NÚMERO 403

RICARDO SMITH RAMÍREZ

FIML Estimation of Treatment Effect Models with Endogenous Selection and Multiple Censored Responses via a Monte Carlo EM Algorithm

DICIEMBRE 2007

www.cide.edu

Las colecciones de **Documentos de Trabajo** del **CIDE** representan un medio para difundir los avances de la labor de investigación, y para permitir que los autores reciban comentarios antes de su publicación definitiva. Se agradecerá que los comentarios se hagan llegar directamente al (los) autor(es).

Fax: 5727•9800 ext.6314
Correo electrónico: publicaciones@cide.edu
www.cide.edu

Producción a cargo del (los) autor(es), por lo que tanto el contenido así como el estilo y la redacción son su responsabilidad.

Abstract

We formulate a Monte Carlo EM algorithm to estimate treatment effect models involving multiple censored responses. The algorithm has at least three advantages with respect to traditional methods. First, it does not require integrating the unobserved information out from the likelihood function, which reduces the estimation time dramatically and permits to solve problems involving a high number of latent variables. Second, it reduces the estimation of the vector of slopes to the calculation of a GLS estimator, and numerical techniques are required only to estimate the elements in the disturbance covariance matrix. Third, it has low sensitivity to the selection of starting values and fragile identification. We illustrate the method by estimating a 3-equation treatment model; then we compare the performance of our algorithm against a quasi-Newton optimization that uses numerical integration.

Resumen

Este artículo presenta un algoritmo Monte Carlo EM para estimar modelos de tratamientos con respuestas censuradas múltiples. El algoritmo tiene al menos tres ventajas respecto a métodos tradicionales. Primero, no requiere integración numérica, lo cual reduce dramáticamente el tiempo de estimación. Segundo, la estimación del vector de pendientes se reduce al cálculo de un estimador GLS, y métodos numéricos son requeridos solamente para estimar los elementos en la matriz de covarianzas de los errores. Tercero, el método tiene una baja sensibilidad a la selección de valores iniciales y a identificación frágil. Como ilustración, el algoritmo es aplicado a la estimación de un modelo de tratamiento de tres ecuaciones. Luego, las propiedades del algoritmo son comparadas contra técnicas que combinan integración numérica y métodos de optimización tradicionales.

Introduction

Usually, a treatment effect model includes a single participation (selection) equation and one or more response equations, where the last ones include an endogenous dummy for participation. All the equations are linked through the correlations between their error terms. When only one response exists, the two-equation system can be estimated either by the 2-step Heckman method or by full information maximum likelihood (FIML), being the later more efficient than the former. Multi-objective programs, however, involve not one but several individual responses that program administrators aim to influence. In cases where individuals determine jointly the optimal levels of the different responses, a multivariate framework handling all the responses simultaneously is more adequate than analyzing each response variable independently (Dorfman, 2001; Cooper, 2003). FIML estimation of a multivariate treatment model can be challenging for two reasons. First, responses are often limited dependent or discrete, which requires the inclusion of latent variables in the model. Second, most optimization algorithms are sensitive to the selection of starting values and problems of "fragile identification" when latent variables are involved. We discuss these issues in the next paragraphs.

Systems of equations with latent structures are abundant in applied economics literature. Typically, econometricians use latent structures when dealing with variables that are partially observed but which behavior can be modeled by assuming the existence of a continuous unobserved counterpart that behaves linearly. Thus, latent variables allow us handling dichotomous, polytomous, or censored variables without much trouble. Nonetheless, the use of latent variables can be very costly computationally or even make the estimation unfeasible in a multivariate framework. If the model is going to be estimated by FIML, then both the number and the dimensionality of the integral terms in the likelihood function increase with the number of latent variables in the model. High dimensional integration slows the estimation down and it can even make the estimation impossible.

A second and recurrent issue in FIML estimation is the selection of starting values in order to initiate the optimization routine. OLS estimates are the most frequent choice for slope vector; yet, they are biased since the dependent variable is not observed fully and, in some models, it is not observed at all. The main challenge in a multivariate framework, however, is not finding starting values for the slopes but for the elements in the disturbance covariance matrix. When the starting values are not in the approximation area of the optimum, commercial routines frequently cannot identify the covariance terms; the usual evidence of this problem is that algorithms cannot keep the covariance estimates in the parameter space (*e.g.*

the magnitude of the correlation coefficients are greater than one, or the covariance matrix is not longer positive definite).

Finally, a third 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 since line-search methods cannot progress to the optimum.

For equation systems involving three or more latent variables, FIML procedures that use quadrature integration are often impractical. This "curse of dimensionality" has, however, been partially overcome in the last decade by the use of probability simulators (McFadden, 1989; Börsch-Supan and Hajivassiliou, 1993; Geweke *et al.*, 1994) and Monte Carlo and Quasi-Monte Carlo integration methods (Sobol, 1998). Yet, the focus of these approaches is to make the integration of the likelihood function feasible, while the issues of starting values and fragile identification remain.

Instead of placing the attention on calculating the integrals in the likelihood function, a Monte Carlo EM (MCEM) algorithm focuses on the latent continuum underlying the observed information. Similar to its deterministic version (the EM algorithm), a MCEM algorithm executes two steps iteratively: the expectation step and the maximization step. In the expectation step, a Gibbs sampler circumvents the integration problem by imputing the unobserved information (Casella and George, 1992). Then, after the continuum has been "restored", the maximization step does not differ much from maximizing the likelihood function of a set of seemingly unrelated regressions (SUR). We proceed sequentially. First, the vector of slopes is obtained from the generalized linear squares (GLS) estimator of a modified SUR model. Then, the elements in the disturbance covariance matrix are estimated numerically and conditional on the slopes calculated previously. We show that this procedure reduces dramatically the sensibility of the algorithm to the selection of starting values and fragile identification.

The remaining of this article is organized in the following way. The next section formulates the MCEM algorithm and exemplifies how it works by estimating a system of three equations and three latent variables. The second section solves the same problem by numerical integration. The outputs of both approaches are then compared focusing on the sensibility to starting values of the two methods. The only reason to use a 3-equation model is to allow comparing between the two approaches; the extension to higher dimensions of the MCEM algorithm is straightforward, but making traditional numerical integration feasible is not. The third and last section gives final remarks.

1. The Monte Carlo Expectation-Maximization (MCEM) algorithm

The Expectation-Maximization (EM) algorithm as an iterative procedure to compute maximum likelihood estimates when "… observations can be viewed as incomplete data" Dempster *et al*. (1977). To give a flavor of how the algorithm works consider the many-to-one mapping $z \in Z \rightarrow y = y(z) \in Y$. The information z in Z is not observed directly but through its observed realization y in Y. In words, z is only know 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 $(\boldsymbol{y}, \boldsymbol{z})$, where $\boldsymbol{z}$ is the unobserved information. Then the log-likelihood function of the observed information can be written as

$$\ell(\theta \mid \boldsymbol{y}) = \ln L(\theta \mid \boldsymbol{y}) = \ln \int_{Z(y)} L(\theta \mid \boldsymbol{y}, \boldsymbol{z})\, d\boldsymbol{z} \qquad (1)$$

The integrals present in (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 (1) directly, the EM algorithm focuses on the complete-information log-likelihood $\ell^c(\theta \mid \boldsymbol{y}, \boldsymbol{z})$ and maximizes $E\left[\ell^c(\theta \mid \boldsymbol{y}, \boldsymbol{z})\right]$ by executing iteratively two steps. The first one is called Expectation step or E-step, which at iteration *m*+1 computes $Q\left(\theta \mid \theta^{(m)}, \boldsymbol{y}\right) = E\left[\ell^c(\theta \mid \boldsymbol{y}, \boldsymbol{z})\right]$, where $E\left[\ell^c(\theta \mid \boldsymbol{y}, \boldsymbol{z})\right]$ is the expectation (on $\boldsymbol{z}$) of the complete-information log-likelihood conditional on the observed information and provided that the conditional density $f\left(\boldsymbol{z} \mid \boldsymbol{y}, \theta^{(m)}\right)$ 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 avoids troublesome computations in the E-step by imputing the unobserved information conditional on what is observed and 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 $f\left(\boldsymbol{z} \mid \theta^{(m)}, \boldsymbol{y}\right)$ (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.

1.1. Implementing the Monte Carlo EM algorithm

To formulate the MCEM algorithm consider the *k*-equation system

$$\begin{aligned} y_{1i}^{*} &= x_{1i}\beta_1 + \varepsilon_{1i} \\ y_{2i}^{*} &= \gamma_2 y_{1i} + x_{2i}\beta_2 + \varepsilon_{2i} \\ &\vdots \\ y_{ki}^{*} &= \gamma_k y_{1i} + x_{ki}\beta_k + \varepsilon_{ki} \end{aligned} \qquad (2)$$

Where:

$$y_{1i} = \begin{cases} 1 \text{ if } y_{1i}^{*} > 0 \\ 0 \text{ if } y_{1i}^{*} \le 0 \end{cases} \qquad y_{ji} = \begin{cases} b_j & \text{if } y_{ji}^{*} \ge u \\ y_{ji}^{*} & \text{if } l < y_{ji}^{*} < u \\ a_j & \text{if } y_{ji}^{*} \le l \end{cases} \qquad j = 2,...k$$

i.e. y_{1i} is dichotomous and variables y_{ji} ($j = 2,...k$) are censored from below at a_j and from above at b_j (in case a response is not censored, just consider $a_j = -\infty$ and $b_j = \infty$). Thus, the variables y_{ji}^{*} are not observed fully and the problem can be considered as one with incomplete data; therefore, an EM approach applies. The disturbance terms in (2) are assumed to follow a *k*-variate normal distribution $N(0,\Sigma)$ with covariance

$$\Sigma = \begin{pmatrix} 1 & \sigma_{\varepsilon_1\varepsilon_2} & \cdots & \sigma_{\varepsilon_1\varepsilon_k} \\ \sigma_{\varepsilon_1\varepsilon_2} & \sigma_{\varepsilon_2}^2 & \cdots & \sigma_{\varepsilon_2\varepsilon_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\varepsilon_1\varepsilon_k} & \sigma_{\varepsilon_2\varepsilon_k} & \cdots & \sigma_{\varepsilon_k}^2 \end{pmatrix} \qquad (3)$$

Where $\sigma_{\varepsilon_1}^2 = 1$ is the usual restriction to ensure identification of the coefficients in an equation with a dichotomous dependent variable.

The complete-data likelihood function (i.e. as if the latent variables were observed) for the equation system is:

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

Where $\boldsymbol{\theta} = (\beta_1, \gamma_2, \beta_2, \cdots, \gamma_k, \beta_k)$, and $\boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \vdots \\ \varepsilon_{ki} \end{pmatrix} = \begin{pmatrix} y_{1i}^{*} - X_{1i}\beta_1 \\ y_{2i}^{*} - \gamma_2 y_{1i} - X_{2i}\beta_2 \\ \vdots \\ y_{ki}^{*} - \gamma_k y_{1i} - X_{ki}\beta_k \end{pmatrix}$

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

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

$$E\left[\ell^c\left(\boldsymbol{\theta},\Sigma\mid \boldsymbol{z}\right)\right]=-\frac{kN}{2}\ln\left(2\pi\right)-\frac{N}{2}\ln\left|\Sigma\right|-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_i E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\right]\right) \qquad (4)$$

Where N is the total number of observations and the expectation operator indicates expectation conditional on observed information and distribution assumptions. The E-step is straightforward from (4) and, at iteration $m+1$, requires the calculation of:

$$Q_i\left(\boldsymbol{\theta}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)=E\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^{'}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]=E\left[\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1\\ y_{2i}^{*}-\gamma_2 y_{1i}-X_{2i}\beta_2\\ \vdots\\ y_{ki}^{*}-\gamma_k y_{1i}-X_{ki}\beta_k\end{pmatrix}\begin{pmatrix} y_{1i}^{*}-X_{1i}\beta_1\\ y_{2i}^{*}-\gamma_2 y_{1i}-X_{2i}\beta_2\\ \vdots\\ y_{ki}^{*}-\gamma_k y_{1i}-X_{ki}\beta_k\end{pmatrix}^{'}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]$$

$$=\sigma_i^{2(m)}+\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)}-X_{1i}\beta_1\\ \mu_{y_{2i}^{*}}^{(m)}-\gamma_2 y_{1i}-X_{2i}\beta_2\\ \vdots\\ \mu_{y_{ki}^{*}}^{(m)}-\gamma_k y_{1i}-X_{ki}\beta_k\end{pmatrix}\begin{pmatrix} \mu_{y_{1i}^{*}}^{(m)}-X_{1i}\beta_1\\ \mu_{y_{2i}^{*}}^{(m)}-\gamma_2 y_{1i}-X_{2i}\beta_2\\ \vdots\\ \mu_{y_{ki}^{*}}^{(m)}-\gamma_k y_{1i}-X_{ki}\beta_k\end{pmatrix}^{'} \qquad (5)$$

Where $\sigma_i^{2(m)}=\mathrm{Cov}\left(y_{1i}^{*},\cdots,y_{ki}^{*}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)=\begin{pmatrix}\sigma_{y_{1i}^{*}}^{2(m)} & \cdots & \sigma_{y_{1i}^{*}y_{ki}^{*}}^{(m)}\\ \vdots & \ddots & \vdots\\ \sigma_{y_{1i}^{*}y_{ki}^{*}}^{(m)} & \cdots & \sigma_{y_{ki}^{*}}^{2(m)}\end{pmatrix}$ (6)

and $\begin{pmatrix}\mu_{y_{1i}^{*}}^{(m)}\\ \vdots\\ \mu_{y_{ki}^{*}}^{(m)}\end{pmatrix}=\begin{pmatrix}E\left[y_{1i}^{*}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]\\ \vdots\\ E\left[y_{ki}^{*}\mid\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right]\end{pmatrix}$ (7)

The covariance matrix $\sigma_i^{2(m)}$ in (6) and the vector of means in (7) can be estimated by Gibbs sampling (Casella and George, 1992) from the joint distribution of $\left(y_{1i}^{*},\cdots,y_{ki}^{*}\right)$ conditional on parameters $\left(\boldsymbol{\theta}^{(m)},\Sigma^{(m)}\right)$ and the observed information $\boldsymbol{y}$. Additionally, given the distribution of the disturbances in (3), the distribution of $\left(y_{1i}^{*},\cdots,y_{ki}^{*}\right)$ at iteration m is $N\left(\mu_i^{(m)},\Sigma^{(m)}\right)$, where:

$$\mu_i^{(m)} = \begin{pmatrix} X_{1i}\beta_1^{(m)} \\ \gamma_2^{(m)} y_{1i} + X_{2i}\beta_2^{(m)} \\ \vdots \\ \gamma_k^{(m)} y_{1i} + X_{ki}\beta_k^{(m)} \end{pmatrix} \quad \text{and} \quad \Sigma^{(m)} = \begin{pmatrix} 1 & \sigma_{\varepsilon_1\varepsilon_2}^{(m)} & \cdots & \sigma_{\varepsilon_1\varepsilon_k}^{(m)} \\ \sigma_{\varepsilon_1\varepsilon_2}^{(m)} & \sigma_{\varepsilon_2}^{(m)} & \cdots & \sigma_{\varepsilon_2\varepsilon_k}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\varepsilon_1\varepsilon_k}^{(m)} & \sigma_{\varepsilon_2\varepsilon_k}^{(m)} & \cdots & \sigma_{\varepsilon_k}^{(m)} \end{pmatrix} \tag{8}$$

1.2. The Gibbs sampler

The moments in (6) and (7) 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 hefty 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. Before proceeding, let consider the following notation (Natarajan *et al.* 2000)

$$\boldsymbol{y}_{i|-j}^* = \begin{pmatrix} y_{1i}^* \\ \vdots \\ y_{j-1i}^* \\ y_{j+1i}^* \\ \vdots \\ y_{ki}^* \end{pmatrix} \quad \boldsymbol{X}_{i|-j} = \begin{pmatrix} X_{1i} \\ \vdots \\ X_{j-1i} \\ X_{j+1i} \\ \vdots \\ X_{ki} \end{pmatrix} \quad \boldsymbol{\gamma}_{-j}^{(m)} = \begin{pmatrix} \gamma_1^{(m)} \\ \vdots \\ \gamma_{j-1}^{(m)} \\ \gamma_{j+1}^{(m)} \\ \vdots \\ \gamma_k^{(m)} \end{pmatrix} \quad \boldsymbol{\beta}_{-j}^{(m)} = \begin{pmatrix} \beta_1^{(m)} \\ \vdots \\ \beta_{j-1}^{(m)} \\ \beta_{j+1}^{(m)} \\ \vdots \\ \beta_k^{(m)} \end{pmatrix}$$

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 $\boldsymbol{y}_{i|-j}^*$. Under the normality assumption, these conditional distributions are univariate normal. Thus, means $\mu_{ji|i(-j)}$ and variances $\sigma_{j|-j}^2$ at the $m+1$ iteration can be estimated by

$$\begin{aligned} \mu_{ji|i(-j)}^{(m)} &= E\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}\right) \\ &= X_{ji}\beta_j^{(m)} + \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Sigma^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^* \middle| \Sigma^{(m)}\right)\right]^{-1}\left(\boldsymbol{y}_{i|-j}^* - \boldsymbol{\gamma}_{-j}^{(m)} - \boldsymbol{X}_{i|-j}\boldsymbol{\beta}_{-j}^{(m)}\right) \end{aligned} \tag{9}$$

$$\begin{aligned} \sigma_{j|-j}^{2(m)} &= \operatorname{var}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \boldsymbol{\theta}^{(m)}, \Sigma^{(m)}\right) \\ &= \operatorname{var}\left(y_{ji}^* \middle| \Sigma^{(m)}\right) - \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Sigma^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^* \middle| \Sigma^{(m)}\right)\right]^{-1} \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Sigma^{(m)}\right)' \end{aligned} \tag{10}$$

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 (6) and (7). 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) 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_{ji}^{*} ($j=2,...k$) are observed in the interval $\left(a_j, b_j\right)$. Consequently, it is only necessary to simulate them when $y_{ji}=a_j$ or $y_{ki}=b_j$. 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 a_j when $y_{ji}=a_j$ and truncated below at b_j when $y_{ki}=b_j$. When $a_j < y_{ji} < b_j$, we set $y_{ji}^{*}=y_{ji}$.

We use the inverse distribution method to sample from a truncated normal distribution. According to Devroye (1986, p39), a random draw from a normal distribution $N\left(\mu,\sigma^2\right)$ limited to the interval $[l,u]$ is given by

$$y=\mu+\sigma\Phi^{-1}\left(P_l+U\left(P_u-P_l\right)\right) \qquad (11)$$

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. A complete set of starting vectors $\boldsymbol{y}_i^{*}$ is necessary to start 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. Then we eliminate a number k_{burn} of simulations from the beginning of the sequence. The remaining observations in the sequence are used to obtain sample estimate for $\sigma_{ji}^{2(m)}$ and $\mu_{ji}^{(m)}$ in (6) and (7). Notice that when y_{ji}^{*} is fully observed (i.e. when $y_{ji}^{*}=y_{ji}$) then $\sigma_{y_{ri}^{*}y_{si}^{*}}^{(m)}=0$ and $\mu_{y_{ji}^{*}}^{(m)}=y_{ij} \quad \forall\, m,r,s$.

1.3. Maximization Step

After obtaining $\sigma_{ji}^{2(m)}$ and $\mu_{ji}^{(m)}$, we move to the Maximization step. From (4) and (5) we maximize:

$$E\left[\ell^c\left(\boldsymbol{\theta},\Sigma\,|\,\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right]=-\frac{kN}{2}\ln(2\pi)-\frac{N}{2}\ln|\Sigma|-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_i Q_i\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right) \quad (12)$$

Notice that, except for the covariance matrices $\sigma_i^{2(m)}$ present in the $Q_i\left(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)$ terms, the expression in (12) is the log-likelihood function of a system of linear equations, where the unobserved information has been replaced by its expected values. As in Meg and Rubin (1993), we use two conditional maximization steps in order to maximize the expression in (12) with respect to $\boldsymbol{\theta}$ and the elements in Σ. The first maximization step maximizes (12) 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 (5) that the maximizer in the first conditional maximization is the generalized least square estimator

$$\boldsymbol{\theta}^{(m+1)}=\left[\tilde{X}_d'\left(\Sigma^{(m)}\otimes I_N\right)^{-1}\tilde{X}_d\right]^{-1}\tilde{X}_d'\left(\Sigma^{(m)}\otimes I_N\right)^{-1}\mu_{y^*}^{(m)} \quad (13)$$

Where:

$$\tilde{X}_d=\begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & \tilde{X}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{X}_k \end{bmatrix}, \quad \tilde{X}_j=\begin{bmatrix} y_{1i} \vdots X_j \end{bmatrix} \ (j=2,\ldots k),$$

I_N is a *N*-dimensional identity matrix and $\mu_{y^*}^{(m)}$ is a column vector of dimension Nk constructed by stacking vertically the elements $\mu_{y_{ji}^*}^{(m)}$ from (7). After plugging (13) in (12), $\Sigma^{(m+1)}$ is obtained by maximizing

$$E\left[\ell^c\left(\Sigma\,|\,\boldsymbol{\theta}^{(m+1)},\Sigma^{(m)},\boldsymbol{y}\right)\right]=-\frac{kN}{2}\ln(2\pi)-\frac{N}{2}\ln|\Sigma|-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_i Q_i\left(\boldsymbol{\theta}^{(m+1)}\,|\,\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)\right) \quad (14)$$

with respect to the $k(k-1)/2+k-1$ different elements in Σ and subject to $\sigma_{\varepsilon_2}^2=1$. The maximization of (14) can be accomplished by standard commercial routines. The objective function is simple enough to obtain a simple analytical expression for the first order conditions, which are

$$\Sigma-\frac{1}{N}\sum_i Q_i\left(\boldsymbol{\theta}^{(m+1)}\,|\,\boldsymbol{\theta}^{(m)},\Sigma^{(m)},\boldsymbol{y}\right)+\frac{\lambda}{N}\Sigma A\Sigma=0 \quad (15)$$

where λ is a scalar Lagrange multiplier and A is a matrix whose only non-zero element is a one in the position $(1,1)$ An estimate of Σ can be obtained by solving numerically either (14) or (15). In our simulations we found that solving (15) is faster and shows fewer convergence problems.

1.4. Monitoring convergence and stopping rules

It is inefficient to begin with large Gibbs samples since MCEM estimates are likely to be distant from the true maximizer during the first iterations (Wei and Tanner, 1990). It is more reasonable to begin with small samples and to make $K^{(m)}$ an increasing function of m in order to reduce the Monte Carlo error as the algorithm approaches the maximizer. Alternatively, convergence monitoring can be accomplished by plotting the expected log-likelihood versus iteration number and the algorithm is stopped manually when the process is observed to stabilize.

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 (Chan and Ledolter, 1995; Eickhoff *et al*., 2004; Booth and Hobert, 1999). In this study, we use 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:

$$\sum_{j=M-J}^{M}\left|\frac{E\left[\ell^c\left(\vartheta^{(j)}\right)\right]-E\left[\ell^c\left(\vartheta^{(j-1)}\right)\right]}{E\left[\ell^c\left(\vartheta^{(j-1)}\right)\right]}\right|<10^{-3}\ ,\ \max_r\left[\sum_{j=M-J}^{M}\left|\frac{\vartheta_r^{(j)}-\vartheta_r^{(j-1)}}{\vartheta_r^{(j-1)}}\right|\right]<10^{-3} \qquad (16)$$

Where $\vartheta_r^{(j)}$ is the estimate of the r component of the parameter vector at iteration *j*, *M* is the current number of iterations, and *J* is a researcher choice. In this work, *J* was set equal to $0.2\times M$. The algorithm was stopped only when both criteria were satisfied simultaneously for at least five 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 (16) are simple to implement and more stringent than those presented in the articles cited above and might increase unnecessarily the number of iterations required for convergence. However, given the speed of today's computer, the computational cost is not very high.

1.5. Estimation of the Information matrix

The asymptotic standard errors of the estimates are not among the outputs of the EM algorithm and additional code needs to be appended to the algorithm in order to estimate them. Let the complete information log-likelihood function be $\ell^c(\boldsymbol{\theta};\boldsymbol{z})$, where $\boldsymbol{\theta}$ is the full set of parameters to estimate. Then, according to the "missing information principle" (Orchard and Woodbury, 1972), the information matrix of the observed information is:

$$I(\boldsymbol{\theta};\boldsymbol{y}) = I^c(\boldsymbol{\theta};\boldsymbol{z}) - I^m(\boldsymbol{\theta};\boldsymbol{z}\,|\,\boldsymbol{y}) \qquad (17)$$

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

$$I^m(\boldsymbol{\theta};\boldsymbol{z}\,|\,\boldsymbol{y}) = -E\left[\frac{\partial^2 \ell^m(\boldsymbol{\theta};\boldsymbol{z}\,|\,\boldsymbol{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] = \mathrm{Var}\left[S^c(\boldsymbol{\theta};\boldsymbol{z})\right] = E\left[S^c(\boldsymbol{\theta};\boldsymbol{z})S^c(\boldsymbol{\theta};\boldsymbol{z})'\right] - E\left[S^c(\boldsymbol{\theta};\boldsymbol{z})\right]E\left[S^c(\boldsymbol{\theta};\boldsymbol{z})'\right] \qquad (18)$$

Where $S^c(\boldsymbol{\theta};\boldsymbol{z}) = \dfrac{\partial \ell^c(\boldsymbol{\theta};\boldsymbol{z})}{\partial\boldsymbol{\theta}}$ is the complete information score vector. All the expectations are taken with respect to the distribution $f\left(\boldsymbol{z}\,|\,\boldsymbol{y},\boldsymbol{\theta}^{EM}\right)$, where $\boldsymbol{\theta}^{EM}$ is the final MCEM estimator. The evaluation of all the expectations involved commonly prevents the estimation of the observed information matrix in (18) by direct calculation. Monte Carlo estimates of the expected complete information Hessian and score can be used to circumvent the problem. 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:

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

$$I^m\left(\boldsymbol{\theta}^{EM};\boldsymbol{x}\,|\,\boldsymbol{y}\right) = \sum_{i=1}^{N}\left\{E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{z}_i\right)S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{z}_i\right)'\right] - E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{z}_i\right)\right]E\left[S_i^c\left(\boldsymbol{\theta}^{EM};\boldsymbol{z}_i\right)'\right]\right\}$$

$$\cong \sum_{i=1}^{N}\left\{\frac{1}{R}\sum_{r=1}^{R} S_i^{c(r)}\left(\boldsymbol{\theta}^{EM},\boldsymbol{y}_i^{*(r)}\,|\,\boldsymbol{y}_i\right)S_i^{c(r)}\left(\boldsymbol{\theta}^{EM},\boldsymbol{y}_i^{*(r)}\,|\,\boldsymbol{y}_i\right)' - \frac{1}{R}\sum_{r=1}^{R} S_i^{c(r)}\left(\boldsymbol{\theta}^{EM};\boldsymbol{y}_i^{*(r)}\,|\,\boldsymbol{y}_i\right)\frac{1}{R}\sum_{r=1}^{R} S_i^{c(r)}\left(\boldsymbol{\theta}^{EM};\boldsymbol{y}_i^{*(r)}\,|\,\boldsymbol{y}_i\right)'\right\}$$

Expressions for the contributions from each observation to the Hessian and score are standard results from the theory of the multivariate normal

distribution. A sample with $R = 3300$ and $r_{burn} = 300$ was used in this study. Results of the Monte Carlo EM estimation of equation system (2) are presented in Table 1.

2. The numerical integration approach

In this section we solve by numerical integration the same model and data used to illustrate the MCEM algorithm. The performances of the two approaches are then compared.

The observed-information log-likelihood function for the equation system (2) with $a_j = 0$, $b_j = \infty$, and $k = 3$ is

$$L = \sum_{\substack{y_{1i}=0\\y_{2i}=0\\y_{3i}=0}} \ln \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{-\infty}^{0} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{2i}^* dy_{3i}^* + \sum_{\substack{y_{1i}=1\\y_{2i}=0\\y_{3i}=0}} \ln \int_{-\infty}^{0}\int_{-\infty}^{0}\int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{2i}^* dy_{3i}^* + \sum_{\substack{y_{1i}=0\\y_{2i}=0\\y_{3i}>0}} \ln \int_{-\infty}^{0}\int_{-\infty}^{0} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{2i}^* +$$

$$\sum_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}=0}} \int_{-\infty}^{0}\int_{-\infty}^{0} \ln f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{3i}^* + \sum_{\substack{y_{1i}=1\\y_{2i}=0\\y_{3i}>0}} \ln \int_{-\infty}^{0}\int_{0}^{\infty} \ln f\left(y_{1i}^*, y_{2i}^*, y_{3i}\right) dy_{1i}^* dy_{2i}^* + \sum_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}=0}} \ln \int_{-\infty}^{0}\int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* dy_{3i}^* +$$

$$\sum_{\substack{y_{1i}=0\\y_{2i}>0\\y_{3i}>0}} \ln \int_{-\infty}^{0} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* + \sum_{\substack{y_{1i}=1\\y_{2i}>0\\y_{3i}>0}} \ln \int_{0}^{\infty} f\left(y_{1i}^*, y_{2i}^*, y_{3i}^*\right) dy_{1i}^* \qquad (19)$$

Where $f(\cdot)$ is a trivariate normal probability density function (pdf.). It turns out that all the integrals in the likelihood function can be written as products of the trivariate, bivariate, and univariate standard normal cumulative density function (cdf) and pdf. Instead of using three-dimensional quadratures, we programmed the cdf functions according to the methodology pioneered by Steck (1958) and Sowden and Ashford (1969) to accelerate the estimation. This methodology allows reducing the high-dimensional normal cdf to functions involving only 1-dimensional integrals and the univariate normal cdf. As mentioned before, an alternative to handle the integrals is to use probability simulators. The information matrix was calculated from a finite-difference estimation of the Hessian of the objective function.

3. Comparison between numerical integration and MCEM estimates

3.1. Simulation study

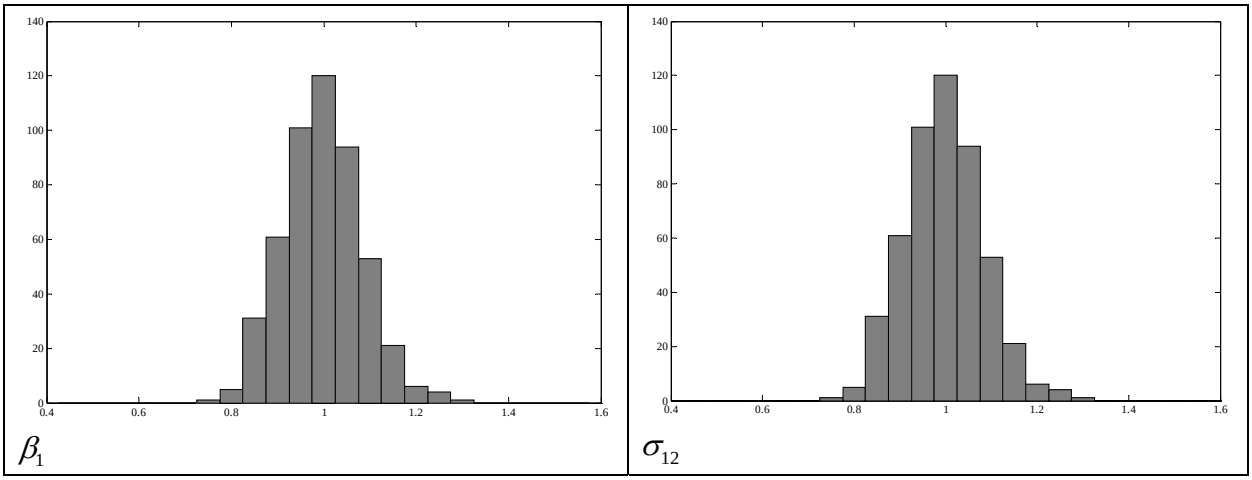
In order to compare both estimation methods, we run a simulation study. We consider dos regressors in each of the three equations: a constant and a continuous variable, which we simulated from a uniform distribution. We simulated the continuous regressor in the range $[-2, 2]$ for the selection

equation, in $[1,2]$ for the first response equation, and in $[-1,1]$ for the second response equation; all of regressors were kept fixed during the simulation study. We simulated the latent dependent variables from a trivariate normal distribution $N(\mu,\Sigma)$, where:

$$\mu=\begin{pmatrix} X_1\beta_1 \\ X_2\beta_2 \\ X_3\beta_3 \end{pmatrix} \quad \beta_1=\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \beta_2=\begin{pmatrix} 1 \\ -0.5 \end{pmatrix} \quad \beta_3=\begin{pmatrix} -1 \\ 0.5 \end{pmatrix} \quad \Sigma=\begin{pmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{pmatrix}$$

The dependent observed variables were then obtained by imposing that the dependent variable is dichotomous for the selection equation and censored for the response equations. We replicated the estimation 500 times, which allows considering samples with a broad variation in the number of censored observations. We considered samples of 500 simulated observations and OLS estimates as starting values in each one of the 500 estimations. Estimation times ranged between 5 and 15 minutes depending mainly on the level of censoring in the simulated samples.[1] To evaluate the sensibility to the selection of starting values, we run a smaller simulation study using vectors of zeros and vectors of numbers randomly generated from the uniform distribution for the slopes and disturbance covariance matrices calculated from the corresponding residuals. We observed no effect whatsoever on the performance of the MCEM algorithm. Figure 1 shows histograms for the estimates obtained for selected parameters.

FIGURE 1. SIMULATION RESULTS FOR SELECTED PARAMETERS



We were unable to perform a similar analysis for the combination of numerical integration and quasi-Newton optimization since the estimation

[1] On a Pentium IV, 1 GB RAM, Windows XP, Matlab 7.01

time (45 to 140 minutes) made it computationally prohibitive. Instead, we replicated the estimation only twenty times starting from the OLS estimates. Only a fourth of the estimations converged correctly to the maximum, while all the rest stopped in a neighborhood of the optimum because the algorithm was unable to make further progress. All the unsuccessful estimations generated finite-difference Hessians that were not negative definite and, consequently, they failed in providing a correct covariance matrix for the estimates. Finally, to evaluate the sensibility to starting values, we try several random combinations of starting values including vector of zeros for the slopes of some or all the equations. For almost all the combinations we tried, the algorithms based on quasi-Newton methods failed to converge.

3.2. Study on real data

We applied the method on data from a survey administered to Maryland farmers in order to evaluate a multi-objective conservation cost-sharing program. The program is voluntary, and the parameters of primary interest are γ_2 and γ_3. Thus, the first equation in our three-equation system models the participation decision (only a dichotomous response —yes or no— variable is observed), while the second and third equations model two conservation responses that program administrators want to influence. Variable y_{2i} and y_{3i} are the proportion of the farm under permanent vegetative cover and the proportion of the area cropped on which cover crops are planted, respectively. Both responses are proportions and thus they are censored from below at cero for individuals who find the use of these practices not profitable (i.e. we observe $y_{2i}=0$ and/or $y_{3i}=0$ for them). Additionally, it might be possible to observe censoring at one for some individuals finding the practices highly profitable. Our sample is abundant in observations censored at zero, but not a single case censored at one was detected. The matrices of explanatory variables X_1, X_2, and X_3 include information about farmer characteristics (such as age and formal education), farm topography, farm size, cropping patterns, distance to water bodies, previous history of participation in conservation programs, and location dummies.

The iteration paths for the expected log-likelihood function and selected parameters are presented in Figure 2. OLS estimates were used as starting values. The routine converged after 535 iterations. The Gibbs sampler was started with 300 simulations and increased in an amount of 15 simulations per iteration, i.e. $K^{(m)}=300+15(m-1)$. The number of dismissed cases, k_{burn}, was kept constant at 150. Estimates are presented in Table 1.

TABLE 1. FIML ESTIMATES BY THE MONTE CARLO EM ALGORITHM

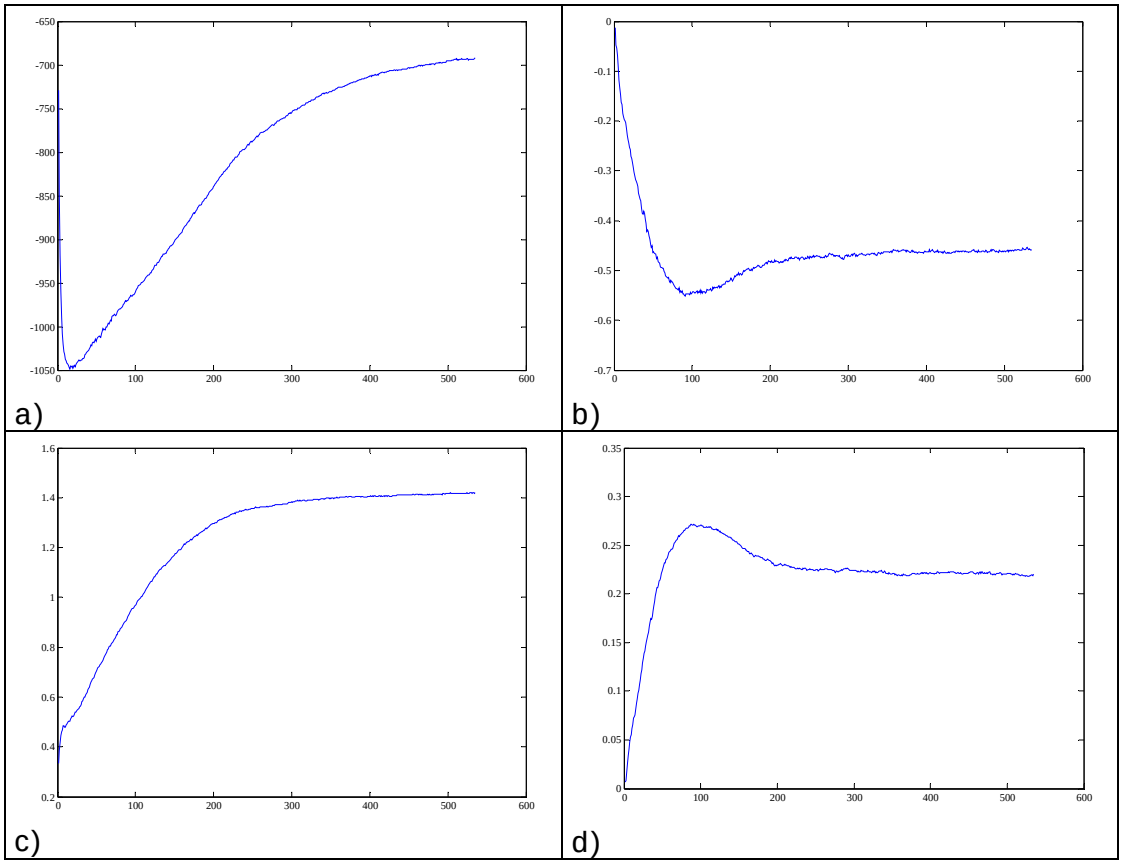
Equation 1			Equation 2			Equation 3		
	Estimate	Std. error		Estimate	Std. error		Estimate	Std. error
			γ_2	-0.1442	0.1730	γ_3	0.1833	0.1822
β_1	-0.9004	0.5406	β_1	-0.4581	0.2029	β_1	1.4197	0.1535
β_2	-2.0180	0.9021	β_2	-0.1678	0.2711	β_2	-0.4695	0.2872
β_3	0.3352	0.2081	β_3	0.2190	0.0673	β_3	-0.1113	0.0784
β_4	0.0193	0.3706	β_4	0.0072	0.1135	β_4	-0.3973	0.1493
β_5	0.1183	0.2550	β_5	0.1334	0.0867	β_5	-0.1474	0.1066
β_6	0.2423	0.1449	β_6	0.1782	0.0432	β_6	0.0460	0.0493
β_7	0.4204	0.2106	β_7	-0.0247	0.1087	β_7	0.0735	0.1113
β_8	0.1903	0.1949	β_8	-0.0441	0.1088	β_8	0.0298	0.0743
β_9	0.9313	0.2258	β_9	0.0039	0.0084	β_9	-0.0059	0.0055
			σ_{12}	0.2191	0.0754	σ_{13}	-0.5158	0.0544
σ_{23}	-0.0358	0.0236	σ_{22}	0.2439	0.0318	σ_{33}	0.3403	0.0441

The expected treatment effect for practice $j = 2,3$ and farm i, conditional on program participation and after controlling for censoring, is:

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

Where X_{ji} and β_j are the set of regressors and the slopes in the equation of practice j respectively; $\Phi_2(\cdot)$ and $\Phi(\cdot)$ are correspondingly the bivariate and univariate standard normal cdf, and ρ_{1j} is the correlation between the error terms in the selection equation and the equation for practice j. The average expected treatment effect on the whole sample is –0.3436 for permanent vegetative cover and 0.1143 for the use of cover crops. The corresponding standard errors are 0.0333 and 0.0002, which were estimated by the delta method. These results agree with theoretical results (*e.g.* Khanna *et al.*, 2002), which indicates that cost sharing programs reduces set-aside land and increases the use of conservation practices linked to production activities.

FIGURE 2. ITERATION PATHS FOR
A) EXPECTED LOG-LIKELIHOOD, B) γ_2, C) γ_3 AND D) σ_{12}



Conclusions

This article presented a MCEM algorithm to estimate treatment effect models involving multiple censored responses. The algorithm has a number of advantages over traditional methods. First, it does not require integrating the unobserved information out from the likelihood function, which reduces the estimation time dramatically as no numerical integration is needed and 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 methods are 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, is easier to keep the whole set of parameters in the parameter space during the procedure and almost no time is consumed in the Maximization step. Additionally, the MCEM reduces substantially the problems of "fragile" identification and selection of starting values, which are serious limitations of traditional approaches. Finally, the accuracy of the estimates of the standard errors can be improved inexpensively by increasing the number of simulations of the Information matrix.

References

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.

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

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

Cooper, J. 2003. "A Joint framework for Analysis of Agri-Environmental Payment Programs". Am. J. of Agr. Econ 85(4): 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. pp. 843.

Dorfman, J. 1996. "Modeling Multiple Adoption Decisions in a Joint Framework". Am. J. of Agr. Econ 78(August 1996): 547-557.

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

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.

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.

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

Khanna M., M. Isik and D. Zilberman. "Cost-effectiveness of alternative green payment policies for conservation technology adoption with heterogeneous land quality". *Agricultural Economics* 27 (2002) 157–174.

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

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.

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.

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.

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

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.

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.

Novedades

DIVISIÓN DE ADMINISTRACIÓN PÚBLICA

Cejudo, Guillermo, *Critical Junctures or Slow-Moving Processes? The Effects of Political and Economic Transformations…*, DTAP-186

Sour, Laura, *Un repaso de conceptos sobre capacidad y esfuerzo fiscal, y su aplicación para los gobiernos locales mexicanos*, DTAP-187

Santibañez, Lucrecia, School-Based Management Effects on Educational Outcomes: A Literature Review and Assessment of the Evidence Base, DTAP-188

Cejudo, Guillermo y Sour Laura, *¿Cuánto cuesta vigilar al gobierno federal?*, DTAP-189

Cejudo, Guillermo, *New Wine in Old Bottles: How New Democracies Deal with Inherited Bureaucratic Apparatuses…*, DTAP-190

Arellano, David, *Fallas de transparencia: hacia una incorporación efectiva de políticas de transparencia en las organizaciones públicas*, DTAP-191

Sour, Laura y Munayer Laila, *Apertura política y el poder de la Cámara de Diputados durante la aprobación presupuestaria en México*, DTAP-192

Casar, Ma. Amparo, *La cultura política de los políticos en el México democrático*, DTAP-193

Arellano, David y Lepore Walter, *Economic Growth and Institutions: The Influence of External Actors*, DTAP-194

Casar, Ma. Amparo, *Los gobiernos sin mayoría en México: 1997-2006*, DTAP-195

DIVISIÓN DE ECONOMÍA

Castañeda, Alejandro y Villagómez Alejandro, *Ingresos fiscales petroleros y política fiscal óptima*, DTE-382

Dam, Kaniska, *A Two-Sided Matching Model of Monitored Finance*, DTE-383

Dam, Kaniska, Gautier Axel y Mitra Manipushpak, *Efficient Access Pricing and Endogenous Market Structure*, DTE-384

Dam, Kaniska y Sánchez Pagés Santiago, *Deposit Insurance, Bank Competition and Risk Taking*, DTE-385

Carreón, Víctor, Di Giannatale Sonia y López Carlos, *Mercados formal e informal de crédito en Mexico: Un estudio de caso*, DTE-386

Villagómez, Alejandro y Roth Bernardo, Fiscal Policy and National Saving in Mexico, 1980-2006, DTE-387

Scott, John, *Agricultural Policy and Rural Poverty in Mexico*, DTE-388

Hogan, William, Rosellón Juan y Vogeslang Ingo, *Toward a Combined Merchant-Regulatory Mechanism for Electricity Transmission Expansion*, DTE-389

Roa, Ma. José y Cendejas José Luis, *Crecimiento económico, estructura de edades y dividendo demográfico*, DTE-390

Kristiansen, Tarjei y Rosellón Juan, *Merchant Electricity Transmission Expansion: A European Case Study*, DTE-391

DIVISIÓN DE ESTUDIOS INTERNACIONALES

Schiavon, Jorge y Velázquez Rafael, *El 11 de septiembre y la relación México-Estados Unidos: ¿Hacia la securitización de la agenda?*, DTEI-150

Velázquez, Rafael, *La paradiplomacia mexicana: Las relaciones exteriores de las entidades federativas*, DTEI-151

Meseguer, Covadonga, *Do Crises Cause Reform? A New Approach to the Conventional Wisdom*, DTEI-152

González, Guadalupe y Minushkin Susan, Líderes, opinión pública y política exterior en México, Estados Unidos y Asia: un estudio comparativo, DTEI-153

González, Guadalupe y Minushkin Susan, *Leaders, public opinion and foreign policy in Mexico, the United States, and Asia: a comparative study*, DTEI-154

González, Guadalupe y Minushkin Susan, *Opinión pública y política exterior en México*, DTEI-155

González, Guadalupe y Minushkin Susan, *Public opinion and foreign policy in Mexico*, DTEI-156

Ortiz Mena, Antonio, *El Tratado de Libre Comercio de América del Norte y la política exterior de México: lo esperado y lo acontecido*, DTEI-157

Ortiz Mena, Antonio y Fagan Drew, *Relating to the Powerful One: Canada and Mexico's Trade and Investment Relations with the United States*, DTEI-158

Schiavon, Jorge, *Política exterior y opinión pública: México ante el mundo*, DTEI-159

DIVISIÓN DE ESTUDIOS JURÍDICOS

Fondevila Gustavo, *Estudio de percepción de usuarios del servicio de administración de justicia familiar en el Distrito Federal,* DTEJ-14

Pazos, Ma. Inés, *Consecuencia lógica derrotable: análisis de un concepto de consecuencia falible,* DTEJ-15

Posadas, Alejandro y Hugo E. Flores, *Análisis del derecho de contar con un juicio justo en México*, DTEJ-16

Posadas, Alejandro, *La Responsabilidad Civil del Estado /Análisis de un caso hipotético*, DTEJ-17

López, Sergio y Posadas Alejandro, *Las pruebas de daño e interés público en materia de acceso a la información. Una perspectiva comparada,* DTEJ-18

Magaloni, Ana Laura, *¿Cómo estudiar el derecho desde una perspectiva dinámica?*, DTEJ-19

Fondevila, Gustavo, *Cumplimiento de normativa y satisfacción laboral: un estudio de impacto en México*, DTEJ-20

Posadas, Alejandro, *La educación jurídica en el CIDE (México). El adecuado balance entre la innovación y la tradición*, DTEJ-21

Ingram, Matthew C., *Judicial Politics in the Mexican States: Theoretical and Methodological Foundations*, DTEJ-22

Fondevila, Gustavo e Ingram Matthew, *Detención y uso de la fuerza*, DTEJ-23

DIVISIÓN DE ESTUDIOS POLÍTICOS

Lehoucq, Fabrice E., Structural Reform, *Democratic Governance and Institutional Design in Latin America*, DTEP-188

Schedler, Andreas, *Patterns of Repression and Manipulation. Towards a Topography of Authoritarian Elections*, *1980-2002,* DTEP-189

Benton, Allyson, *What Makes Strong Federalism Seem Weak? Fiscal Resources and Presidencial-Provincial Relations in Argentina*, DTEP-190

Crespo, José Antonio, *Cultura política y consolidación democrática (1997-2006)*, DTEP-191

Lehoucq, Fabrice, *Policymaking, Parties and Institutions in Democratic Costa Rica*, DTEP-192

Benton, Allyson, *Do Investors Assess the Credibility of Campaign Commitments? The Case of Mexico's 2006 Presidential Race*, DTEP-193

Nacif, Benito, *Para entender las instituciones políticas del México democrático*, DTEP-194

Lehoucq, Fabrice, *Why is Structural Reform Stangnating in Mexico? Policy Reform Episodes from Salinas to Fox*, DTEP-195

Benton, Allyson, *Latin America's (Legal) Subnational Authoritarian Enclaves: The Case of Mexico*, DTEP-196

Hacker, Casiano y Jeffrey Thomas, *An Antitrust Theory of Group Recognition*, DTEP-197

DIVISIÓN DE HISTORIA

Pipitone, Ugo, *Aperturas chinas (1889, 1919, 1978),* DTH-34

Meyer, Jean, *El conflicto religioso en Oaxaca*, DTH-35

García Ayluardo Clara, *El privilegio de pertenecer. Las comunidades de fieles y la crisis de la monarquía católica*, DTH-36

Meyer, Jean, *El cirujano de hierro (2000-2005),* DTH-37

Sauter, Michael, *Clock Watchers and Stargazers: On Time Discipline in Early-Modern Berlin,* DTH-38

Sauter, Michael, *The Enlightenment on Trial…,* DTH-39

Pipitone, Ugo, *Oaxaca prehispánica*, DTH-40

Medina Peña, Luis, *Los años de Salinas: crisis electoral y reformas*, DTH-41

Sauter, Michael, *Germans in Space: Astronomy and Anthropologie in the Eighteenth Century*, DTH-42

Meyer, Jean, *La Iglesia católica de los Estados Unidos frente al conflicto religioso en México, 1914-1920*, DTH-43

Ventas

El **Centro de Investigación y Docencia Económicas / CIDE,** es una institución de educación superior especializada particularmente en las disciplinas de Economía, Administración Pública, Estudios Internacionales, Estudios Políticos, Historia y Estudios Jurídicos. **El CIDE publica**, como producto del ejercicio intelectual de sus investigadores, **libros, documentos de trabajo, y cuatro revistas especializadas:** *Gestión y Política Pública***,** *Política y Gobierno***,** *Economía Mexicana Nueva Época* e *Istor***.**

Para adquirir alguna de estas publicaciones, le ofrecemos las siguientes opciones:

VENTAS DIRECTAS:

Tel. Directo: 5081-4003
Tel: 5727-9800 Ext. 6094 y 6091
Fax: 5727 9800 Ext. 6314

Av. Constituyentes 1046, 1er piso, Col. Lomas Altas, Del. Álvaro Obregón, 11950, México, D.F.

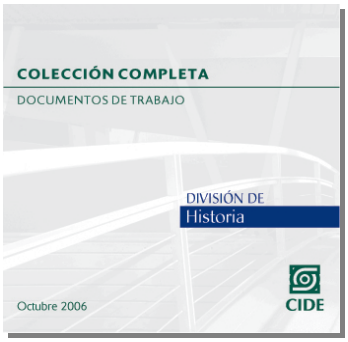
VENTAS EN LÍNEA:

Librería virtual: **www.e-cide.com**

Dudas y comentarios:
publicaciones@cide.edu

¡Nuevo!

Adquiera el **CD de las colecciones completas** de los documentos de trabajo de la **División de Historia** y de la **División de Estudios Jurídicos.**



¡Próximamente! los CD de las colecciones completas de las Divisiones de Economía, Administración Pública, Estudios Internacionales y Estudios Políticos.