NÚMERO 425

RICARDO SMITH RAMÍREZ

A Monte Carlo EM algorithm for FIML estimation of multivariate endogenous switching models with censored and discrete responses

JUNIO 2008

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

A Monte Carlo EM algorithm for FIML estimation of multivariate endogenous switching models with censored and discrete responses

This article presents a Monte Carlo EM algorithm to estimate multivariate endogenous switching regression models with censored and/or discrete responses and heteroscedastic errors. Advantages of the algorithm include: (1) *it does not require numerical integration;* (2) *it reduces the estimation of the vector of slopes to the calculation of a GLS estimator and* (3) *numerical techniques are required only to estimate the parameters in the disturbance covariance matrix. Extensions to panel data are discussed. The algorithm is illustrated on both simulated data and on real data from an agricultural conservation program.*

Keywords: Monte Carlo EM, Gibbs sampling, switching regression.

Resumen

Estimación de modelos cambiantes multivariados con respuestas dicotómicas y censuradas mediante un algoritmo Monte Carlo EM

Este artículo presenta un algoritmo Monte Carlo EM para estimar modelos de regresión cambiante con respuestas múltiples. El algoritmo permite el análisis de modelos que contengan respuestas censuradas y/o discretas, y permite controlar por heteroscedasticidad. La estimación se realiza por máxima verosimilitud con información completa. Algunas ventajas del algoritmo son: (1) *no requiere integración numérica,* (2) *la estimación del vector de pendientes se reduce al cálculo de un estimador de mínimos cuadrado generalizados, y (3) la optimización numérica es requerida sólo para estimar los elementos de la matriz de covarianza de las perturbaciones. El uso del algoritmo es ilustrado sobre datos simulados y sobre datos reales tomados desde un programa de conservación de suelos.*

Palabras clave: Monte Carlo EM, muestra Gibbs, regresión cambiante.

Introduction

The endogenous switching regressions (SR) model is widely used to separate the components of a mixture of normal distributions when a discrete endogenous indicator determines the sample separation. The central idea behind the model can be tracked back to Roy's (1951) discussion about self-selection and earnings inequality, while the first econometric formulations belong to Gronau (1974), Lewis (1974) and Maddala and Nelson (1974). Regarding to estimation, Heckman (1976) and Lee *et al.* (1980) developed two-stage methods, whereas Griliches *et al.* (1978) and Kenny *et al.* (1979) were the first ones in using maximum likelihood (ML). Currently, SR models are utilized in almost every area of applied economics. An example is treatment evaluation, where SR models are used as an alternative to matching methods when the conditional independence assumption does not hold (*i.e.* when, after controlling for the observables, the participation decision is not independent from response outcomes). In these cases, participation is endogenous and a SR framework applies (Cameron and Trivedi, 2005).

Computational burden, however, limits full-information ML (FILM) estimation of multivariate SR models with latent structures, as it happens in evaluation of treatments aiming to influence more than one behavioral aspect of the treated subpopulation. In such cases, a joint estimation considering all the responses simultaneously is more efficient. Yet, the estimation of such models is challenging since, as in any problem involving latent variables, multidimensional integrals show up in the likelihood function. Although integrals can be solved either by Monte Carlo o quasi Monte Carlo integration, or by probability simulators (McFadden, 1989; Börsch-Supan and Hajivassiliou, 1993; Geweke *et al*., 1994), traditional optimization algorithms often have some serious drawbacks when working on models involving multiple latent variables. Apart from computation time, common problems are: high sensitivity to the selection of starting values, inability of line-search methods to progress to the optimum, and/or algorithms cannot keep the covariance estimates in the parameter space (e.g. the magnitude of the correlation coefficients become greater than one, or the disturbance covariance matrix is not longer positive definite).

The Monte Carlo EM (MCEM) algorithm that we propose here follows the previous work of Natarajan *et al.* (2000), who formulated an MCEM algorithm to estimate the multinomial probit model. Our formulation uses Gibbs sampling and takes advantage of the latent continuum to circumvent multivariate integration and reduce the FIML estimation of a multivariate SR model to the ML estimation of a system of seemingly unrelated regressions. Thus, in each of the algorithm iterations, the slope vector is estimated as a

generalized least square (GLS) estimator, while numerical optimization is used in the estimation of the covariance parameters only.

The remaining of this article is organized in the following way. The next section formulates the algorithm for censored responses and heteroscedastic disturbances. The second section illustrates the algorithm on both a simulated sample and real data from an agricultural conservation program. The third section gives indications about how handling different types of limited-dependent and discrete responses, and extends the algorithm to panel data. The fourth and last section gives final remarks.

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

Let consider the many-to-one mapping $z \in Z \to y = y(z) \in Y$. In words, z is only known to lie in $Z(y)$, the subset of Z determined by the equation $y = y(z)$, where $\boldsymbol{y}$ is the observed data and $\boldsymbol{z}$ is the complete but not fully observed information. Thus, the log-likelihood of the observed information is $l(\vartheta \mid \boldsymbol{y}) = \ln L(\vartheta \mid \boldsymbol{y}) = \ln \int_{Z(y)} L(\vartheta \mid \boldsymbol{z})\, d\boldsymbol{z}$, where ϑ is the parameter to be estimated. Instead of solving the integrals in $l(\vartheta \mid \boldsymbol{y})$ directly, the deterministic version of the EM algorithm (Dempster *et al*., 1977) focuses on the complete-information log-likelihood $\ell^c(\vartheta \mid \boldsymbol{z})$ and maximizes the expectation $\boldsymbol{E}\left[l^c(\vartheta \mid \boldsymbol{z}) \mid \boldsymbol{y} \right]$ by executing two steps iteratively: (1) the expectation step (E-step), which computes $Q\left(\vartheta \mid \vartheta^{(m)}, \boldsymbol{y}\right) = E\left[l^c(\vartheta \mid \boldsymbol{z}) \mid y \right]$ at iteration *m*+1, and (2) the maximization step (M-step), which maximizes $Q\left(\vartheta \mid \vartheta^{(m)}, \boldsymbol{y}\right)$ to find $\vartheta^{(m+1)}$. The procedure is repeated until convergence is attained.

The Monte Carlo version of the EM algorithm (Wei and Tanner, 1990) circumvents high-dimension integrals in the E-step by imputing the unobserved information conditional on what is observed and on distribution assumptions. In this approach, the term $Q\left(\vartheta \mid \vartheta^{(m)}, \boldsymbol{y}\right)$ is approximated by the mean $\frac{1}{K}\sum_{k=1}^{K} Q\left(\vartheta, \boldsymbol{z}^{(k)} \mid \boldsymbol{y}\right)$, where the $\boldsymbol{z}^{(k)}$ are random draws from $f\left(\boldsymbol{z} \mid \vartheta^{(m)}, \boldsymbol{y}\right)$.

1.1. The algorithm

Let consider the *k*-equation system:

$$y_{1i}^{*} = X_{1i}\beta_1 + \varepsilon_{1i} \qquad \text{selection (switching) equation}$$

$$\begin{array}{cc} y_{1i}=0 \; [&] \; y_{1i}=1 \\ y_{2i}^{*} = X_{2i}^{0}\beta_2^{0} + \varepsilon_{2i}^{0} & y_{2i}^{*} = X_{2i}^{1}\beta_2^{1} + \varepsilon_{2i}^{1} \\ y_{3i}^{*} = X_{3i}^{0}\beta_3^{0} + \varepsilon_{3i}^{0} & y_{3i}^{*} = X_{3i}^{1}\beta_3^{1} + \varepsilon_{3i}^{1} \\ \text{M} & \text{M} \\ y_{ki}^{*} = X_{ki}^{0}\beta_k^{0} + \varepsilon_{ki}^{0} & y_{ki}^{*} = X_{ki}^{1}\beta_k^{1} + \varepsilon_{ki}^{1} \end{array} \Bigg\} \text{ response equations} \qquad (1)$$

where X_{1i} and X_{ji}^{s} ($j = 2,...k \quad s = 0,1$) are vectors of explanatory variables, the ε_{ji}^{s} are error terms, and the mappings between the observed dependent variables y_{ji} and their latent counterparts y_{ji}^{*} are:

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

Thus, y_{1i} is dichotomous and variables y_{ji} 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 $b_j = \infty$). Variables y_{1i}^{*} and y_{ji}^{*} are not observed fully and, therefore, the problem can be considered as one with incomplete data. Disturbance vectors $\boldsymbol{\varepsilon}_i^{0} = \left(\varepsilon_i, \varepsilon_{2i}^{0}, ..., \varepsilon_{ki}^{0}\right)'$ and $\boldsymbol{\varepsilon}_i^{1} = \left(\varepsilon_i, \varepsilon_{2i}^{1}, ..., \varepsilon_{ki}^{1}\right)'$ follow *k*-variate normal distributions $N\left(0, \Omega_{s,i}\right)$, where covariance matrices $\Omega_{s,i}$ are heteroscedastic with moments

$$\begin{aligned} &\text{var}\left[\varepsilon_{1i}\right] = \sigma_{11}\exp\left(Z_{1i}\delta_1\right) = \exp\left(Z_{1i}\delta_1\right) \qquad E\left[\varepsilon_{1i}\varepsilon_{ji}^{s}\right] = \sigma_{1j}^{s}\exp\left(\frac{Z_{1i}\delta_1 + Z_{ji}\delta_j^{s}}{2}\right) \\ &E\left[\varepsilon_{j_1 i}^{s}\varepsilon_{j_2 i}^{s}\right] = \sigma_{j_1 j_2}^{s}\exp\left(\frac{Z_{j_1 i}\delta_{j_1}^{s} + Z_{j_2 i}\delta_{j_2}^{s}}{2}\right) \qquad j, j_1, j_2 = 2,3,4 \quad s = 0,1 \end{aligned} \qquad (2)$$

We have assumed multiplicative heteroscedasticity in (2) since it is one of the most flexible and commonly used forms for heteroskedasticity. Vectors, Z_{1i} and Z_{ji} are sets of heteroscedasticity determinants for the selection and response equations respectively, σ_{1j}^{s}, $\sigma_{j_1 j_2}^{s}$, δ_1, and δ_j^{s} are parameters to estimate, while $\sigma_{11} = 1$ to ensure identification in the dichotomous selection equation.

The complete information log-likelihood function for the equation system is

$$\ell^c(\boldsymbol{\beta},\boldsymbol{\theta}\,|\,\boldsymbol{y}) = -\frac{Nk}{2}\ln(2\pi) - \frac{1}{2}\left(\sum_{y_{1i}=0}\ln\left|\Omega_{0,i}\right| + \sum_{y_{1i}=1}\ln\left|\Omega_{1,i}\right|\right) - \frac{1}{2}\operatorname{tr}\left(\sum_{y_{1i}=0}\Omega_{0,i}^{-1}\boldsymbol{\varepsilon}_i^0\boldsymbol{\varepsilon}_i^{0'} + \sum_{y_{1i}=1}\Omega_{1,i}^{-1}\boldsymbol{\varepsilon}_i^1\boldsymbol{\varepsilon}_i^{1'}\right)$$

where $\boldsymbol{\beta} = \left(\beta_1 \quad \beta_2^0 \quad \cdots \quad \beta_k^0 \quad \beta_2^1 \quad \cdots \quad \beta_k^1\right)'$, N is the number of observations, $\Omega_{s,i}$ $(s=0,1)$ is the covariance matrix of $\boldsymbol{\varepsilon}_i^s$ defined according to the moments in (2), $\boldsymbol{\varepsilon}_i^s = \begin{pmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}^s\\ \vdots\\ \varepsilon_{ki}^s\end{pmatrix} = \begin{pmatrix} y_{1i}^* - X_{1i}\beta_1\\ y_{2i}^* - X_{2i}^s\beta_2^s\\ \vdots\\ y_{ki}^* - X_{ki}^s\beta_k^s\end{pmatrix}$, and $\boldsymbol{\theta} = \left(\sigma_{12}^0,\cdots,\sigma_{kk}^0,\sigma_{12}^1,\cdots,\sigma_{kk}^1,\ \delta_1,\delta_2^0,\cdots,\delta_k^0,\delta_2^1,\cdots,\delta_k^1\right)$.

E-Step. The conditional expectation of the log-likelihood can be written as:

$$\ell^c(\boldsymbol{\beta},\boldsymbol{\theta}\,|\,\boldsymbol{y}) = -\frac{Nk}{2}\ln(2\pi) - \frac{1}{2}\left(\sum_{y_{1i}=0}\ln\left|\Omega_{0,i}\right| + \sum_{y_{1i}=1}\ln\left|\Omega_{1,i}\right|\right) - \frac{1}{2}\operatorname{tr}\left(\sum_{y_{1i}=0}\Omega_{0,i}^{-1}E\left[\boldsymbol{\varepsilon}_i^0\boldsymbol{\varepsilon}_i^{0'}\right] + \sum_{y_{1i}=1}\Omega_{1,i}^{-1}E\left[\boldsymbol{\varepsilon}_i^1\boldsymbol{\varepsilon}_i^{1'}\right]\right)$$

The E-step at iteration $m+1$, requires the calculation of

$$Q_i^s\left(\boldsymbol{\beta}\,|\,\boldsymbol{\beta}^{(s,m)},\Omega_{s,i}^{(m)},\boldsymbol{y}_i\right) = E\left[\boldsymbol{\varepsilon}_i^s\boldsymbol{\varepsilon}_i^{s'}\,|\,\boldsymbol{\beta}^{(s,m)},\Omega_{s,i}^{(m)},\boldsymbol{y}_i\right] = \sigma_i^{2(s,m)} + \begin{pmatrix}\mu_{y_{1i}^*}^{(s,m)} - X_{1i}\beta_1\\ \mu_{y_{2i}^*}^{(s,m)} - X_{2i}^s\beta_2^s\\ \vdots\\ \mu_{y_{ki}^*}^{(s,m)} - X_{ki}^s\beta_k^s\end{pmatrix}\begin{pmatrix}\mu_{y_{1i}^*}^{(s,m)} - X_{1i}\beta_1\\ \mu_{y_{2i}^*}^{(s,m)} - X_{2i}^s\beta_2^s\\ \vdots\\ \mu_{y_{ki}^*}^{(s,m)} - X_{ki}^s\beta_k^s\end{pmatrix}'$$

where $\sigma_i^{2(s,m)} = \operatorname{Cov}\left(y_{1i}^*,...,y_{ki}^*\,|\,\boldsymbol{\beta}^{(m)},\Omega_{s,i}^{(m)},\boldsymbol{y}_i\right)$ and $\mu_{y_{ji}^*}^{(s,m)} = E\left[y_{ji}^*\,|\,\boldsymbol{\beta}^{(m)},\Omega_{s,i}^{(m)},\boldsymbol{y}_i\right]\quad j=1,...,k$

The covariance matrices $\sigma_i^{2(s,m)}$ and the vectors of means $\mu_{y_{ji}^*}^{(s,m)}$ can be estimated from Gibbs samples obtained from the joint distribution of $\boldsymbol{y}_i^* = \left(y_{1i}^*,\cdots,y_{ki}^*\right)'$ conditional on parameters $\left(\boldsymbol{\beta}^{(m)},\boldsymbol{\Omega}_{s,i}^{(m)}\right)$ and the observed information $\boldsymbol{y}_i$. Additionally, the distribution of $\left(y_{1i}^*,\cdots,y_{ki}^*\right)$ at iteration m is:

$$N\left(\mu_i^{(s,m)},\Omega_{s,i}^{(m)}\right), \text{ where } \mu_i^{(s,m)} = \begin{pmatrix} X_{1i}\beta_1^{(m)}\\ X_{2i}^s\beta_2^{(s,m)}\\ \vdots\\ X_{ki}^s\beta_k^{(s,m)}\end{pmatrix}$$

1.2. The Gibbs sampler

The moments $\sigma_i^{2(s,m)}$ and $\mu_{y_{ji}^*}^{(s,m)}$ could be calculated easily if the conditional (on observed information) densities of y_{ji}^* $(j=1,...,k)$ were known. However, obtaining those marginal densities may require solving high-dimensional integrals. Instead of tackling the problem by integration, the Gibbs sampler (Casella and George, 1992) provides a way to simulate samples 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)

$$\boldsymbol{y}_{i|-j}^* = \begin{pmatrix} y_{1i}^* \\ \vdots \\ y_{j-1i}^* \\ y_{j+1i}^* \\ \vdots \\ y_{ki}^* \end{pmatrix} \qquad \boldsymbol{X}_{i|-j}^s = \begin{pmatrix} X_{1,i} \\ \vdots \\ X_{j-1,i}^s \\ X_{j+1,i}^s \\ \vdots \\ X_{k,i}^s \end{pmatrix} \qquad \boldsymbol{\beta}_{-j}^{(s,m)} = \begin{pmatrix} \beta_1^{(m)} \\ \vdots \\ \beta_{j-1}^{(s,m)} \\ \beta_{j+1}^{(s,m)} \\ \vdots \\ \beta_k^{(s,m)} \end{pmatrix}$$

The implementation of the sampler begins with determining the distribution of each y_{ji}^* conditional on the vector of the rest of the dependent variables $\boldsymbol{y}_{i|-j}^*$. Under the normality assumption, these conditional distributions are univariate normal $N\left(\mu_{ji|i(-j)}^{(s)}, \sigma_{j|-j}^{2(s)}\right)$ and, at iteration $m+1$, means $\mu_{ji|i(-j)}^{(s)}$ and variances $\sigma_{j|-j}^{2(s)}$ can be estimated from

$$\begin{aligned}
\mu_{ji|i(-j)}^{(s,m)} &= E\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \beta^{(m)}, \Omega_{s,i}^{(m)}\right) \\
&= X_{ji}^s \beta_j^{(s,m)} + \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Omega_{s,i}^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^* \middle| \Omega_{s,i}^{(m)}\right)\right]^{-1}\left(\boldsymbol{y}_{i|-j}^* - \boldsymbol{X}_{i|-j}^s \boldsymbol{\beta}_{-j}^{(s,m)}\right) \\
\sigma_{j|-j}^{2(s,m)} &= \operatorname{var}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \beta^{(m)}, \Omega_{s,i}^{(m)}\right) \\
&= \operatorname{var}\left(y_{ji}^* \middle| \Omega_{s,i}^{(m)}\right) - \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Omega_{s,i}^{(m)}\right)\left[\operatorname{cov}\left(\boldsymbol{y}_{i|-j}^* \middle| \Omega_{s,i}^{(m)}\right)\right]^{-1} \operatorname{cov}\left(y_{ji}^* \middle| \boldsymbol{y}_{i|-j}^*, \Omega_{s,i}^{(m)}\right)'
\end{aligned}$$

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}^*. 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 mappings connecting $\boldsymbol{y}_i$ and $\boldsymbol{y}_i^*$.

The observed counterpart of y_{1i}^* in the first equation in (1) 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 $\mu_{1i|i(-1)}^{(s,m)}$ and variance $\sigma_{1i|-1}^{2(s,m)}$ truncated below at zero (see Devroye, 1986) 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 (a_j, b_j). Consequently, it is only necessary to simulate them when $y_{ji}=a_j$ or $y_{ji}=b_j$. We make the simulations from normal distributions with means $\mu_{ji|i(-j)}^{(s,m)}$ and variances $\sigma_{ji|-j}^{2(s,m)}$ truncated above at a_j when $y_{ji}=a_j$ and truncated below at b_j when $y_{ji}=b_j$. When $a_j < y_{ji} < b_j$, we set $y_{ji}^*=y_{ji}$.

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 estimates for $\sigma_i^{2(s,m)}$ and $\mu_{y_{ji}^*}^{(s,m)}$. Notice that when y_{ji}^* is fully observed (*i.e.* when $y_{ji}^*=y_{ji}$) then $\sigma_{y_{ri}^* y_{ti}^*}^{(s,m)}=0$ and $\mu_{y_{ji}^*}^{(s,m)}=y_{ij}$ $\forall\, m,r,s,t$.

1.3. Maximization Step

After obtaining $\sigma_i^{2(s,m)}$ and $\mu_i^{(s,m)}$, we move to the Maximization step. We maximize

$$\ell^c\left(\boldsymbol{\beta},\boldsymbol{\theta}\,|\,\beta^{(m)},\boldsymbol{\theta}^{(m)},\boldsymbol{y}\right)=-\frac{Nk}{2}\ln(2\pi)-\frac{1}{2}\left(\sum_{y_{1i}=0}\ln\left|\Omega_{0,i}\right|+\sum_{y_{1i}=1}\ln\left|\Omega_{1,i}\right|\right)-$$
$$\frac{1}{2}\mathrm{tr}\left(\sum_{y_{1i}=0}\Omega_{0,i}^{-1}Q_i^0\left(\boldsymbol{\beta}\,|\,\boldsymbol{\beta}^{(0,m)},\Omega_{0,i}^{(m)},\boldsymbol{y}_i\right)+\sum_{y_{1i}=1}\Omega_{1,i}^{-1}Q_i^1\left(\boldsymbol{\beta}\,|\,\boldsymbol{\beta}^{(1,m)},\Omega_{1,i}^{(m)},\boldsymbol{y}_i\right)\right) \tag{3}$$

with respect to the elements in $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. Notice that, except for the covariance matrices $\sigma_i^{2(s,m)}$ present in the $Q_i^s\left(\boldsymbol{\beta}\,|\,\boldsymbol{\beta}^{(s,m)},\Omega_{s,i}^{(m)},\boldsymbol{y}_i\right)$ terms, the expression in (3) resembles closely the sum of the log-likelihood functions of two systems of linear equations, where the unobserved information has been replaced by its expected values. The two equation systems correspond to the regimes $s=0$ and $s=1$, where the only parameters common to both regimes

are the slope vector β_1 and the heteroscedasticity term δ_1 in the selection equation. We split the M-step in two conditional maximization steps (Meng and Rubin, 1993). At iteration $m+1$, the first conditional maximization maximizes (3) with respect to $\boldsymbol{\beta}$ conditional on $\boldsymbol{\theta}^{(m)}$ to generate $\boldsymbol{\beta}^{(m+1)}$. The second maximization maximizes (3) with respect to the parameters in $\boldsymbol{\theta}$ conditional on the updated $\boldsymbol{\beta}^{(m+1)}$ to obtain $\boldsymbol{\theta}^{(m+1)}$.

From the first order conditions of (3), the maximizer in the first conditional maximization can be written as the GLS estimator

$$\boldsymbol{\beta}^{(m+1)} = \left[X_d^{'} \left(\sum_{y_{1i}=0} \tilde{\Omega}_i^{(0,m)} \otimes I_i^0 + \sum_{y_{1i}=1} \tilde{\Omega}_i^{(1,m)} \otimes I_i^1 \right) X_d \right]^{-1} X_d^{'} \left(\sum_{y_{1i}=0} \tilde{\Omega}_i^{(0,m)} \otimes I_i^0 + \sum_{y_{1i}=1} \tilde{\Omega}_i^{(1,m)} \otimes I_i^1 \right) \mu_{y^*}^{(m)}$$

where I_i^s is a $N \times N$ matrix which only non-zero element is $I_{ii}^s = 1$ if $y_{1i} = s$ and $I_{ii}^s = 0$ otherwise. Matrix X_d is block diagonal with X_1, X_2^0, $\cdots$, X_k^0, X_2^1, $\cdots$, X_k^1 in the diagonal. Additionally, $\mu_{y^*}^{(m)} = \left(\mu_{y_1^*}^{(m)} \quad \mu_{y_2^*}^{(0,m)} \quad \cdots \quad \mu_{y_6^*}^{(1,m)} \quad \mu_{y_2^*}^{(0,m)} \quad \cdots \quad \mu_{y_6^*}^{(1,m)} \right)^{'}$, where $\mu_{y_j^*}^{(s,m)} = \left(\mu_{y_{j1}^*}^{(s,m)} \quad \cdots \quad \mu_{y_{jN}^*}^{(s,m)} \right)^{'}$, and $\mu_{y_{ji}^*}^{(s,m)} = 0$ if $y_{1i} \neq s$. Finally, the matrices $\tilde{\Omega}_i^{(s,m)}$ are square matrices of dimension $2k-1$ defined by

$$\tilde{\Omega}_i^{(0,m)} = \begin{bmatrix} \tilde{\omega}_{11,i}^{(0,m)} & \tilde{\omega}_{12,i}^{(0,m)} & \cdots & \tilde{\omega}_{1k,i}^{(0,m)} & 0 & \cdots & 0 \\ \tilde{\omega}_{12,i}^{(0,m)} & \tilde{\omega}_{22,i}^{(0,m)} & \cdots & \tilde{\omega}_{2k,i}^{(0,m)} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ \tilde{\omega}_{1k,i}^{(0,m)} & \tilde{\omega}_{2k,i}^{(0,m)} & \cdots & \tilde{\omega}_{kk,i}^{(0,m)} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad \tilde{\Omega}_i^{(1,m)} = \begin{bmatrix} \tilde{\omega}_{11,i}^{(1,m)} & 0 & \cdots & 0 & \tilde{\omega}_{12,i}^{(1,m)} & \cdots & \tilde{\omega}_{1k,i}^{(1,m)} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \tilde{\omega}_{12,i}^{(1,m)} & 0 & \cdots & 0 & \tilde{\omega}_{22,i}^{(1,m)} & \cdots & \tilde{\omega}_{2k,i}^{(1,m)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\omega}_{1k,i}^{(1,m)} & 0 & \cdots & 0 & \tilde{\omega}_{2k,i}^{(1,m)} & \cdots & \tilde{\omega}_{kk,i}^{(1,m)} \end{bmatrix},$$

where $\tilde{\omega}_{rt,i}^{(s,m)}$ is the element on the *r*-th row and *t*-th column of the inverses of $\Omega_{s,i}^{(m)}$ and $\Omega_{s,i}^{(m)}$.

After plugging $\boldsymbol{\beta}^{(m+1)}$ in (3), parameters in $\boldsymbol{\theta}$ are estimated by maximizing

$$\ell^c\left(\boldsymbol{\theta} \mid \beta^{(m+1)}, \boldsymbol{\theta}^{(m)}, \mathbf{y}\right) = -\frac{Nk}{2}\ln(2\pi) - \frac{1}{2}\left(\sum_{y_{1i}=0} \ln\left|\Omega_{0,i}\right| + \sum_{y_{1i}=1} \ln\left|\Omega_{1,i}\right| \right) - \frac{1}{2}\operatorname{tr}\left(\sum_{y_{1i}=0} \Omega_{0,i}^{-1} Q_i^0\left(\boldsymbol{\beta}^{(0,m+1)} \mid \boldsymbol{\beta}^{(0,m)}, \Omega_{0,i}^{(m)}, \mathbf{y}_i\right) + \sum_{y_{1i}=1} \Omega_{1,i}^{-1} Q_i^1\left(\boldsymbol{\beta}^{(1,m+1)} \mid \boldsymbol{\beta}^{(1,m)}, \Omega_{1,i}^{(m)}, \mathbf{y}_i\right) \right) \tag{4}$$

with respect to the $k^2 + 3(k-1)$ different elements in $\boldsymbol{\theta}$. The maximization of (4) can be accomplished numerically and, since the expression in (4) is simple enough to obtain analytical formulae for the gradient and Hessian, they do not need to be estimated numerically during the optimization procedure. When the covariance matrices of the two regimes are independent from each other (as it happens under homoskedasticity), it is possible to split the maximization on $\boldsymbol{\theta}$ in two conditional maximizations: the first one maximizes

(4) on θ^0 conditional on $\beta^{(m+1)}$ and $\theta^{(1,m)}$, while the second one maximizes (4) on θ^1 conditional on $\beta^{(m+1)}$ and $\theta^{(0,m+1)}$. Our simulations show that this procedure accelerates the algorithm and reduces the possibility that estimates escape from the parameter space during the process.

1.4. Convergence and stopping rules

Wei and Tanner (1990) find more efficient to start 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. 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; Booth and Hobert 1999; Eickhoff *et al*., 2004).

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 applying the following criteria at iteration m:

$$\frac{1}{J}\sum_{j=m-J}^{m}\left|\frac{E\left[\ell^c\left(\vartheta^{(j)}\right)\right]-E\left[\ell^c\left(\vartheta^{(j-1)}\right)\right]}{E\left[\ell^c\left(\vartheta^{(j-1)}\right)\right]}\right|<10^{-3},\quad \max_r\left[\frac{1}{J}\sum_{j=m-J}^{m}\left|\frac{\vartheta_r^{(j)}-\vartheta_r^{(j-1)}}{\vartheta_r^{(j-1)}}\right|\right]<10^{-3} \qquad (5)$$

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 chosen by the researcher. In this work, J was set equal to $\min(50, 0.2m)$. These stopping rules require that the average change (both in the expected log-likelihood and in each component of the vector of estimates) in the last J iterations of the algorithm must be smaller than 10^{-3}. The algorithm was stopped only when both criteria were satisfied simultaneously for at least ten consecutive iterations to avoid false convergence.

1.5. Estimation of the Information matrix

We estimate the information matrix from a stochastic version of Louis' (1982) result. Let the complete information log-likelihood function be $\ell^c(\vartheta;\boldsymbol{z})$, where ϑ 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 given by

$I(\vartheta;\boldsymbol{y}) = I^c(\vartheta;\boldsymbol{z}) - I^m(\vartheta;\boldsymbol{z}\,|\,\boldsymbol{y})$, where $I^c(\vartheta;\boldsymbol{z}) = -E\left[H^c(\vartheta;\boldsymbol{z})\right]$ is the complete information matrix, $H^c(\vartheta;\boldsymbol{z})$ is the complete information Hessian, and $I^m(\vartheta;\boldsymbol{z}\,|\,\boldsymbol{y})$ is the "missing information" matrix. Louis showed that this last matrix can be written as $I^m(\vartheta;\boldsymbol{z}\,|\,\boldsymbol{y}) = \mathrm{Var}\left[S^c(\vartheta;\boldsymbol{z})\right] = E\left[S^c(\vartheta;\boldsymbol{z})S^c(\vartheta;\boldsymbol{z})'\right] - E\left[S^c(\vartheta;\boldsymbol{z})\right]E\left[S^c(\vartheta;\boldsymbol{z})'\right]$, where $S^c(\vartheta;\boldsymbol{z})$ is the complete information score vector. All the expectations are taken with respect to the distribution $f\left(\boldsymbol{z}\,|\,\boldsymbol{y},\vartheta^{EM}\right)$, where ϑ^{EM} is the final MCEM estimator. The evaluation of all the expectations involved commonly prevents the estimation of the observed information matrix by direct calculation. Monte Carlo estimates of the expected complete-information Hessian and score can be used to circumvent the problem. We proceed in three steps. First, simulate a Gibbs sequence $\boldsymbol{y}_i^{*(1)},\ldots,\boldsymbol{y}_i^{*(r_{burn}+2)}$ while holding $\vartheta = \vartheta^{EM}$ (we use $r_{burn} = 50k$ in this study). Second, let $\boldsymbol{y}_i^{*(r_{burn}+1)}$ be the sequence 1 and $\boldsymbol{y}_i^{*(r_{burn}+1)}, \boldsymbol{y}_i^{*(r_{burn}+2)}$ the sequence 2; calculate the information matrices for these two sequences according to

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

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

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

where analytical expressions for the contributions from each observation to the Hessian and score are standard results from the normal distribution theory. Third, calculate the corresponding vectors of standard errors for the two sequences; then use the second convergence criterion in (5) with $m = 2$ and using the standard error estimates instead of the parameter estimates. If the criterion is met, stop the procedure and take the standard errors calculated from sequence 2 as the final estimates. If the criterion is not met, add a new simulation $\boldsymbol{y}_i^{*(r_{burn}+3)}$ such that sequences 1 and 2 are now $\boldsymbol{y}_i^{*(r_{burn}+1)}, \boldsymbol{y}_i^{*(r_{burn}+2)}$ and $\boldsymbol{y}_i^{*(r_{burn}+1)},\cdots,\boldsymbol{y}_i^{*(r_{burn}+3)}$ respectively. Check convergence with $m = 3$. Keep adding sample simulations until convergence is attained. We found this procedure more effective that the one proposed by Natarajan (2000) and Ibrahim *et al.* (2001), who suggest simulating a sequence $\boldsymbol{y}_i^{*(1)},\ldots,\boldsymbol{y}_i^{*(r_{burn}+R)}$, calculating the information matrix for each one of the last R simulations, and then taking the average as the final estimator. Since no clue about how R should be estimated is provided, we found that, even for

relatively large R, some of the standard errors might not have converged or even the information matrix might not be positive definite (because convergence has not been attained). Our approach, on the other hand, guarantees convergence in all the standard errors.

2. lustration on simulated and real data

2.1. Analysis on simulated data

For our simulation, we considered that, for individuals in regime s, $\boldsymbol{y}_i^*$ comes from a 3-variate normal with mean $\left(X_{1i}\beta_1 \quad X_{2i}^s\beta_2^s \quad X_{3i}^s\beta_3^s\right)'$ and heteroscedastic covariance matrix $\boldsymbol{\Omega}_{(s,i)}$. Each equation contains two explanatory variables: an intercept and a variable, which was simulated from a standard normal distribution. We choose the following values for the model parameters: $\beta_1=(1,2)'$, $\beta_2^0=(1,0.5)'$, $\beta_3^0=(-1,\ 1)'$, $\left(\sigma_{12}^0,\sigma_{13}^0,\sigma_{22}^0,\sigma_{23}^0,\sigma_{33}^0\right)=(-0.5,-0.5,1,0,1)$, $\beta_2^1=(0.5,-1)'$, $\beta_3^1=(1,-1)'$, $\left(\sigma_{12}^1,\sigma_{13}^1,\sigma_{22}^1,\sigma_{23}^1,\sigma_{33}^1\right)=(0.5,0.5,1,0,1)$, and $\left(\delta_1,\delta_2^0,\delta_3^0,\delta_2^1,\delta_3^1\right)=(0.5,0.2,0.1,-0.1,-0.2)$. We run 500 simulated experiments; in each case, the estimates were obtained from samples of 300 simulated observations. Graphical outcomes in Figure 1 show that the algorithm works fairly well. All the estimations used OLS estimates as starting values; the Gibbs sampler was started with 300 simulations and increased in an amount of 15 Gibbs simulations per iteration, *i.e.* $K^{(m)}=300+15(m-1)$. All routines were programmed in Matlab.

2.2. Analysis on real data

We illustrate the algorithm by analyzing data from a survey administered to Maryland farmers to evaluate the performance of a multi-objective voluntary conservation program. The program provides funds to cost share the implementation of practices for the conservation of soil and wildlife habitat.

The first equation in our 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} is the proportion of the farm under permanent vegetative cover, while y_{3i} and y_{4i} are the proportions of the area cropped on which contour cropping and cover crops are used, respectively. All the responses are proportions and thus one or more of them are censored from below at cero for individuals who find the use of these practices unprofitable. Additionally, it might be possible to observe censoring at one for some individuals that find the practices highly profitable. The

matrices of explanatory variables X_1, X_2^0, X_3^0, and X_4^0 include information about farmer characteristics (such as age and formal education), farm topography, farm size, distance to water bodies, previous history of participation in conservation programs, and location dummies. Matrices X_2^1, X_3^1, and X_4^1 contain similar variables, but they also include the time since program funds were awarded to the farms.

OLS estimates were used as starting values. The Gibbs sampler was started with 400 simulations and increased in an amount of 15 simulations per iteration. Marginal effects and covariance matrix estimates are reported in Table 1. We evaluate the possibility that participants behave differently from non-participants by a joint test of the hypotheses $H_0: \beta_j^0 = \beta_j^1,\ \sigma_{1j}^0 = \sigma_{1j}^1,\ \sigma_{j_i j_2}^0 = \sigma_{j_i j_2}^1,\ \delta_j^0 = \delta_j^1 \quad (j, j_1, j_2 = 2,3,4)$. Additionally, we evaluate the existence of heteroscedasticity by testing $H_0:\ \delta_j^0 = 0 \quad (j = 2,3,4)$. Two Wald tests rejected both hypotheses ($w = 69.0,\ p < 0.01$, and $w = 24.0,\ p < 0.01$, respectively), which supports the use of a heteroscedastic switching regression approach.

3. Extensions to other types of responses and to panel data

3.1. Extensions for dichotomous and ordinal response variables

As it will become clear, only small adjustments to the Gibbs sampler and the maximization step are needed to generalize the algorithm to handle any type of response. Adjustments to the Gibbs sampler follow. First, if y_{ji} is dichotomous we proceed like in the selection equation, *i.e.*, we simulate y_{ji}^* from a normal distribution with mean $\mu_{ji|i(-j)}^{(s,m)}$ and variance $\sigma_{ji|-j}^{2(s,m)}$ truncated below at zero if y_{ji} equals one and truncated above at zero if y_{ji} equals zero. If y_{ji} is ordinal, we have:

$$y_{ji} = \begin{cases} l_{j,1} & \alpha_{j,1} < y_{ji}^* \le \alpha_{j,2} \\ l_{j,2} & \alpha_{j,2} < y_{ji}^* \le \alpha_{j,3} \\ \vdots & \vdots \\ l_{j,r} & \alpha_{j,r} < y_{ji}^* \le \alpha_{j,r+1} \end{cases} \quad \text{if}$$

where $l_{j,1} < l_{j,2} \cdots < l_{j,r}$ are consecutive integer values, $\alpha_{j,1} = -\infty$, $\alpha_{j,r+1} = \infty$. To satisfy spatial identification we need to set one of the thresholds (or the equation intercept) equal to a known value; we choose $\alpha_{k,2} = 0$, while thresholds $\alpha_{j,3} < \ldots < \alpha_{j,r}$ are extra parameters to estimate. Consequently, we

simulate y_{ji}^{*} from a normal distribution with mean $\mu_{ji|i(-j)}^{(s,m)}$, and variance $\sigma_{ji|-j}^{2(s,m)}$ truncated above at $\alpha_{j,t+1}$ and truncated below at $\alpha_{j,t}$ when y_{ji} equals $l_{j,t}$ $(t=1,...,r)$.

Regarding to the M-step, thresholds $\alpha_{j,3}<...<\alpha_{j,r}$ are not present in the complete-information likelihood function; therefore, they are not obtained from the maximization of (3). Instead, we proceed as it follows: i) at every round of the Gibbs sampler during iteration m, keep the minimum value of every sequence obtained when simulating the observations having $y_{ji}=l_{j,t}$; this produces a set of $K^{(m)}-k_{burn}$ values; ii) keep the maximum value of every sequence obtained when simulating the observations $y_{ji}=l_{j,t-1}$; iii) take the medians of the two sets obtained in i) and ii); and finally iv) take the average of the two medians to produces a consistent estimator of $\alpha_{j,t}$. Finally, either y_{ji} is dichotomous or ordinal, variances σ_{jj}^{0} and σ_{jj}^{1} must be set equal to 1 for scale identification.

3.2. Extension to panel data

Similar to the previous section, handling panel data requires some modifications to the Gibbs sampler and the maximization step. Consider the multivariate random effects model

$$y_{1i}^{*}=X_{1i}\beta_1+\varphi_{1i}+\varepsilon_{1it} \qquad \text{selection (switching) equation}$$

$$\begin{matrix} y_{1i}=0 \\ [\end{matrix} \qquad \begin{matrix} \\ \ ] \end{matrix} \qquad \begin{matrix} \\ y_{1i}=1 \end{matrix}$$

$$\left.\begin{array}{ll} y_{2i}^{*}=X_{2it}^{0}\beta_2^{0}+\varphi_{2i}^{0}+\varepsilon_{2it}^{0} & y_{2i}^{*}=X_{2it}^{1}\beta_2^{1}+\varphi_{2i}^{1}+\varepsilon_{2ti}^{1} \\ \text{M} & \text{M} \\ y_{ki}^{*}=X_{kit}^{0}\beta_k^{0}+\varphi_{ki}^{0}+\varepsilon_{kit}^{0} & y_{ki}^{*}=X_{kit}^{1}\beta_k^{1}+\varphi_{ki}^{1}+\varepsilon_{kti}^{1} \end{array}\right\} \text{ response equations}$$

where $\boldsymbol{\varphi}_i=\begin{pmatrix}\varphi_i^{0}\\ \varphi_i^{1}\end{pmatrix}\sim N(\mathbf{0},\Theta)$, $\varphi_i^{s}=\left(\varphi_{1i},\varphi_{2i}^{s},...,\varphi_{ki}^{s}\right)'$, and $\operatorname{var}(\boldsymbol{\varphi}_i)=\Theta=\begin{bmatrix}\Theta^{0} & \Theta_0^{1}\\ \Theta_0^{1} & \Theta^{1}\end{bmatrix}$, and the covariance between the random effects of the two regimes, Θ_0^{1}, can be identified only if there are enough observations changing regimen during the time of observation. The vector of latent responses for individual *i* is now $\boldsymbol{y}_i^{*}=\left(y_{1i}^{*'},...,y_{ji}^{*'},...,y_{ki}^{*'}\right)'$, where $\boldsymbol{y}_{ji}^{*}=\left(y_{ji1}^{*},\cdots,y_{jiT_i}^{*}\right)$, $j=1,...,k$. Although the Gibbs sampling is entirely analogous to the cross-sectional case, it must be taken into account that now $\operatorname{var}\left(\boldsymbol{y}_i^{*}\right)=\Lambda_i=\operatorname{var}\left(\boldsymbol{\varphi}_i+\boldsymbol{\varepsilon}_i\right)$ is a square matrix of dimension kT_i involving elements from Θ and $\boldsymbol{\theta}$. Thus, since the observations for individual i are connected through the individual random effects, the

simulation for y^*_{jit} must be made conditional on observed information and $\boldsymbol{y}^*_{i|-j,-t}$. The first conditional maximization produces:

$$\boldsymbol{\beta}^{(m+1)} = \left[\sum_{i=1}^{N} x_i' \left[\Lambda_i^{(m)} \right]^{-1} x_i \right]^{-1} \sum_{i=1}^{N} x_i' \left[\Lambda_i^{(m)} \right]^{-1} \mu_{y_i^*}^{(m)} ,$$

where x_i is a kT_i-column matrix constructed by taking the columns corresponding to individual i from X_d and concatenating them horizontally for periods $t = 1,...T_i$. After constructing the $Q_i\left(\boldsymbol{\beta}^{(m+1)} \mid \beta^{(m)}, \Theta^{(m)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{y}\right) = E\left[\left(\boldsymbol{\varphi}_i + \boldsymbol{\varepsilon}_i\right)\left(\boldsymbol{\varphi}_i + \boldsymbol{\varepsilon}_i\right)' \mid \boldsymbol{\beta}^{(m)}, \Theta^{(m)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{y}_i \right]$, the second conditional maximization optimizes

$$\ell^c\left(\Theta, \boldsymbol{\theta} \mid \beta^{(m+1)}, \Theta^{(m)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{y}\right) = -\frac{k}{2}\ln(2\pi)\sum_i T_i - \frac{1}{2}\sum_i \ln|\Lambda_i| - \frac{1}{2}\mathrm{tr}\left(\sum_i \Lambda_i^{-1} Q_i\left(\boldsymbol{\beta}^{(m+1)} \mid \beta^{(m)}, \Theta^{(m)}, \boldsymbol{\theta}^{(m)}, \boldsymbol{y}\right)\right)$$

with respect to Θ and $\boldsymbol{\theta}$. Since individuals can move from one regime to another as time goes, Λ_i is not necessarily a block diagonal matrix. Therefore we cannot split the sums in the expected log-likelihood function according to the two regimes like in (4), and the Λ_i must be constructed according to the behavior dynamics of each individual.

Conclusions

This article presented a MCEM algorithm for FILM estimation of endogenous switching models involving multiple censored responses and heteroscedastic disturbances. The algorithm does not require numerical integration; 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 have closed forms, it 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. The extension of the algorithm to handle different types of discrete responses is straightforward, as only small changes to the Gibbs sampler in the expectation step are needed. The extension to panel data is also simple, although it requires a modification of the maximization step as well.

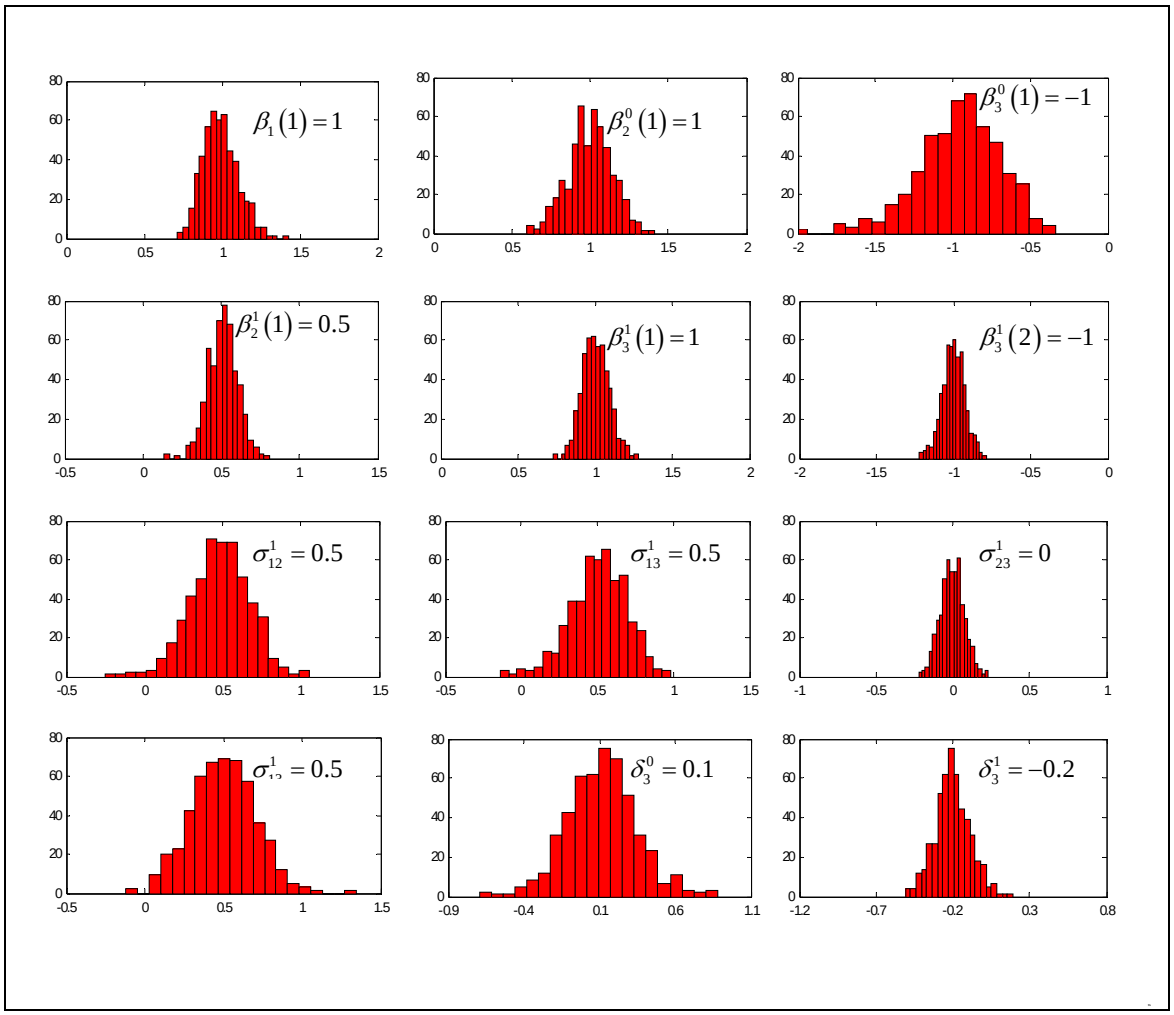
TABLE 1. MARGINAL EFFECTS AND COEFFICIENTS OF THE DISTURBANCE COVARIANCE MATRIX

Equation	Variable	Estimate	Std. error
Cost-Sharing	Age	-0.0035 [b]	0.0016
	College	0.0040	0.0615
	Highly	-0.0337	0.1336
	Moderate	0.0772	0.0703
	Land	0.8247 [a]	0.1698
	Waterbody	0.0717 [b]	0.0331

		Regime $s=0$		Regime $s=1$	
Equation	**Variable**	**Estimate**	**Std. error**	**Estimate**	**Std. error**
Vegetative Cover	Age	0.0013	0.0012	-0.0004	0.0018
	College	0.0688 [b]	0.0316	-0.0116	0.0429
	Highly	0.0606	0.0821	0.2525 [a]	0.0806
	Moderate	0.2073 [a]	0.0382	0.1003 [c]	0.0576
	Land	-0.4961	0.7372	0.5148	0.3182
	Time			-0.0324 [a]	0.0134
	Time^2			0.0025 [a]	0.0011
Contour cropping	Age	-0.0035 [a]	0.0010	-0.0037	0.0040
	College	-0.0337	0.0259	0.0126	0.0951
	Highly	0.0072	0.0702	0.3978 [a]	0.1315
	Moderate	0.1938 [a]	0.0335	0.1990	0.1897
	Land	-1.2992	0.8775	-1.2605 [b]	0.4998
	Time			-0.0575 [c]	0.0310
	Time^2			0.0044 [c]	0.0025
Cover crops	Age	-0.0017 [c]	0.0009	-0.0025	0.0042
	College	0.0308	0.0254	-0.0685	0.0878
	Highly	-0.1115	0.0725	-0.2117	0.1838
	Moderate	-0.0281	0.0329	-0.0054	0.1181
	Land	-0.1285	0.7534	-0.1653	1.1699
	Time			-0.0470 [c]	0.0294
	Time^2			0.0033	0.0024
Covariance matrix	σ_{12}	0.0994	0.0691	0.0047	0.0373
	σ_{13}	-0.2321	0.2201	0.4174 [a]	0.1468
	σ_{14}	-0.2427	0.2526	0.0283	0.2395
	σ_{22}	0.0694 [a]	0.0219	0.0102	0.0064
	σ_{23}	-0.0382 [b]	0.0180	0.0293	0.0210
	σ_{24}	0.0013	0.0083	0.0155	0.0113
	σ_{33}	0.2675 [a]	0.0995	0.2839 [c]	0.1576
	σ_{34}	0.0830 [b]	0.0384	0.0906	0.0816
	σ_{44}	0.1253 [b]	0.0573	0.0914	0.0568
	δ_1	1.1219 [a]	0.4082	1.1219 [a]	0.4082
	δ_2	-0.2127	0.1297	-0.8532 [a]	0.3018
	δ_3	-0.0207	0.2151	0.1901	0.2474
	δ_4	-0.3151 [c]	0.1638	-0.2132	0.3411

[a] Significant at 1% significance; [b] significant at 5% significance; [c] significant at 10% significance.

FIGURE 1. HISTOGRAMS FOR ESTIMATORS FROM SIMULATED DATA



References

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

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

Cameron, A., Trivedi P., 2005. Microeconometrics, Methods and Applications. Cambridge University Press, pp. 862-871.

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

Chan, K., Ledolter J., 1995. Monte Carlo EM Estimation for Time Series Models Involving Counts. J Am Stat Assoc 90(429): 242-252.

Dempster, A., Laird N., Rubin D., 1977. Maximum Likelihood Estimation from incomplete observations. J R Stat Soc B. 39:1-38.

Devroye, L., 1986. Non-uniform random variate generation. Springer-Verlag, pp. 38-39.

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.

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

Griliches, Z., Hall B., Hausman J., 1978. Missing Data and Self-Selection in Large Panels. Annals de l'INSEE 30-31:137-76.

Gronau, R., 1974. Wage Comparisons. A Selectivity Bias. *Journal of Political Economy*, 82(6): 1119-1143.

Heckman, J., 1976. The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5:475-92

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

Kenny, L., Lee F., Maddala G., Trost R., 1979. Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data. *International Economic Review*, 20(3): 775-789.

Lee, F., Maddala G., Trost R., 1980. Asymptotic covariance Matrices of two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity. *Econometrica*, 48(2): 491-503.

Lewis, H., 1974. Comments on Selectivity Biases in Wage Comparisons. *Journal of Political Economy*, 82(6): 1145-1155.

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

Maddala, G., Nelson F., 1974. Switching Regression Models with Exogenous and Endogenous Switching. Proceedings of the American Statistical Association (Bussiness and Economics Section), pp. 423-426.

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

Meng, X., Rubin D., 1993. Maximum Likelihood Estimation Via the ECM algorithm: a General Framework. *Biometrika*, 80: 267-278.

Natarajan, R., McCulloch C., Kiefer N., 2000. A Monte Carlo EM Method for estimating Multinomial Probit Models. Computational Statistics and Data Analysis 34: 33-50.

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

Roy, A., 1951. Some Thoughts on the Distribution of Earnings. Oxford Economics Papers 3: 135-46.

Wei, C., Tanner M., 1990. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. J Am Stat. Assoc 85(411): 699-704.

Novedades

DIVISIÓN DE ADMINISTRACIÓN PÚBLICA

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

De Angoitia, Regina, *La evolución de los portales del gobierno federal: la experiencia de ocho dependencias*, DTAP-196

Cabrero, Enrique, *De la descentralización como aspiración a la descentralización como problema*, DTAP-197

Sour, Laura y Eunises Rosillo, *¿Cuáles son los resultados del presupuesto por resultados?*, DTAP-198

Arellano, David y Walter Lepore, *Prevención y control de conflictos de interés: lecciones para la Administración Pública Federal en México…*, DTAP-199

Sour, Laura y Fredy Girón, *El efecto flypaper de las transferencias intergubernamentales del ramo 28…*, DTAP-200

Mariscal, Judith, *Convergencia tecnológica y armonización regulatoria en México: una evaluación de los instrumentos regulatorios*, DTAP-201

Mariscal, Judith, *Market Structure in the Latin American Mobile Sector*, DTAP-202

DIVISIÓN DE ECONOMÍA

Elbittar, Alexander, Rodrigo Harrison y Roberto Muñoz, *Network Structure in a Link-formation Game: An Experimental Study*, DTE-405

Carreón, Víctor y Malena Svarch, *Impacto del crédito en el bienestar*, DTE-406

Cermeño, Rodolfo, *Median-Unbiased Estimation in Panel Data…*, DTE-407

Unger, Kurt, *Apertura y empleos: la economía de los sectores comerciables y no comerciables de las regiones de México*, DTE-408

Gómez Aurora y Aldo Musacchio, *Public Finance, Foreign Creditors, and the Costs of the Mexican Revolution*, DTE-409

Gómez, Aurora, *Networks and Entrepreneurship: The Modernization of Textile Production and Distribution in Porfirian Mexico*, DTE-410

Gómez Aurora y Rodrigo Parral, *La revolución mexicana y la creación y disolución de las empresas*, DTE-411

Gómez, Aurora, *From Company Stores to Consumer's Cooperatives and the Worker's Bank in the Orizaba Valley Textile Milles*, DTE-412

Gómez, Aurora y César Guerrero, *To Open or Not to Open: The Causes, Timing and Consequences of Protectionism*, DTE-413

Scott, John, *Salud, sistemas de salud y pobreza: una síntesis de la literatura*, DTE-414

DIVISIÓN DE ESTUDIOS INTERNACIONALES

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

Sotomayor, Arturo, *Los métodos cualitativos en la ciencia política contemporánea*, DTEI-160

González, Guadalupe, *Democratización y política exterior: ¿el fin del predominio presidencial?*, DTEI-161

González, Guadalupe, *Percepciones sociales sobre la migración en México y Estados Unidos: ¿hay espacios para cooperar?*, DTEI-162

Bernhard, William y David Leblang, *Standing Tall When the Wind Shifts: Financial Market Responses to Elections, Disasters and Terrorist Attacks*, DTEI-163

Velázquez, Rafael, *La relación entre el Ejecutivo y el Congreso en materia de política exterior durante el sexenio de Vicente Fox…*, DTEI-164

Ruano, Lorena, *De la exaltación al tedio: las relaciones entre México y la Unión Europea…*, DTEI-165

Martínez, Ferrán e Ignacio Lago Peñas, *Why new Parties? Changes in the number of Parties over time within Countries*, DTEI-166

Sotomayor, Arturo, *México y la ONU en momentos de transición: entre el activismo internacional, parálisis interna y crisis internacional*, DTEI-167

DIVISIÓN DE ESTUDIOS JURÍDICOS

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

Magaloni, Ana Laura y Ana María Ibarra Olguín, *La configuración jurisprudencial de los derechos fundamentales…*, DTEJ-24

Magaloni, Ana Laura, *¿Por qué la Suprema Corte no ha sido un instrumento para la defensa de derechos fundamentales?*, DTEJ-25

Magaloni, Ana Laura, *Arbitrariedad e ineficiencia de la procuración de justicia: dos caras de la misma moneda*, DTEJ-26

Ibarra, Ana María, *Los artificios de la Dogmática Jurídica*, DTEJ-27

DIVISIÓN DE ESTUDIOS POLÍTICOS

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

Hacker, Casiano y Jeffrey Thomas, *Operationalizing and Reconstructing the Theory of Nationalism*, DTEP-198

Langston, Joy y Allyson Benton, *"A ras de suelo": Candidate Appearances and Events in Mexico's Presidential Campaign*, DTEP-199

Negretto, Gabriel, *The Durability of Constitutions in Changing Environments…*, DTEP-200

Langston, Joy, *Hasta en las mejores familias: Madrazo and the PRI in the 2006 Presidential Elections*, DTEP-201

DIVISIÓN DE HISTORIA

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

Barrón, Luis, *Revolucionarios sí, pero Revolución no*, DTH-44

Pipitone, Ugo, *Oaxaca: comunidad, instituciones, vanguardias*, DTH-45

Barrón, Luis, *Venustiano Carranza: un político porfiriano en la Revolución*, DTH-46

Tenorio, Mauricio y Laurencio Sanguino, *Orígenes de una ciudad mexicana: Chicago y la ciencia del* Mexican Problem *(1900-1930)*, DTH-47

Rojas, Rafael, *José María Heredia y la tradición republicana*, DTH-48

Rojas, Rafael, *Traductores de la libertad: el americanismo de los primeros republicanos*, DTH-49

Sánchez, Mónica Judith, *History* vs. *the Eternal Present or Liberal Individualism and the Morality of Compassion and Trust*, DTH-50

Medina, Luis, *Salida: los años de Zedillo*, DTH-51

Sauter, Michael, *The Edict on Religion of 1788 and the Statistics of Public Discussion in Prussia*, DTH-52

Ventas

El **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 Centro 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 cualquiera de estas publicaciones, le ofrecemos las siguientes opciones:

VENTAS DIRECTAS:	VENTAS EN LÍNEA:
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.	Librería virtual: **www.e-cide.com** Dudas y comentarios: **publicaciones@cide.edu**

¡¡Colecciones completas!!

Adquiere los **CDs de las colecciones completas** de los documentos de trabajo de todas las divisiones académicas del CIDE: **Economía**, **Administración Pública**, **Estudios Internacionales**, **Estudios Políticos**, **Historia** y **Estudios Jurídicos.**

¡Nuevo! ¡¡Arma tu CD!!

Visita nuestra Librería Virtual www.e-cide.com y selecciona entre 10 y 20 documentos de trabajo. A partir de tu lista te enviaremos un CD con los documentos que elegiste.