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

• D.R. © 2008. Centro de Investigación y Docencia Económicas, carretera México-Toluca 3655 (km. 16.5), Lomas de Santa Fe, 01210, México, D.F.
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 estimator 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 or 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 FILM 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 \( z \in Z \rightarrow y = y(z) \in Y \). In words, \( z \) is only known to lie in \( Z(y) \), the subset of \( Z \) determined by the equation \( y = y(z) \), where \( y \) is the observed data and \( z \) is the complete but not fully observed information. Thus, the log-likelihood of the observed information is \( l(y|y) = \ln L(y|y) = \ln \int_{Z(y)} L(y|z)dz \), where \( \theta \) is the parameter to be estimated. Instead of solving the integrals in \( l(y|y) \) directly, the deterministic version of the EM algorithm (Dempster et al., 1977) focuses on the complete-information log-likelihood \( \ell^c(\theta|z) \) and maximizes the expectation \( E[1^c(\theta|z)|y] \) by executing two steps iteratively: (1) the expectation step (E-step), which computes \( Q(\theta|\theta^m,y) = E[1^c(\theta|z)|y] \) at iteration \( m+1 \), and (2) the maximization step (M-step), which maximizes \( Q(\theta|\theta^m,y) \) to find \( \theta^{m+1} \). 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 \( Q(\theta|\theta^m,y) \) is approximated by the mean \( \frac{1}{K} \sum_{k=1}^{K} Q(\theta,z^{(k)}|y) \), where the \( z^{(k)} \) are random draws from \( f(z|\theta^m,y) \).
1.1. The algorithm

Let consider the \( k \)-equation system:

\[
\begin{align*}
    y_{1i}^* &= X_{1i}\beta_1 + \varepsilon_{1i} \\
    y_{2i}^* &= X_{2i}\beta_2^0 + \varepsilon_{2i}^0 \\
    y_{3i}^* &= X_{3i}\beta_3^0 + \varepsilon_{3i}^0 \\
    \vdots \\
    y_{ki}^* &= X_{ki}\beta_k^0 + \varepsilon_{ki}^0 \\
\end{align*}
\]

where \( y_{1i}^* \) is the selection (switching) equation and \( y_{2i}^*, y_{3i}^*, \ldots, y_{ki}^* \) are the response equations.

The system can be rewritten as:

\[
\begin{align*}
    y_{1i}^* &= X_{1i}\beta_1 + \varepsilon_{1i} \\
    y_{2i}^* &= X_{2i}\beta_2^0 + \varepsilon_{2i}^0 \\
    y_{3i}^* &= X_{3i}\beta_3^0 + \varepsilon_{3i}^0 \\
    \vdots \\
    y_{ki}^* &= X_{ki}\beta_k^0 + \varepsilon_{ki}^0 \\
\end{align*}
\]

where \( X_{1i}, X_{2i}, \ldots, X_{ki} \) are vectors of explanatory variables, the \( \varepsilon_{1i}, \varepsilon_{2i}^0, \ldots, \varepsilon_{ki}^0 \) are error terms, and the mappings between the observed dependent variables \( y_{ji} \) and their latent counterparts \( y_{ji}^* \) are:

\[
\begin{align*}
    y_{ji} &= \begin{cases} 
        1 & \text{if } y_{ji}^* > 0 \\
        0 & \text{if } y_{ji}^* \leq 0
    \end{cases} \\
    y_{ji}^* &= \begin{cases} 
        b_j & \text{if } y_{ji}^* \geq b_j \\
        a_j & \text{if } y_{ji}^* \leq a_j \quad j = 2, \ldots, k
    \end{cases}
\end{align*}
\]

Thus, \( y_{ji} \) 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 \( y_{ji}^* \) and \( y_{ji}^* \) are not observed fully and, therefore, the problem can be considered as one with incomplete data.

Disturbance vectors \( \varepsilon_i = (\varepsilon_{1i}, \varepsilon_{2i}^0, \ldots, \varepsilon_{ki}^0) \) and \( \varepsilon_i^1 = (\varepsilon_{1i}^1, \varepsilon_{2i}^1, \ldots, \varepsilon_{ki}^1) \) follow \( k \)-variate normal distributions \( N(0, \Omega_{s,j}) \), where covariance matrices \( \Omega_{s,j} \) are heteroscedastic with moments

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

We have assumed multiplicative heteroscedasticity in (2) since it is one of the most flexible and commonly used forms for heteroscedasticity. Vectors, \( Z_{1i} \) and \( Z_{ji} \), are sets of heteroscedasticity determinants for the selection and response equations respectively, \( \sigma_{11}^\delta, \sigma_{11}^\delta^s, \delta_1, \) and \( \delta_j^s \) 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

\[
\ell^c (\beta, \theta \mid y) = -\frac{Nk}{2} \ln (2\pi) - \frac{1}{2} \left( \sum_{y_{ij} = 0} \ln |\Omega_{ij,0}^o| + \sum_{y_{ij} = 1} \ln |\Omega_{ij,1}^o| \right) - \frac{1}{2} \text{tr} \left( \sum_{y_{ij} = 0} \Omega_{ij,0}^o \epsilon_i^o \epsilon_i^o' + \sum_{y_{ij} = 1} \Omega_{ij,1}^o \epsilon_i^1 \epsilon_i^1' \right)
\]

where \( \beta = (\beta_1, \beta_2^0, \ldots, \beta_k^0, \beta_2^1, \ldots, \beta_k^1) \), \( N \) is the number of observations, \( \Omega_{s,i} \) \( (s=0,1) \) is the covariance matrix of \( \epsilon_i^s \) defined according to the moments in (2),

\[\epsilon_i^s = \begin{pmatrix} \epsilon_{ii}^s \\ \vdots \\ \epsilon_{ki}^s \end{pmatrix} = \begin{pmatrix} y_{ii}^* - X_{ii} \beta_1^s \\ y_{2i}^* - X_{2i} \beta_2^s \\ \vdots \\ y_{ki}^* - X_{ki} \beta_k^s \end{pmatrix}, \quad \text{and} \quad \theta = (\sigma_0^0, \ldots, \sigma_k^0, \sigma_{12}^1, \ldots, \sigma_{kk}^1, \delta_1, \delta_{21}, \ldots, \delta_k, \delta_{12}, \ldots, \delta_{kk}).\]

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

\[
\ell^c (\beta, \theta \mid y) = -\frac{Nk}{2} \ln (2\pi) - \frac{1}{2} \left( \sum_{y_{ij} = 0} \ln |\Omega_{ij,0}^o| + \sum_{y_{ij} = 1} \ln |\Omega_{ij,1}^o| \right) - \frac{1}{2} \text{tr} \left( \sum_{y_{ij} = 0} \Omega_{ij,0}^o E[\epsilon_i^0 \epsilon_i^0'] + \sum_{y_{ij} = 1} \Omega_{ij,1}^o E[\epsilon_i^1 \epsilon_i^1'] \right)
\]

The E-step at iteration \( m+1 \), requires the calculation of

\[
Q^c (\beta \mid \beta^{(s,m)}, \Omega_{s,i}^{(m)}, y_i) = E[\epsilon_i^s \epsilon_i^s' \mid \beta^{(s,m)}, \Omega_{s,i}^{(m)}, y_i] = \sigma_i^{(s,m)} + \begin{pmatrix} \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_1^s \\ \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_2^s \\ \vdots \\ \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_k^s \end{pmatrix}^T \begin{pmatrix} \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_1^s \\ \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_2^s \\ \vdots \\ \mu_{y_{ij}}^{(s,m)} - X_{ij} \beta_k^s \end{pmatrix}
\]

where \( \sigma_i^{(s,m)} = \text{Cov}(y_{ij}^*, \ldots, y_{kj}^* \mid \beta^{(s,m)}, \Omega_{s,i}^{(m)}, y_i) \) and \( \mu_{y_{ij}}^{(s,m)} = E[y_{ij}^* \mid \beta^{(m)}, \Omega_{s,i}^{(m)}, y_i] \) for \( j = 1, \ldots, k \).

The covariance matrices \( \sigma_i^{2(s,m)} \) and the vectors of means \( \mu_{y_{ij}}^{(s,m)} \) can be estimated from Gibbs samples obtained from the joint distribution of \( y_i^* = (y_{i1}^*, \ldots, y_{ki}^*) \) conditional on parameters \( (\beta^{(m)}, \Omega_{s,i}^{(m)}) \) and the observed information \( y_i \). Additionally, the distribution of \( (y_{i1}^*, \ldots, y_{ki}^*) \) at iteration \( m \) is:

\[
N(\mu_i^{(s,m)}, \Omega_{s,i}^{(m)}), \quad \text{where} \quad \mu_i^{(s,m)} = \begin{pmatrix} X_{1i} \beta_1^{(m)} \\ X_{2i} \beta_2^{(s,m)} \\ \vdots \\ X_{ki} \beta_k^{(s,m)} \end{pmatrix}
\]
1.2. The Gibbs sampler

The moments $\sigma^2_{j, (s, m)}$ and $\mu_{j, (s, m)}$ could be calculated easily if the conditional (on observed information) densities of $y^*_j$ ($j = 1, \ldots, k$) 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)

$$
\begin{align*}
\mathbf{y}^*_{i-j} &= \begin{pmatrix}
y^*_{i1} \\
\vdots \\
y^*_{ij} \\
\vdots \\
y^*_{ik}
\end{pmatrix} \\
\mathbf{X}^*_{i-j} &= \begin{pmatrix}
X^*_{i1} \\
\vdots \\
X^*_{ij} \\
\vdots \\
X^*_{ik}
\end{pmatrix} \\
\mathbf{\beta}^{(s, m)}_{i-j} &= \begin{pmatrix}
\beta^1_{i-j} \\
\vdots \\
\beta^k_{i-j}
\end{pmatrix}
\end{align*}
$$

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

$$
\begin{align*}
\mu_{j|-(i-j)}^{(s, m)} &= E\left(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j}\right) \\
&= X^t_{j,i-j} \mathbf{\beta}^{(m)}_{i-j} + \text{cov}(y^*_{j|i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \Omega^{(m)}_{i-j})^{-1} (y^*_{j|i-j} - X^t_{j,i-j} \mathbf{\beta}^{(s, m)}_{i-j}) \\
\sigma^2_{j|-(i-j)} &= \text{var}\left(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j}\right) \\
&= \text{var}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j}) - \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \\
&= \text{var}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j}) - \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \\
&= \text{var}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j}) - \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1} \text{cov}(y^*_{j|i-j}, \mathbf{X}_{i-j}, \mathbf{\beta}^{(m)}_{i-j}, \Omega^{(m)}_{i-j})^{-1}
\end{align*}
$$

The next step is to sample iteratively from these conditional distributions in order to simulate a sample for the unobserved values of each $y^*_j$. Since the simulations for $y^*_j$ must be done conditional on its corresponding observed information $y_i$, the implementation procedure depends on the mappings connecting $y_i$ and $y^*_i$.
The observed counterpart of $y_{i1}$ in the first equation in (1) is dichotomous with $y_{i1}^*$ being positive if $y_{i1}$ equals one and non-positive if $y_{i1}$ equals zero. Accordingly, we simulate $y_{i1}^*$ from a normal distribution with mean $\mu_{i1|i-1}^{(s,m)}$ and variance $\sigma_{i1|i-1}^{2(s,m)}$ truncated below at zero (see Devroye, 1986) if $y_{i1}$ equals one and truncated above at zero if $y_{i1}$ equals zero.

Variables $y_{ij}^*$ ($j=2,...,k$) are observed in the interval $(a_j, b_j)$. Consequently, it is only necessary to simulate them when $y_{ij} = a_j$ or $y_{ij} = b_j$. We make the simulations from normal distributions with means $\mu_{ji|i-j}^{(s,m)}$ and variances $\sigma_{ji|i-j}^{2(s,m)}$ 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 $y_i^*$ 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 $y_{i1}^{(1)},...,y_{i1}^{(k(m))}$, 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 $\mu_{i}^{(s,m)}$. Notice that when $y_{ji}^*$ is fully observed (i.e. when $y_{ji}^* = y_{ji}$) then $\sigma_{ji}^{2(s,m)} = 0$ and $\mu_{ji}^{(s,m)} = y_{ji}^*$ $\forall m,r,s,t$.

1.3. Maximization Step

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

$$
\ell^c(\beta, \theta | \beta^{(m)}, \theta^{(m)}, y) = -\frac{Nk}{2} \ln(2\pi) \left( \frac{1}{2} \sum_{j=0}^{n} \ln |\Omega_{\theta_{j}}| + \sum_{j=1}^{m} \ln |\Omega_{\beta_{j}}| \right) - \frac{1}{2} \sum_{j=0}^{n} \ln |\Omega_{\theta_{j}}| \prod_{j=1}^{m} |\Omega_{\beta_{j}}| - \frac{1}{2} \sum_{j=0}^{n} \ln |\Omega_{\beta_{j}}| \prod_{j=1}^{m} |\Omega_{\theta_{j}}| \left( \beta^{(m)}, \theta^{(m)}, y \right)
$$

with respect to the elements in $\beta$, and $\theta$. Notice that, except for the covariance matrices $\sigma_{i}^{2(s,m)}$ present in the $Q_{i}^\prime \left( \beta^{(s,m)}, \theta^{(s,m)} | y \right)$ 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 $\beta_i$ and the heteroscedasticity term $\delta_i$ 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 $\beta$ conditional on $\theta^{(m)}$ to generate $\beta^{(m+1)}$. The second maximization maximizes (3) with respect to the parameters in $\theta$ conditional on the updated $\beta^{(m+1)}$ to obtain $\theta^{(m+1)}$.

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

$$
\beta^{(m+1)} = \left(X_d' \left( \sum_{j_i = 0} \Omega_i^{(0,m)} \otimes I_i^0 + \sum_{j_i = 1} \Omega_i^{(1,m)} \otimes I_i^1 \right) X_d \right)^{-1} X_d' \left( \sum_{j_i = 0} \Omega_i^{(0,m)} \otimes I_i^0 + \sum_{j_i = 1} \Omega_i^{(1,m)} \otimes I_i^1 \right) \mu^{(m)}_{j_i}
$$

where $I_i^s$ is a $N \times N$ matrix which only non-zero element is $I_i^s = 1$ if $y_{li} = s$ and $I_i^s = 0$ otherwise. Matrix $X_d$ is block diagonal with $X_1$, $X_2^0$, $X_2^1$, $X_3^0$, $X_3^1$, in the diagonal. Additionally, $\mu^{(m)}_{j_i} = \left( \mu^{(m)}_{j_{i1}, \ldots, \mu^{(m)}_{j_{iN}}} \right)$, where $\mu^{(s,m)}_{j_i} = \left( \mu^{(s,m)}_{j_{i1}, \ldots, \mu^{(s,m)}_{j_{iN}}} \right)$, and $\mu^{(s,m)}_{j_i} = 0$ if $y_{li} \neq s$. Finally, the matrices $\Omega_i^{(s,m)}$ are square matrices of dimension $2k-1$ defined by

$$
\Omega_i^{(s,m)} = \begin{bmatrix}
\alpha^{(s,m)}_{11} & \ldots & \alpha^{(s,m)}_{12} & \ldots & \alpha^{(s,m)}_{1k} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\alpha^{(s,m)}_{k1} & \ldots & \alpha^{(s,m)}_{kk} & \ldots & \alpha^{(s,m)}_{kk}
\end{bmatrix}
$$

where $\alpha^{(s,m)}_{ri,j}$ is the element on the $r$-th row and $t$-th column of the inverses of $\Omega_{si}^{(m)}$ and $\Omega_{ti}^{(m)}$.

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

$$
L^C(\theta | \beta^{(m+1)}, \theta^{(m)}, y) = -\frac{NK}{2} \ln (2\pi) - \frac{1}{2} \sum \ln |\Omega_{si}^{(m)}| + \sum \ln |\Omega_{ti}^{(m)}| -
\frac{1}{2} \ln \left( \sum_{j_i = 0} \Omega_{si}^{(s,m)} | \beta^{(s,m)}, \Omega_{ti}^{(m)} |, y_{ij} \right) + \sum_{j_i = 1} \Omega_{ti}^{(s,m)} | \beta^{(s,m)}, \Omega_{si}^{(m)} |, y_{ij} \right)
$$

with respect to the $k^2 + 3(k-1)$ different elements in $\theta$. 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 $\theta$ in two conditional maximizations: the first one maximizes...
(4) on $\theta^0$ conditional on $\beta^{(m+1)}$ and $\theta^{(1,m)}$, while the second one maximizes (4) on $\theta^i$ 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 $\left| E[\ell^c(\varphi^{(j)})] - E[\ell^c(\varphi^{(j-1)})] \right|$ or parameter differences $\left| \varphi^{(j)} - \varphi^{(j-1)} \right|$, where $\varphi^{(j)}$ 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 rules 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 $m$:

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

where $\varphi_r^{(j)}$ 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 $\ell^c(\varphi;z)$, where $\varphi$ 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
A Monte Carlo EM Algorithm for FIML estimation of multivariate endogenous switching models with censored and discrete responses

$I(\mathcal{G}; y) = I^c(\mathcal{G}; z) - I^m(\mathcal{G}; z | y)$, where $I^c(\mathcal{G}; z) = -E\left[H^c(\mathcal{G}; z)\right]$ is the complete information matrix, $H^c(\mathcal{G}; z)$ is the complete information Hessian, and $I^m(\mathcal{G}; z | y)$ is the “missing information” matrix. Louis showed that this last matrix can be written as

$I^m(\mathcal{G}; z | y) = \text{Var}\left[S^c(\mathcal{G}; z)\right] = E\left[S^c(\mathcal{G}; z)S^c(\mathcal{G}; z)^\prime\right] - E\left[S^c(\mathcal{G}; z)^\prime\right]E\left[S^c(\mathcal{G}; z)\right]$, where $S^c(\mathcal{G}; z)$ is the complete information score vector. All the expectations are taken with respect to the distribution $f(z | y, \mathcal{G}^{EM})$, where $\mathcal{G}^{EM}$ 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 $y_i^{*(1)}, \ldots, y_i^{*(m+1)}$ while holding $\mathcal{G} = \mathcal{G}^{EM}$ (we use $r_{\text{burn}} = 50k$ in this study). Second, let $y_i^{*(m+1)}, \ldots, y_i^{*(m+2)}$ be the sequence 2; calculate the information matrices for these two sequences according to

$I^c(\mathcal{G}^{EM}; z) = -\sum_{i=1}^{N} E\left[H^c(\mathcal{G}^{EM}; z_i)\right] \approx -\sum_{i=1}^{N} \frac{1}{R} \sum_{r=1}^{R} H^c_r(\mathcal{G}^{EM}; y_i^{(r)} | y_i)$

$I^m(\mathcal{G}^{EM}; x | y) = \sum_{i=1}^{N} \left\{ E\left[S^c(\mathcal{G}^{EM}; z_i)S^c(\mathcal{G}^{EM}; z_i)^\prime\right] - E\left[S^c(\mathcal{G}^{EM}; z_i)^\prime\right]E\left[S^c(\mathcal{G}^{EM}; z_i)\right] \right\} \approx \sum_{i=1}^{N} \left\{ \frac{1}{R} \sum_{r=1}^{R} S^c_r(\mathcal{G}^{EM}; y_i^{(r)} | y_i)S^c_r(\mathcal{G}^{EM}; y_i^{(r)} | y_i)^\prime - \frac{1}{R} \sum_{r=1}^{R} S^c_r(\mathcal{G}^{EM}; y_i^{(r)} | y_i)^\prime S^c_r(\mathcal{G}^{EM}; y_i^{(r)} | y_i) \right\}$,

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 $y_i^{*(m+3)}$ such that sequences 1 and 2 are now $y_i^{*(m+1)}, \ldots, y_i^{*(m+2)}$ and $y_i^{*(m+1)}, \ldots, y_i^{*(m+3)}$ 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 $y_i^{*(1)}, \ldots, y_i^{*(m+R)}$, 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. Illustration on simulated and real data

2.1. Analysis on simulated data

For our simulation, we considered that, for individuals in regime $s$, $y^*_i$ comes from a 3-variate normal with mean $(X_i \beta_1 X'_i \beta_2 X'_i \beta_3')$ and heteroscedastic covariance matrix $\Omega_{(s,i)}$. 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: $\beta_1 = (1,2)'$, $\beta_2^0 = (1,0.5)'$, $\beta_3^0 = (-1, 1)'$, $(\sigma_{12}^0, \sigma_{13}^0, \sigma_{22}^0, \sigma_{23}^0, \sigma_{33}^0) = (-0.5, -0.5, 1, 0, 1)$, $\beta_1^0 = (0.5, -1)'$, $\beta_2^0 = (1, -1)'$, $(\sigma_{12}^1, \sigma_{13}^1, \sigma_{22}^1, \sigma_{23}^1, \sigma_{33}^1) = (0.5, 0.5, 1, 0, 1)$, and $(\delta_1, \delta_2, \delta_3, \delta_2, \delta_3) = (0.5, 0.2, 0.1, -0.1, -0.2)$. 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 zero 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 $X_1, X_0, X_3^0$, 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 $X_1^1, X_1^1$, 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

$$H_0 : \beta_j^0 = \beta_j^1, \sigma_{ij}^0 = \sigma_{ij}^1, \sigma_{ij,j}^0 = \sigma_{ij,j}^1, \delta_j^0 = \delta_j^1 \quad (j, j_1, j_2 = 2, 3, 4).$$

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 $\mu_{ji}^{s(m)}$ and variance $\sigma_{ji}^{2(s,m)}$ 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:

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

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 $\alpha_{j,3} < \cdots < \alpha_{j,r}$ are extra parameters to estimate. Consequently, we
simulate \( y_{ji}^* \) from a normal distribution with mean \( \mu_{ji}^{(s,m)} \), and variance \( \sigma_{ji}^{2(s,m)} \) truncated above at \( \alpha_{j,i+1} \) and truncated below at \( \alpha_{j,i} \) when \( y_{ji} \) equals \( l_{j,i} \) \( (t=1,\ldots,r) \).

Regarding to the M-step, thresholds \( \alpha_{j,3}<\ldots<\alpha_{j,r} \) 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 \( y_{ji} = l_{j,i} \); this produces a set of \( K^{(m)} - k_{burn} \) values; ii) keep the maximum value of every sequence obtained when simulating the observations \( y_{ji} = l_{j,i-1} \); iii) take the medians of the two sets obtained in i) and ii); and finally iv) take the average of the two medians to produce a consistent estimator of \( \alpha_{j,i} \). Finally, either \( y_{ji} \) is dichotomous or ordinal, variances \( \sigma_{ji}^0 \) and \( \sigma_{ji}^1 \) 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

\[
\begin{align*}
    y_{i1}^* &= X_{i1} \beta_1 + \varphi_{i1} + \epsilon_{i1}^0 \quad \text{selection (switching) equation} \\
    y_{2i}^* &= X_{2i}^0 \beta_2^0 + \varphi_{2i}^0 + \epsilon_{2i}^0 \\
    y_{2i}^+ &= X_{2i}^1 \beta_2^1 + \varphi_{2i}^1 + \epsilon_{2i}^1 \\
    y_{ki}^* &= X_{kri} \beta_k^0 + \varphi_{kri}^0 + \epsilon_{kri}^0 \\
    y_{ki}^+ &= X_{kri}^1 \beta_k^1 + \varphi_{kri}^1 + \epsilon_{kri}^1 \\
    y_i &= \begin{bmatrix} y_{i1}^+ \cdots y_{ki}^+ \end{bmatrix}'
\end{align*}
\]

where \( \varphi_i = \begin{pmatrix} \phi_i^0 \\ \phi_i^1 \end{pmatrix} \sim N(0, \Theta) \), \( \varphi_i^* = \begin{pmatrix} \phi_{i1}, \phi_{2i1}, \ldots, \phi_{ki1} \end{pmatrix} \), and \( \text{var}(\varphi_i) = \Theta = \begin{bmatrix} \Theta_0^0 & \Theta_1^1 \\ \Theta_1^0 & \Theta_1^1 \end{bmatrix} \),

and the covariance between the random effects of the two regimes, \( \Theta_0^1 \), 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 \( y_i^* = \begin{pmatrix} y_{i1}^* \cdots y_{ki}^* \end{pmatrix} \), where \( y_{ji}^* = \begin{pmatrix} y_{j1i}^* \cdots y_{jki}^* \end{pmatrix} \), \( j = 1,\ldots,k \). Although the Gibbs sampling is entirely analogous to the cross-sectional case, it must be taken into account that now \( \text{var}(y_i^*) = \Lambda_i = \text{var}(\varphi_i + \epsilon_i) \) is a square matrix of dimension \( kT_i \) involving elements from \( \Theta \) and \( \theta \). Thus, since the observations for individual \( i \) are connected through the individual random effects, the
simulation for $y^*_j$ must be made conditional on observed information and $y^*_{i,j-1}$. The first conditional maximization produces:

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

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

$$Q_i(\beta^{(m+1)} | \beta^{(m)}, \Theta^{(m)}, \theta^{(m)}, y) = E \left[ (\varphi_i + \epsilon_i)(\varphi_i + \epsilon_i) | \beta^{(m)}, \Theta^{(m)}, \theta^{(m)}, y_i \right] ,$$

the second conditional maximization optimizes

$$\ell^* \left( \Theta, \theta | \beta^{(m+1)}, \Theta^{(m)}, \theta^{(m)}, y \right) = -k \ln(2\pi) \sum_{i} \sum_{t=1}^{T_i} -\frac{1}{2} \ln |\Lambda_i| - \frac{1}{2} \text{tr} \left( \sum_{i} \Lambda_i^{-1} Q_i(\beta^{(m+1)} | \beta^{(m)}, \Theta^{(m)}, \theta^{(m)}, y) \right)$$

with respect to $\Theta$ and $\theta$. Since individuals can move from one regime to another as time goes, $\Lambda_i$ 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 $\Lambda_i$ 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

<table>
<thead>
<tr>
<th>Equation</th>
<th>Variable</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost-Sharing</strong></td>
<td>Age</td>
<td>-0.0035</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>College</td>
<td>0.0040</td>
<td>0.0615</td>
</tr>
<tr>
<td></td>
<td>Highly</td>
<td>-0.0337</td>
<td>0.1336</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>0.0772</td>
<td>0.0703</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td>0.8247</td>
<td>0.1698</td>
</tr>
<tr>
<td></td>
<td>Waterbody</td>
<td>0.0717</td>
<td>0.0331</td>
</tr>
<tr>
<td><strong>Vegetative Cover</strong></td>
<td>Age</td>
<td>0.0013</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>College</td>
<td>0.0688</td>
<td>0.0316</td>
</tr>
<tr>
<td></td>
<td>Highly</td>
<td>0.0606</td>
<td>0.0821</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>0.2073</td>
<td>0.0382</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td>-0.4961</td>
<td>0.7372</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>-0.0324</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>Time²</td>
<td>0.0025</td>
<td>0.0011</td>
</tr>
<tr>
<td><strong>Contour cropping</strong></td>
<td>Age</td>
<td>-0.0035</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>College</td>
<td>-0.0337</td>
<td>0.0259</td>
</tr>
<tr>
<td></td>
<td>Highly</td>
<td>0.0072</td>
<td>0.0702</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>0.1938</td>
<td>0.0335</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td>-1.2992</td>
<td>0.8775</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>-0.0575</td>
<td>0.0310</td>
</tr>
<tr>
<td></td>
<td>Time²</td>
<td>0.0044</td>
<td>0.0025</td>
</tr>
<tr>
<td><strong>Cover crops</strong></td>
<td>Age</td>
<td>-0.0017</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>College</td>
<td>0.0308</td>
<td>0.0254</td>
</tr>
<tr>
<td></td>
<td>Highly</td>
<td>-0.1115</td>
<td>0.0725</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>-0.0281</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td>Land</td>
<td>-0.1285</td>
<td>0.7534</td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td>-0.0470</td>
<td>0.0294</td>
</tr>
<tr>
<td></td>
<td>Time²</td>
<td>0.0033</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Covariance matrix</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{12})</td>
<td>0.0994</td>
<td>0.0691</td>
</tr>
<tr>
<td>(\sigma_{13})</td>
<td>-0.2321</td>
<td>0.2201</td>
</tr>
<tr>
<td>(\sigma_{14})</td>
<td>-0.2427</td>
<td>0.2526</td>
</tr>
<tr>
<td>(\sigma_{22})</td>
<td>0.0694</td>
<td>0.0219</td>
</tr>
<tr>
<td>(\sigma_{23})</td>
<td>-0.0382</td>
<td>0.0180</td>
</tr>
<tr>
<td>(\sigma_{33})</td>
<td>0.013</td>
<td>0.0083</td>
</tr>
<tr>
<td>(\sigma_{34})</td>
<td>0.2675</td>
<td>0.0995</td>
</tr>
<tr>
<td>(\sigma_{44})</td>
<td>0.0830</td>
<td>0.0384</td>
</tr>
<tr>
<td>(\sigma_{55})</td>
<td>0.1253</td>
<td>0.0573</td>
</tr>
<tr>
<td>(\delta_{0})</td>
<td>1.1219</td>
<td>0.4082</td>
</tr>
<tr>
<td>(\delta_{1})</td>
<td>-0.2127</td>
<td>0.1297</td>
</tr>
<tr>
<td>(\delta_{2})</td>
<td>-0.8532</td>
<td>0.4082</td>
</tr>
<tr>
<td>(\delta_{3})</td>
<td>-0.2132</td>
<td>0.1901</td>
</tr>
<tr>
<td>(\delta_{4})</td>
<td>-0.3151</td>
<td>0.1638</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Significance</th>
<th>Estimate</th>
<th>Std. error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^a) Significant at 1% significance; (^b) significant at 5% significance; (^c) significant at 10% significance.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 1. HISTOGRAMS FOR ESTIMATORS FROM SIMULATED DATA
References


Novedades

**DIVISIÓN DE ADMINISTRACIÓN PÚBLICA**

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

**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
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 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 Mill*, 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


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


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


DIVISIÓN DE HISTORIA


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

**Ventas**


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

<table>
<thead>
<tr>
<th>VENTAS DIRECTAS:</th>
<th>VENTAS EN LÍNEA:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tel. Directo: 5081-4003</td>
<td>Librería virtual: <a href="http://www.e-cide.com">www.e-cide.com</a></td>
</tr>
<tr>
<td>Tel: 5727-9800 Ext. 6094 y 6091</td>
<td>Dudas y comentarios: <a href="mailto:publicaciones@cide.edu">publicaciones@cide.edu</a></td>
</tr>
<tr>
<td>Fax: 5727 9800 Ext. 6314</td>
<td></td>
</tr>
<tr>
<td>Av. Constituyentes 1046, 1er piso, Col. Lomas Altas, Del. Álvaro Obregón, 11950, México, D.F.</td>
<td></td>
</tr>
</tbody>
</table>

¡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.