are distributed according to

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{pmatrix} \sim N(0,\Sigma), \text{ where } \Sigma = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{12} & 1 & \sigma_{23} & \sigma_{24} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \sigma_{34} \\ \sigma_{14} & \sigma_{24} & \sigma_{34} & \sigma_{44} \end{bmatrix} \tag{4.12}$$

where σ_{33} and σ_{44} have been set equal to one according to the usual normalization required for identification. Then the complete information likelihood function can be written as

$$L^c(\boldsymbol{\alpha},\boldsymbol{\beta},\Sigma \mid \boldsymbol{y}) = \prod_{i,y_{1i}^*<0} \left[\frac{1}{(2\pi)^{3/2} |\Gamma_1| |\Sigma_1|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}_i' \Sigma_1^{-1} \boldsymbol{\varepsilon}_i \right) \right] \times \prod_{i,y_{1i}^* \geq 0} \left[\frac{1}{(2\pi)^{3/2} |\Gamma_2| |\Sigma_2|^{1/2}} \exp\left(-\frac{1}{2} \boldsymbol{\varepsilon}_i' \Sigma_2^{-1} \boldsymbol{\varepsilon}_i \right) \right]$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$, $\boldsymbol{\alpha} = (\alpha_3, \alpha_4)$, $\Sigma_1 = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}$, and

$$\Sigma_2 = \begin{bmatrix} 1 & \sigma_{12} & \sigma_{14} \\ \sigma_{12} & 1 & \sigma_{24} \\ \sigma_{14} & \sigma_{24} & \sigma_{44} \end{bmatrix}.$$

Since Γ_1 and Γ_2 are identity matrices, the complete information log-likelihood function reduces to

$$\ell^c(\boldsymbol{\alpha},\boldsymbol{\beta},\Sigma \mid \boldsymbol{y}) = -\frac{3N}{2}\ln(2\pi) - \frac{1}{2}\sum_{y_{1i}^*<0}\ln|\Sigma_1| - -\frac{1}{2}\sum_{y_{1i}^*\geq 0}\ln|\Sigma_2| - \frac{1}{2}\operatorname{tr}\left(\Sigma_1^{-1}\sum_{y_{1i}^*<0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right) - \frac{1}{2}\operatorname{tr}\left(\Sigma_2^{-1}\sum_{y_{1i}^*\geq 0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right)$$
$$= -\frac{3N}{2}\ln(2\pi) - \frac{1}{2}\sum_{y_{1i}^*<0}\ln|\Sigma_1| - -\frac{1}{2}\sum_{y_{1i}^*\geq 0}\ln|\Sigma_2| - \frac{1}{2}\operatorname{tr}\left(\Sigma_1^{-1}\sum_{y_{1i}^*<0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'} + \Sigma_2^{-1}\sum_{y_{1i}^*\geq 0}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right) \quad (4.13)$$

where N is the total number of observations. Note that the parameter σ_{34} in (4.12) cannot be estimated since there are no observations in both regimes simultaneously. Also note that the last two terms between brackets in (4.13) cannot be written in practice because, as y_{1i}^* is not observed, we do not know the regime in which each observation must be included. This is standard in switching models with unobserved switching points (Dickens and Lang, 1985). In the classical maximum likelihood approach the log-likelihood function for each individual is the weighted sum of the likelihoods of being in each regime, where the weights are the conditional probabilities of being in the respective regimes. In (4.13) we still have a sum of two terms; however, instead of weighting the sum, the idea is to simulate y_{1i}^* as if the individual were in one regime in order to calculate the first sum and then simulate y_{1i}^* as if the individual were in the other regime in order to calculate the second sum. Details are given below when describing the implementation of the Gibbs sampler.

The Expectation step is straightforward from (4.13) and it requires the calculation of

$$\underset{y_{1i}^{*}\le 0}{Q_i}\left(\boldsymbol{\alpha},\boldsymbol{\beta}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)=E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}\le 0\right]=E\left[\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-X_{2i}\beta_2 \\ y_{3i}^{*}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-X_{2i}\beta_2 \\ y_{3i}^{*}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}^{'}\right] \quad (4.14)$$

$$=\sigma_i^{2(m)-}+\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)-}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)-}-X_{2i}\beta_2 \\ y_{3i}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)-}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)-}-X_{2i}\beta_2 \\ y_{3i}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}^{'}$$

Analogously

$$\underset{y_{1i}^{*}>0}{Q_i}\left(\boldsymbol{\alpha},\boldsymbol{\beta}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)=E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}>0\right]=E\left[\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-X_{2i}\beta_2 \\ y_{3i}^{*}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1 \\ y_{2i}^{*}-X_{2i}\beta_2 \\ y_{3i}^{*}-\alpha_3 y_{2i}-X_{3i}\beta_3\end{pmatrix}^{'}\right] \quad (4.15)$$

$$=\sigma_i^{2(m)+}+\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)+}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)+}-X_{2i}\beta_2 \\ y_{3i}-\alpha_4 y_{2i}-X_{4i}\beta_4\end{pmatrix}\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)+}-X_{1i}\beta_1 \\ \mu_{y_{2i}^{*}}^{(m)+}-X_{2i}\beta_2 \\ y_{3i}-\alpha_4 y_{2i}-X_{4i}\beta_4\end{pmatrix}^{'}$$

where

$$\sigma_i^{2(m)-}=\mathrm{Cov}\left(y_{1i}^{*},y_{2i}^{*},y_{3i}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}\le 0\right)$$
$$\sigma_i^{2(m)+}=\mathrm{Cov}\left(y_{1i}^{*},y_{2i}^{*},y_{3i}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}>0\right)$$

and

$$\mu_{y_{ji}^{*}}^{(m)-}=E\left[y_{ji}^{*}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}\le 0\right] \qquad j=1,\dots,3$$

$$\mu_{y_{ji}^{*}}^{(m)+}=E\left[y_{ji}^{*}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y},y_{1i}^{*}>0\right] \qquad j=1,\dots,3$$

Recall that $E\left[y_{ji}^{*}\mid\boldsymbol{\alpha}^{(m)},\boldsymbol{\beta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]$ equals y_{ji} if y_{ji} is observed and must be estimated by Gibbs sampling otherwise.

I replace the M-step by two conditional M-steps. Given the simplicity of expressions (4.14) and (4.15), it is useful to define the vector $\theta = \left(\beta_1, \beta_2, \alpha_3, \beta_3, \alpha_4, \beta_4\right)$. Thus, the first conditional M-step maximizes

$$-\frac{3N}{2}\ln\left(2\pi\right)-\frac{1}{2}\sum_{y_{1i}^*\le 0}\ln\left|\Sigma_1\right|--\frac{1}{2}\sum_{y_{1i}^*>0}\ln\left|\Sigma_2\right|-\frac{1}{2}\operatorname{tr}\left(\Sigma_1^{-1}\sum_{y_{1i}^*\le 0}Q_i\left(\theta\,|\,\theta^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right)-\frac{1}{2}\operatorname{tr}\left(\Sigma_2^{-1}\sum_{y_{1i}^*>0}Q_i\left(\theta\,|\,\theta^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right)$$

with respect to θ conditional on the elements of $\Sigma_1^{(m)}$ and $\Sigma_2^{(m)}$ to produce

$$\theta^{(m+1)} = \left[\sum_{y_{1i}^*\le 0} X_i'\tilde{\Sigma}_1^{(m)}X_i + \sum_{y_{1i}^*>0} X_i'\tilde{\Sigma}_2^{(m)}X_i\right]^{-1}\left[\sum_{y_{1i}^*\le 0} X_i'\tilde{\Sigma}_1^{(m)}\mu_{y_i^*}^{(m)} + \sum_{y_{1i}^*>0} X_i'\tilde{\Sigma}_2^{(m)}\mu_{y_i^*}^{(m)}\right] \quad (4.16)$$

where the matrices in (4.16) are define as

$$X_i = \begin{cases} \begin{bmatrix} X_{1i} & 0 & 0 & 0 \\ 0 & X_{2i} & 0 & 0 \\ 0 & 0 & X_{3i} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \text{if } \; y_{1i}^* \le 0 \\ \begin{bmatrix} X_{1i} & 0 & 0 & 0 \\ 0 & X_{2i} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_{4i} \end{bmatrix} & \text{if } \; y_{1i}^* > 0 \end{cases} \quad , \qquad \mu_{y_i^*}^{(m)} = \begin{cases} \begin{pmatrix} \mu_{y_{1i}^*}^{(m)} \\ \mu_{y_{2i}^*}^{(m)} \\ y_{3i}^* \\ 0 \end{pmatrix} & \text{if } \; y_{1i}^* \le 0 \\ \begin{pmatrix} \mu_{y_{1i}^*}^{(m)} \\ \mu_{y_{2i}^*}^{(m)} \\ 0 \\ y_{3i}^* \end{pmatrix} & \text{if } \; y_{1i}^* > 0 \end{cases}$$

$$\tilde{\Sigma}_1^{(m)} = \begin{bmatrix} \tilde{\sigma}_{11}^{(1,m)} & \tilde{\sigma}_{12}^{(1,m)} & \tilde{\sigma}_{13}^{(1,m)} & 0 \\ \tilde{\sigma}_{12}^{(1,m)} & \tilde{\sigma}_{22}^{(1)} & \tilde{\sigma}_{23}^{(1,m)} & 0 \\ \tilde{\sigma}_{13}^{(1,m)} & \tilde{\sigma}_{23}^{(1,m)} & \tilde{\sigma}_{33}^{(1,m)} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad , \quad \tilde{\Sigma}_2^{(m)} = \begin{bmatrix} \tilde{\sigma}_{11}^{(2,m)} & \tilde{\sigma}_{12}^{(2,m)} & 0 & \tilde{\sigma}_{14}^{(2,m)} \\ \tilde{\sigma}_{12}^{(2,m)} & \tilde{\sigma}_{22}^{(2)} & 0 & \tilde{\sigma}_{24}^{(2,m)} \\ 0 & 0 & 0 & 0 \\ \tilde{\sigma}_{14}^{(2,m)} & \tilde{\sigma}_{24}^{(2,m)} & 0 & \tilde{\sigma}_{34}^{(2,m)} \end{bmatrix}$$

and $\tilde{\sigma}_{ij}^{(1,m)}$ and $\tilde{\sigma}_{ij}^{(2,m)}$ are the elements in the i-th row and j-th column of $\Sigma_1^{(m)-1}$ and $\Sigma_2^{(m)-1}$ respectively. The second conditional M-step then maximizes

$$-\frac{3N}{2}\ln(2\pi)-\frac{1}{2}\sum_{y_{1i}^*\le 0}\ln|\Sigma_1|--\frac{1}{2}\sum_{y_{1i}^*>0}\ln|\Sigma_2|-\frac{1}{2}\operatorname{tr}\left(\Sigma_1^{-1}\sum_{y_{1i}^*\le 0}Q_i\left(\boldsymbol{\theta}^{(m+1)}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right)-\frac{1}{2}\operatorname{tr}\left(\Sigma_2^{-1}\sum_{y_{1i}^*>0}Q_i\left(\boldsymbol{\theta}^{(m+1)}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right)$$

with respect to the elements in Σ to obtain an estimate for $\Sigma^{(m+1)}$.

There remains the implementation of the Gibbs sampler necessary to estimate the matrices Q_i in the objective function. Its construction is straightforward from Section 4.2 of Chapter 2 with obvious substitutions in the formulas for the conditional means $\mu_{ji|i(-j)}$ and variances $\sigma_{j|-j}^2$. To simulate the unobserved "observations" of the dependent variables proceed as follows.

The variable y_{1i}^* is fully unobserved. As we do not know what regime y_{1i}^* belongs to, we have to consider the possibility that y_{1i}^* may belong to either. Consequently, in order to estimate $\sigma_i^{2(m)-}$ and $\mu_{y_{1i}^*}^{(m)-}$ in (4.14) we must sample from a normal distribution with mean $\mu_{1i|i(-1)}^{(m)}$ and variance $\sigma_{1|-1}^{2(m)}$ truncated from above at zero. Analogously, the simulation must be performed from a normal distribution with mean

$\mu^{(m)}_{1i|i(-1)}$ and variance $\sigma^{2(m)}_{1|-1}$ truncated from below at zero when estimating $\sigma^{2(m)+}_{i}$ and $\mu^{(m)+}_{y^*_{1i}}$ in (4.15).

The variable y^*_{2i} is binary. Accordingly, we simulate y^*_{2i} from a normal distribution with mean $\mu^{(m)}_{2i|i(-2)}$ and variance $\sigma^{2(m)}_{2|-2}$ truncated below at zero if y^*_{2i} equals one and truncated above at zero if y^*_{2i} equals zero.

The Gibbs sampler was started with 300 simulations and increased by 15 simulations at every iteration of the MCEM algorithm. The number of dismissed simulations, k_{burn}, was kept constant at 150. The routine converged after 420 iterations. A simulation of size $R = 3400$ and $r_{burn} = 400$ was used for the estimation of the information matrix and asymptotic standard errors. Results are presented in Table 4.3 (location dummies COAST and SEVEN were excluded from the estimation as required for identification). Estimation outcomes are presented in Table 4.3

Table 4.3. Estimates of the switching regression model

Equation	Variable	Estimate	As. st. error	As. t-stat.	P-value
Switching	Constant	-0.0197	0.0395	-0.4986	0.6183
	AACRE	0.7208	0.3809	1.8922	0.0594
	ALSAT	-0.2014	0.0992	-2.0293	0.0430
	REV	0.9573	0.2036	4.7012	0.0000
	VALLEY	0.2639	0.2618	1.0081	0.3139
	ANDES	0.1249	11.6548	0.0107	0.9915
Cost share	Constant	-0.1395	0.2367	-0.5893	0.5559
	REV	-0.9582	0.8294	-1.1553	0.2485
	ALSAT	0.8262	0.3517	2.3495	0.0192
	CATTLE	-0.0463	0.0359	-1.2875	0.1986
	VALLEY	-0.9630	0.4981	-1.9332	0.0542
	ANDES	0.8390	1.5384	0.5454	0.5858

Phosphorus pool level HY regime	Constant	0.7499	0.3258	2.3018	0.0218
	SHARE	0.1082	0.0061	17.8456	0.0000
	REV	1.4109	0.1926	7.3238	0.0000
	EIGHT	-0.2377	0.0601	-3.9531	0.0001
	NINE	-0.3188	0.1427	-2.2343	0.0259
	TEN	-0.7226	0.0472	-15.3075	0.0000
	VALLEY	0.5283	0.1188	4.4462	0.0000
	ANDES	-0.2634	0.0626	-4.2078	0.0000
Phosphorus pool level LY regime	Constant	1.8991	0.3977	4.7752	0.0000
	SHARE	0.8270	0.0546	15.1443	0.0000
	REV	0.3741	0.2935	1.2749	0.2030
	EIGHT	-0.1993	0.0402	-4.9600	0.0000
	NINE	-0.5512	0.2098	-2.6270	0.0089
	TEN	-0.9331	0.1937	-4.8177	0.0000
	VALLEY	0.1430	0.1937	0.7384	0.4606
	ANDES	0.5217	0.6220	0.8387	0.4021
	σ_{12}	-0.0116	0.0125	-0.9255	0.3552
	σ_{13}	0.2633	0.0553	4.7629	0.0000
	σ_{23}	-0.4355	0.0486	-8.9554	0.0000
	σ_{33}	1.0256	0.0260	39.4721	0.0000
	σ_{14}	0.6556	0.0721	9.0881	0.0000
	σ_{24}	-0.2472	0.0149	-16.5854	0.0000
	σ_{44}	1.1550	0.0985	11.7285	0.0000

5. Results

Overall, the signs of the coefficients in Table 4.3 correspond closely to what was expected from the theoretical model. In what follows I identify regime $y_{1i}^* \leq 0$ as the "low yield" regime and regime $y_{1i}^* > 0$ as the "high yield" regime.

5.1. Existence of two subpopulations and two fertilization regimes.

To confirm whether there are or not two fertilization regimes, I use a Wald test to compare the beta coefficients of the two Phosphorus-pool-level equations in Table 4.3. The test provides a value $W = 307.6$ (7 df), which permits rejecting the hypothesis of equality between the two sets of coefficients. The test result confirms that there exist two farm sub-populations following different fertilization regimes.

Results from the switching equation suggest that the larger the share of land usable for agricultural purposes and the higher the revenue per hectare, the higher is the likelihood that a farm belongs to the high yield regime. The evidence is particularly strong for the revenue variable, which confirms the importance of financial condition on fertilization strategy. From the same equation, we have that the more intense is the power of the soil to fix phosphorus (measured by variable ALSAT), the higher the probability of being in the low yield regime. This outcome was expected since a higher level of free aluminum in the soil is a signal of greater depletion and thus of a more costly soil remediation.

The coefficients of the regional dummies in the phosphorus equations are negative and increasing in magnitude north to south. This indicates that phosphorus level decreases as we move to south, which was expected since the presence of volcanic ashes in soil parental materials increases north to south in Chile. Soils in the central valley show more phosphorus in the plant-available pool than those close to the Andes or on the Coastal range, which is an indication of the greater productivity of valley soils. This is evidence of farmers' economic rationality: they fertilize more on soils having greater production potential. However, this is true only if farmers are not financially or technologically constrained. From the results for the equation of the low yield regime, we can see that constrained farmers show no significant differences between the phosphorus levels in valley farms and those in farms located on the Andes piedmont or on the Coastal range.

5.2. Determinants of cost share allocation.

The only statistically significant coefficient in the cost share equation is the coefficient of the ALSAT variable. Since the key funding requirement of the program is being below the target level in the plant-available pool, a positive and significant coefficient of our indicator of phosphorus fixation power was expected. None of the remaining variables in the equation is significantly different from zero, indicating that agronomic considerations alone were used to determine cost share awards, with economic considerations playing no role. This result confirms the stated policy of the cost sharing program. Results from the switching equation suggest that financial condition is an important determinant of fertilization strategy, so that it is feasible to improve targeting using observable characteristics, e.g. using farm revenue to determine awards.

5.3. Effects of cost sharing on fertilization.

The final evidence supporting the theoretical model comes from the comparison of program effects between the two regimes, i.e.

$$\begin{aligned}&\left(E\left[y_{3i} \mid y_{1i} \leq 0, y_{2i}=1\right]-E\left[y_{3i} \mid y_{1i} \leq 0, y_{2i}=0\right]\right)-\left(E\left[y_{3i} \mid y_{1i}>0, y_{2i}=1\right]-E\left[y_{3i} \mid y_{1i}>0, y_{2i}=0\right]\right)\\ &\quad=\alpha_4-\alpha_3+E\left[\varepsilon_{3i} \mid \varepsilon_{1i} \leq -X_{1i}\beta_1, \varepsilon_{2i}>-X_{2i}\beta_2\right]-E\left[\varepsilon_{3i} \mid \varepsilon_{1i} \leq -X_{1i}\beta_1, \varepsilon_{2i} \leq -X_{2i}\beta_2\right]-\\ &\quad E\left[\varepsilon_{3i} \mid \varepsilon_{1i}>-X_{1i}\beta_1, \varepsilon_{2i}>-X_{2i}\beta_2\right]+E\left[\varepsilon_{3i} \mid \varepsilon_{1i}>-X_{1i}\beta_1, \varepsilon_{2i} \leq -X_{2i}\beta_2\right]\end{aligned} \tag{4.17}$$

where the last expectations can be written as function of the standard normal pdf and the conditional bivariate normal cdf. After calculating (4.17) for every individual in the sample and taking the average, we obtain an estimate of 0.415 with an asimptotic standard error calculated using the delta method of 0.133. This estimate is positive and different from zero at a 1% significance level. This result supports the hypothesis that cost sharing has a greater effect on the phosphorus level of constrained farmer who would be in the LY regime in the absence of cost sharing.

Overall, these results support the idea that the threshold separating the two sub-populations is actually determined endogenously and that the use of an exogenously determined target phosphorus level should not be used as the sole cost share allocation criterion.

Program administrators should care more about determinants of farmer behavior at the moment to allocate the program budget. By identifying the characteristics that makes a farmer more likely to be in one regime or in the other it would be possible to increase program efficiency by a more careful targeting. The way targeting should be improved, however, is not an easy problem. Determining the actual financial situation of applicants requires collecting sensitive information, which would make the process more complicated and require extra paperwork. As a result we may find that the additional transaction costs of application may discourage participation on the part of farmers for whom cost sharing would do the most good, i.e. financially constrained farmers with low managerial and technical skills.

6. Final remarks.

This chapter analyzed the targeting policy of a soil remediation program aiming to replenish phosphorus fertility of Chilean soils. By using an optimal-control approach it was shown that, depending on soil and farmer characteristics, a farmer may choose between two long-run phosphorus fertilization paths: 1) apply low rates of fertilization and sustain a low yield level of phosphorus in the soil, or 2) apply high fertilization rates initially and then attain and maintain a high yield level of phosphorus in the soil pools.

The existence of two possible equilibria, which are endogenously determined, questions the suitability of using an exogenous target level as the primary determinant in the allocation policy of a fertility remediation program. Thus, if the target is set too high with respect to the individual HY phosphorus level, at least part of cost sharing money becomes a net transfer. On the other hand, if the target is set too low, cost sharing may cause only short run effects and the farmer will move back to the LY equilibrium level in the long run.

An empirical cross-sectional evaluation of the model confirmed that two fertilization regimes do exist. As predicted by the theory, financial conditions are important determinants of the fertilization path followed by the farmer and they must be considered in the targeting policy of the program

Chapter 5: Conclusions

1 Introduction

The study carried out in the previous chapters has two components. The first one is methodological (Chapter 2) and it formulates a computational approach to estimate models involving unobserved information. The second component (chapters 3 and 4) evaluated two conservation cost-sharing programs. Chapter 3 presented an empirical study that uses the estimation method implemented in Chapter 2 to analyze multivariate responses to programs that cost share adoption of conservation practices in Maryland. Chapter 4 analyzes both theoretically and empirically the targeting policy of a soil remediation program that provides cost share for recovering phosphorus fertility in Chilean soils. In what follows, I summarize the main conclusions of this work and discuss some areas of potential research.

2. A Monte Carlo EM algorithm for estimating equation systems with linear latent structures.

The methodological contribution of this work is a Monte Carlo EM (MCEM) algorithm suitable for efficient estimation of systems of equations having a linear-in-parameter latent structure. Although algorithms like this have been formulated before for models containing specific types of latent variables, this study has extended the method to handle potentially any type of missing information and to include the estimation of structural models. The combination of Gibbs sampling in the E-step and sequential maximization in the M-step permits circumventing high-dimensional integration and makes the algorithm more robust to problems of “fragile” identification and starting values. Furthermore, the combination of Gibbs sampling and Louis method allows an easy estimation of the information matrix without the problems associated to the use of finite-difference Hessians.

The MCEM algorithm formulated here revealed very robust when analyzing multivariate responses on cross-sectional data. An obvious extension of the method would be its formulation for panel data including probit and tobit panel models. In these cases, individual heterogeneity can be treated as one more latent variable and thus be simulated via Gibbs sampling. A more advanced extension of the method would be its application to dynamic panel models with latent response variables. Dynamic panel models with latent structures are hard to estimate because not always all the elements in the string $\left(y_{i1}, y_{i2}, ..., y_{it-1}, y_{it}, y_{it+1}, ... y_{iT}\right)$ are observed. For instance, if the variable y_{it} is censored from below, it is frequent that some of the string elements are missing because y_{it} fall below the censoring threshold for some values of t. Current available methods eliminate observations having some of the y_{it} unobserved in order to make the estimation feasible. This waste of information would not happen under a MCEM approach since the string can be "repaired" by simulation.

3 Effect of cost sharing on conservation effort and cropping expansion.

Chapter 3 analyzed the interactions between conservation and cropping decisions in a disaggregated multivariate framework. Several contributions to literature emerged from the analysis. First, it was shown that adoption of soil conservation practices, particularly land-quality augmenting ones, expands cropping both at the intensive and extensive margins. Second, it was provided evidence showing that farms awarded cost share funding implement land-productivity improving practices on a greater share of land than farms that do not receive cost sharing. Third, it was shown the existence of a perverse effect of cost sharing payments: since farmers choose to implement preferentially those practices providing private benefits, the expansion in cropping induced by cost-sharing this type of practices reduces the share of land covered by practices that provide public goods.

Cropping expansion at both the extensive and the intensive margins resulting from adoption of land-quality augmenting practices may imply more agrichemical usage. Additionally, if these practices are implemented on sensitive land, the generation of additional nutrient and soil runoff is possible. Thus, the environmental balance of cost sharing these type of conservation practices seems no clear, especially if we consider the effects on scenery, wildlife, and water quality that substituting grassland and wildlife habitat for cropland may have (particularly on small farms). Cost-sharing payments for practices providing off-farm benefits may need to be increased in order to make them comparable to the opportunity costs of allocating land to them. Simultaneously, program administrators should consider more severe restrictions to cropping expansion and reconsider whether the adoption of land-quality augmenting practices needs to be cost-shared.

Since Chapter 3 is the first study that analyzes cost sharing programs from a multivariate point of view, its scope can be extended in a number of ways in order to get a better picture of the effects of working-land conservation payments. Fist of all, similar analyses can be carried out for sets of practices targeting types of non-point source pollution other than soil runoff. Examples include conservation practices used in nutrient, pest, and animal waste management. More elaborated extensions of this work involve studying short and long run effects of cost sharing payments. Dynamic analysis would allow estimating the effect of cost sharing on the persistence in the use of conservation practices, and how faster is the rate of adoption of conservation practices among program participants when compared with farmers receiving no funding. Other issues possible to investigate using dynamic panel models is determining which practices need cost-sharing as "seed money" and which need a more sustained support in order to guarantee its adoption. Finally, some practices may act as starters of the adoption process, i.e. their adoption may trigger the subsequent adoption of related practices. Cost sharing programs may benefit greatly if "seed" practices are identified because those practices could be awarded preferentially in order to start an adoption sequence. Multivariate dynamic analyses could be useful in detecting such practices (although the econometrics does not look easy in perspective).

4 The rationality of using and exogenous target level in soil fertility remediation programs.

In Chapter 4 I discussed the convenience of using an exogenous target level as the central component of the targeting policy of a soil remediation program. I showed theoretically that, conditional on whether a farm is technologically or financially constrained, farmers may follow either a low-yield phosphorus fertilization path or a high-yield fertilization path. Additionally, I gave theoretical evidence that provision of funding conditional only on being below an exogenously determined target level is inefficient since individual high-yield optimum may be lower than the target, in which case cost sharing becomes a transfer. Alternatively, if the target level is below the high-yield optimum, then cost sharing could fail in providing enough funds to move the farmer on a stable path towards the high yield equilibrium. Predictions from the theoretical model were validated using cross-sectional data from a program that cost-share the recovery of phosphorus fertility in Chilean soils. The econometric estimation confirmed the existence of two fertilization paths, where following one or the other depended on farm's financial condition. Although a greater positive effect of the program was detected on financially constrained farms, data limitations do not allow analyzing the persistence of these effects. Panel data and dynamic analyses are needed to disentangle short and long run effects of the program and determine on which group of farmers (if any) program effects persist in the long run.

Bibliography

Abelson, P.H., and J. W. Rowe. 1987. "A new agricultural frontier". Science. 235:1450-1451

Amemiya, T. 1974. "Multivariate regression and simultaneous equation models when the dependent variables are truncated normal". Econometrica, 42 999-1012.

Barrow, N. J. 1983a. "A mechanistic model for describing the sorption and desorption of phosphate by soil". Journal of Soil Science. 34: 733-750.

Barrow, N. J. 1983b. "On the reversibility of phosphate sorption by soils". Journal of Soil Science. 34: 751-758.

Batie, Sandra S. "Green Payments as Foreshadowed by EQIP". Staff Paper 99-45, Department of Agricultural Economics, Michigan State University, August 1999.

Booth, J. And Hobert, J. 1999. "Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm". J. R. Statist. Soc. B. 61 Part 1, pp. 265-285

Börsch-Supan, A. and V. Hajivassiliou. 1993. "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models". J. of Econometrics 58: 347-368.

Boyles, R. 1983. “On the convergence of the EM algorithm”. J. R. Statist. Soc. B, 45(1) 47-50.

Broyden, C.G. 1970. "The Convergence of a Class of Double-Rank Minimization Algorithms," Journal of the Institute of Mathematics and Its Applications, Vol. 6, pp. 76-90.

Buresh, R, Smithson, P., and D. Hellums. 1997. “Building soil phosphorus capital in Africa. In Replenishing soil fertility in Africa”. Soil Science Society of America, Special Publication Number 51. ICRAF. Madison, WI. 251p.

Casella, G. and George, E. 1992. “Explaining the Gibbs Sampler”. The American Statistician. 46(3): 167-174.

Caswell, Margriet F. and and David Zilberman. 1986. “The Effects of Well Depth and Land Quality on The Choice Of Irrigation Technology”. American Journal of Agricultural Economics 68: 798-811.

Cattaneo, Andrea. 2003. “The Pursuit of Efficiency and Its Unintended Consequences: Contract Withdrawals in the Environmental Quality Incentives Program”. Review of Agricultural Economics 25(2): 449-469.

Chan, K. And Ledolter, J. 1995. “Monte Carlo EM estimation for time series models involving counts”. J. Am. Statist. Assoc., 90(429) 242-252.

Clark, C. 1990. “Mathematical Bioeconomics; the Optimal management of Renewable Resources”, Second Ed. Pure and Applied Mathematics, a Wiley-Interscience Series of Texts, Monographs & Tracts, New York. 386 p.

Classen, R., Breneman, V., Bucholtz, S., Cattaneo, A., Johansson, R. and Morehart, M. 2004. “Environmental Compliance in U.S. Agricultural Policy. Past Performance and Future Potential”. Economic Research Service, USDA, Agricultural Economic Report Number 832, June 2004.

Coleman, T.F. and Li, Y. 1996. "An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds," SIAM Journal on Optimization, Vol. 6, pp. 418-445.

Cooper, Joseph C. “A Joint Framework for Analysis of Agri-Environmental Payment Programs”. American Journal of Agricultural Economics 85(November 2003): 976-987.

Dempster, A.P., Laird, N.M. and Rubin, D.B. 1977. “Maximum Likelihood Estimation from incomplete observations”. J. Roy. Statist. Soc. B 39:1-38.

Devroye, Luc. 1986. “Non-uniform random variate generation”. Springer-Verlag. 843 pp

Dickens, W. and Lang, K. 1985. “A Test of Dual Labor Market Theory”. The American Economic Review 75(4): 792-805.

Eickhoff, J. Zhu, J. Amemiya, Y. 2004. “On the simulation size and the convergence of the Monte Carlo EM algorithm via likelihood-based distances”. Statistics & Probability Letters 67: 161-171.

Escudey, M., Gerardo Galindo, Juan E. Förster , Margarita Briceño, Patricia Diaz, and Andrew Chang. 2001. “Chemical Forms of Phosphorus ff Volcanic Ash-Derived Soils in Chile”. Communications in Soil Science and Plant Analysis 32(5-6): 601-616.

Espinosa, J. 1991. “Efecto residual de fósforo en Andisoles”. Revista de la Facultad de Agronomía. Maracay, Venezuela. 17:39-47.

Espinosa, J. 1992. “Phosphorus diagnosis and recommendations in volcanic ash soils”, pages 109-115. In Proceedings of the Tropsoils Phosphorus Decision Support System Workshop, College Station, TX, 11-12 Mar. 1992. TropSoils Bulletin 92-01. Dep. of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources, Univ. of Hawaii, Honolulu.

Fletcher, R. 1970. "A New Approach to Variable Metric Algorithms," The Computer Journal, Vol. 13, pp. 317-322

Friedman, N.; Ninio, M.; Pe'er, I.; and Pupko, T. 2002. "A Structural EM algorithm for Phylogenetic Inference". Journal of Computational Biology 9(2): 332-353.

Gaetan, C. and Yiao, J. 2003. "A multiple-imputation Metropolis version of the EM algorithm". Biometrika 90(3): 643–654.

Geman, S. and Geman, D. 1984. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". IEEE Transactions on Pattern Analysis and Machine Intelligence, 6: 721-741.

Genz, A. 2004. "Numerical computation of rectangular bivariate and trivariate normal and t probabilities". To appear in Statistics and Computing.

Geweke, J.; M. Keane, and D. Runkle. 1994. "Alternative Computational Approaches to Inference in the Multinomial Probit Model". Review of Economics and Statistics, 76: 609-632.

Gilbert, G., Knight, J., Cance, C., and D. Allan. 1999. "Acid phosphatase activity in phosphorus-deficient white lupin roots". Plant, cell and environment. 22(7):801-810.

Goldfarb, D., "A Family of Variable Metric Updates Derived by Variational Means," Mathematics of Computation, Vol. 24, pp. 23-26, 1970.

Greene, William H. “Econometric Analysis”, Fourth ed. Prentice Hall. 2000, 1004 pp.

Guo, S.W. and Thopson, E.A. 1992. “A Monte Carlo method for combined segregation and linkage analysis”. Amer. J. Hum. Genet. 51: 1111-1126.

Halton, J.H. 1960. “On the eficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals” Numer. Math. 2 (1960), 84-90.

Heckman, J. 1976. “Simultaneous equation models with continuos and discrete endogenous variables and structural shifts”. In S.M. Goldfeld and R. E. Quandt (eds.), Studies in non-linear estimation. Cambridge. 278 p.

Hisinger, P., and R. J. Gilkes. 1996. “Mobilization of phosphate from phosphate rock and alumina-sorbed phosphate by the roots of ryegrass and clover as related to rhyzosphere pH”. European Journal of Soil Science. 47: 533-544.

Ibrahim, J.G., Chen, M., and Lipsitz, S.R. 2001. “Missing Responses in Generalized Linear Mixed Models when the Missing Data Mechanism is Nonignorable”. Biometrika, 88(2): 551-564.

Inoue, J. and Tabushi, K. 2003. "A Generalization of the Deterministic Annealing Em Algorithm by means of Non-Extensive Statistical Mechanics". International Journal of Modern Physics B 17(29): 5525-5539

Jank, W. 2004. "Quasi-Monte Carlo Sampling to improve the Efficiency of Monte Carlo EM". Forthcoming at Computational Statistics and Data Analysis.

Karpinets, T.V., Greenwood, D.J., and Ammons, J.T. 2004. "Predictive mechanistic model of soil-Phosphorus dynamics with readily available inputs". Soil Science Society of America Journal. 68, 644-653.

Keane, Michael P. 1992. "A Note on Identification in the Multinomial Probit Model". Journal of Business and Economic Statistics 10: 193-200.

Kendall, M. and and Stuart, A. 1978. "The advanced Theory of statistics". McMillan, New York.

Kennedy, J.O. 1986. "Rules for Optimal Fertilizer Carryover". Review of Marketing and Agricultural Economics, 54: 3-10.

Klonsky, Karen and Florence Jacquet. "How Well Do Green Payments Fit into the Green Box? Lessons from EQIP". Paper presented at the international conference

Agricultural Policy Reform and the WTO; Where are we Heading?, Capri (Italy), June 23-26, 2003.

Kremen, C, J. O. Niles, M. G. Dalton, G. C. Daily, P. R. Ehrlich, J. P. Fay, D. Grewal, and R. P. Guillery. 2000. “Economic Incentives for Rain Forest Conservation across Scales”. Science 288(5472): 1828-1832.

Lewandrowsky, Jan; Mark Peters; Carol Jones; Robert House; Mark Sperow; Marlen Eve and Keith Paustian. 2004. “Economics of Sequestering Carbon in the US Agricultural Sector”. Technical Bulletin No. 1909, Economic Research Service, US Department of Agriculture, April 2004.

Liao, J.G. 1998. “Variance reduction in Gibbs sampler using quasi random numbers”. Journal of computational and graphical statistics 7(3): 253-266.

Lichtenberg, Erik. 1989. "Land Quality, Irrigation Development and Cropping Patterns in the Northern High Plains". American Journal of Agricultural Economics 71: 187-194.

Lichtenberg, Erik. 2004a. “Are Green Payments Good for the Environment?” Agricultural and Resource Economics Review 33(1), 24-33

Lichtenberg, Erik. 2004b. “Cost-Responsiveness of Conservation Practice Adoption: A Revealed Preference Approach”, Journal of Agricultural and Resource Economics 29(3), 420-435.

Lichtenberg, E. and Smith-Ramírez, R. 2004. “Cost Sharing Transaction Costs and Conservation”. Mimeo, Agr. and Res. Ec. Dep., University of Maryland, College Park.

Louis, T.A. 1982. “Finding the Observed Information Matrix when using the EM Algorithm”. J. Roy. Statist. Soc. B 44:226-233.

Lu, Y and Stocking, M. 2000. “Integrating biophysical and socio-economic aspects of soil conservation on the Loess Plateau, China. Part II. Productivity impact and economic costs of erosion”. Land Degradation & Development 11(2): 141-152.

Ludwick, A. 2002. “The hills and valleys of phosphorus fixation”. Potash and Phosphate Institute. Available at http://www.ppi-far.org/ppiweb/agbrief.nsf/$webcontents.

Lynch, Lori and Robert T. Tjaden. 2004. “When a Landowner Adopts a Riparian Buffer – Benefits and Costs”. Maryland Cooperative Extension Fact Sheet 774. Available at www.riparianbuffers.umd.edu/PDFs/FS774.pdf. Accessed on January 10 2005.

Maddala, G. 1983. “Limited-dependent and qualitative variables in econometrics”. Econometric Society Monographs. 401 p.

Malik, Arun S. and Robbin A. Shoemaker. 1993. “Optimal Cost-Sharing Programs to Reduce Agricultural Pollution”. Technical Bulletin No. 1820, Economic Research Service, US Department of Agriculture, June 1993.

McConnell, K.E. 1983. “An Economic Model of Soil Conservation”. Amer. J. Agr. Econ., February 1983: 83-89.

McCulloch, C. 1994. “Maximum likelihood variance components estimation for binary data”. J. Am. Statist. Assoc., 89 pp. 330-335.

McCulloch, C. 1997. “Maximum likelihood algorithms for generalized linear mixed models”. J. Am. Statist. Assoc., 92 pp. 162-170.

McFadden, D. 1989. “A Method of Simulated Moments for estimation of Discrete Response Models without Numerical Integration”. Econometrica 57: 995-1026.

Meng, X. And Rubin, D. 1993. “Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika, 80 267-278.

Moler, C. 1995. “Random thoughts; 10435 years is a very long time”. Matlab News and Notes, Fall 12-13.

Nanzyo, M., Dahlgren, R. and S. Shoji. 1997. “Chemical characteristics of volcanic soils”. In Replenishing soil fertility in Africa. Soil Science Society of America, Special Publication Number 51. ICRAF. Madison, WI. 251p.

Natarajan, Ranjini; McCulloch, Charles E. and Nicholas Kiefer. 2000. “A Monte Carlo EM Method for estimating Multinomial Probit Models”. Computational Statistics and Data Analysis 34: 33-50.

Nelson, F. and Olson, L. 1978. “Specification and estimation of a simultaneous-equation model with limited dependent variables”. Int. Ec. Rev., 19(3) 695-709.

Orchard, T. and Woodbury, M.A. 1972. “A Missing Information Principle: Theory and Applications”. In Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Vol. 1. Berkeley, California. University of California Press, pp. 697-715.

Roberts, Michael and Ruben N. Lubowski. 2004. “Enduring Benefits of Land Retirement Policies: Evidence from the Conservation Reserve Program”. Economic Research Service, U.S. Department of Agriculture. April 2004.

Rosenbaum, S. 1961. “Moments of a Truncated Bivariate Normal Distribution”. Journal of the Royal Statistical Society B, 23(2): 405-8.

Rowell, D.L. 1994. “Soil science. Methods and applications”. Longman Sci. and Tech., Essex, UK.

Schnitkey, G.D., Hopkins, J.W., and Tweeten, L.G. 1996. “An Economic Evaluation of Precision Fertilizer Applications on Corn-Soybean Fields”. In: Proceedings of the Third International Conference on Precision Agriculture, eds. P. C. Robert. H. R. Rust and W. E. Larson, pp. 977-87. Conf. held in Minneapolis MN, 23-26 June 1996. Madison WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America.

Shanno, D.F. 1970. "Conditioning of Quasi-Newton Methods for Function Minimization," Mathematics of Computation, Vol. 24, pp. 647-656.

Sharpley, A., Daniel, T., Lemunyon, R. and R. Parry. 1999. “Agricultural phosphorus and euthrophication”. USDA, Agricultural Research Service. ARS-149.

Smith-Ramírez, R. 1999. “Sistemas productivos lecheros en Chile”. In: Competitividad de la producción lechera nacional. Facultad de Ciencias Agrarias, Universidad Austral de Chile, Valdivia, Chile. 75 p.

Smith-Ramírez, R., Moreira, V. y Latrille, L. 2002. “Caracterización de Sistemas Productivos Lecheros en la X Región de Chile Mediante Análisis Multivariable”. Agricultura Técnica 62(3): 375-395

Sobol, I.M. 1998. “On quasi-Monte Carlo integration”. Math. Comput. Simulation 47 103-112.

Soule, Meredith J., Abebayehu Tegene, and Keith D. Wiebe. 2000. “Land Tenure and the Adoption of Conservation Practices”. American Journal of Agricultural Economics 82: 993-1005.

Sowden, R. And Ashford, J. 1969. “Computation of the bi-variate normal integral”. Applied Statistics, 18(2) 169-180.

Steck, G.P. 1958. “A Table for Computing Trivariate Normal Probabilities”. Ann. Math. Statisc. 29: 780-800.

Ueda, N. and Nakano, R. 1998. “Deterministic annealing EM algorithm”. Neural Networks. 11: 271-282.

Van der Zaag, P., and C. Kagenzi. 1986. “The phosphorus requirements of five consecutive potato crops on an Andept in Rwanda”. American Potato Journal. 63:121-129.

Vukina, T., Armando Levy, and Michele Marra. 2003. "Do Farmer Value the Environment? Evidence from the Conservation Reserve Program Auctions". Available at http://www2.ncsu.edu/unity/lockers/users/v/vukina/Crp_JEEM1.pdf. Accessed on January 11 2005.

Wei, C. and Tanner, M. 1990. "A Monte Carlo implementation of the EM algorithm and the Poor Man's Data Augmentation Algorithms". J. Am. Statist. Assoc., 85(411) 699-704.

World Trade Organization. "Agreement on Agriculture". 1994. Available at http://www.wto.org/ english/docs_e/legal_e/14-ag.pdf . Accessed on September 10 2004

Wu, C.F.J. 1983. "On the convergence properties of the EM algorithm". The Annals of Statistics. 11(1): 95-108.

Wu, JunJie. 2000. "Slippage Effects of the Conservation Reserve Program". American Journal of Agricultural Economics 82:979-992.