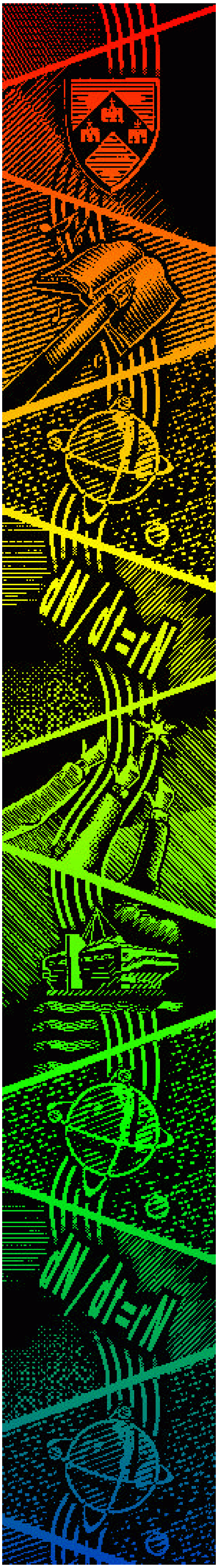

THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2001/04

A Panel Data Simultaneous Equation Model with a Dependent Categorical Variable and Selectivity

by

Roberto Leon Gonzalez

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

A Panel Data Simultaneous Equation Model with a Dependent Categorical Variable and Selectivity.

Roberto Leon Gonzalez
Department of Economics & Center for Health Economics
University of York, YO10 5DD, UK

This paper develops a Bayesian MCMC algorithm to estimate a Panel Data Simultaneous Equations model with a dependent categorical variable and selectivity. In contrast with previous Bayesian analysis of selectivity models, the algorithm does not require the observation of some regressors which do not enter into the likelihood function. This makes the algorithm applicable to studies of the labor market where there are typically missing regressors. In addition, the paper provides an scheme to sample the slope parameters using an analytical approximation of the posterior distribution as a proposal density. Estimation with a simulated dataset illustrates the performance of the algorithm.

January, 2001

Key Words: Bayesian, Markov Chain Monte Carlo, Inverted Wishart.

1 Introduction.

Applications of microeconomic theory to unit record data frequently encounter the joint problems of sample selectivity, categorical variables and simultaneity. Simultaneous equation models with discrete endogenous variables are an important area of research in econometrics, and are reviewed and extended in Blundell and Smith (1993). The model considered in this article is a Bayesian MCMC extension to longitudinal data of one of the models analysed in Blundell and Smith (1993). A random effects specification is chosen to model the longitudinal nature of the data.

Define h_{it} as a categorical variable taking values in the set $\{0, 1, \ldots, J\}$. Assume that the observed value of h_{it} is determined by a continuous unobserved variable h^*_{it} such that $h_{it} = j$ if and only if $h^*_{it} \in (l_j, l_{j+1}]$, with $l_0 = -\infty$ and $l_{J+1} = \infty$. Let w_{it} be a continuous variable which is jointly determined with h^*_{it}. The simultaneous realization of w_{it} and h_{it} and the non-representativeness of the sample can be modelled as follows.

$$
\begin{aligned}
p^*_{it} &= X^T_{1it}\beta_1 + u_{1i} + e_{1it} & i &= 1, \ldots, N \\
w_{it} &= X^T_{2it}\beta_2 + h^*_{it}\delta + u_{2i} + e_{2it} & t &= 1, \ldots, T \\
h^*_{it} &= X^T_{3it}\beta_3 + \gamma w_{it} + u_{3i} + e_{3it} & & \qquad (1)
\end{aligned}
$$

$$
\begin{pmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{pmatrix} \sim N(0; \Sigma) \qquad \Sigma = \begin{pmatrix} 1 & \sigma_{e12} & \sigma_{e13} \\ \sigma_{e12} & \sigma_{e22} & \sigma_{e23} \\ \sigma_{e13} & \sigma_{e23} & 1 \end{pmatrix}
$$

$$
\begin{pmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \end{pmatrix} \sim N(0; D) \qquad D = \begin{pmatrix} \sigma_{u11} & \sigma_{u12} & \sigma_{u13} \\ \sigma_{u12} & \sigma_{u22} & \sigma_{u23} \\ \sigma_{u13} & \sigma_{u23} & \sigma_{u33} \end{pmatrix}
$$

X_{1it}, X_{2it} and X_{3it} are three vectors of explanatory variables of dimensions $(l \times 1)$, $(k \times 1)$ and $(f \times 1)$, respectively. β_1, β_2, and β_3 are three comformably dimensioned parameter vectors. Selectivity effects are modelled by assuming that w_{it} is only observed when the unobserved latent variable p^*_{it} is positive. Since only the sign of p^*_{it} or the category in which h^*_{it} falls is observed, the model is not likelihood identified. As a normalization, σ_{e11} and σ_{e33} are restricted to be one and l_1 restricted to be zero. In addition, the usual conditions for identification in linear simultaneous equation models apply (see Judge et al, 1985, pp. 573-586).

Define $\beta = \left(\beta^T_1, \beta^T_2, \delta, \beta^T_3, \gamma\right)^T$, $Z_{it} = (p^*_{it}, w_{it}, h^*_{it})$ and,

$$
X_{it} = \begin{pmatrix} X_{1it} & 0 & 0 \\ 0 & X_{2it} & 0 \\ 0 & h^*_{it} & 0 \\ 0 & 0 & X_{3it} \\ 0 & 0 & w_{it} \end{pmatrix}
$$

equations in (1) can be expressed as,

$$
Z_{it} = X^T_{it}\beta + U_i + E_{it}
$$

Three main features differentiate the algorithm presented here from other Bayesian MCMC analysis proposed in the literature. Firstly, in contrast with previous analysis of models with selectivity (e.g.

Chib and Hamilton 2000) w is not imputed when it is not observed. This allows the algorithm to be applicable to many situations in which not only w but also some regressors in X_{2it} may not be observed. This situation is common in studies of the labor market in which wages (w) depend on regressors such as size of the company, type of industry and experience in that type of job. In such a situation, wages and some regressors in X_{2it} are not observed for people who stay out of the labor market. In this context, h_{it} could represent self-assesed health, which is usually recorded as a categorical variable and potentially affects and is affected by wages (see for instance Haveman et al. 1994).

Secondly, all the free elements in Σ are sampled directly from their conditional distributions given $\sigma_{\varepsilon 13}$. In addition, the acceptance probability of the Metropolis step which generates Σ only depends on $\sigma_{\varepsilon 13}$. This is made possible by analysing the marginal and conditional distributions of the elements in Σ.

Thirdly, it uses an analytical approximation of the conditional posterior of the slope parameters (β) as a proposal density in a Metropolis-within-Gibbs step. Using this proposal density, the acceptance probability only depends on (γ, δ). Hence, $(\beta_1, \beta_2, \beta_3)$ do not have to be sampled when the proposed value is rejected.

Section 2 describes the algorithm to simulate from the posterior distribution. Section 3 illustrates the performance of the proposed method with simulated data. Section 4 concludes.

2 Sampling from the Posterior Distribution.

A Gibbs sampler algorithm (Gelfand and Smith 1990) with data augmentation (Tanner and Wong 1987) can be followed in order to sample from the posterior distribution. The algorithm proposed in this paper blocks the parameters into six groups so that all the elements in one group are jointly generated conditional on the rest of the groups. Let $\Upsilon_i = \{t : p^*_{it} > 0\}$. The six groups are, $G_1 = \Sigma$, $G_2 = \beta$; $G_3 = \{U_i\}_{i=1}^N$, $G_4 = \left\{\{h^*_{it} : t \in \Upsilon_i\}_{i=1}^N ; \{\mathrm{fl}_j\}_{j=2}^J\right\}$, $G_5 = \{p^*_{it} : t = 1, \ldots, T\}_{i=1}^N$, $G_6 = D$,.

Latent data and individual effects are regarded as parameters following the ideas of data augmentation in Tanner and Wong (1987). Following Cowles (1996), the latent data h^*_{it} and the cutpoints $\{\mathrm{fl}_j\}_{j=2}^J$ are grouped into the same block. This substantially increases the speed of convergence in large datasets.

Non informative priors are chosen for the parameters β, D, $\{\mathrm{fl}_j\}_{j=2}^J$ and Σ,

$$\pi\left(\beta, D, \{\mathrm{fl}_j\}_{j=2}^J, \Sigma\right) = \pi(\beta)\,\pi(D)\,\pi\left(\{\mathrm{fl}_j\}_{j=2}^J\right)\pi(\Sigma) \propto |\Sigma|^{-4/2}|D|^{-4/2}$$

Section 3 estimates the model with simulated data, showing that the algorithm converged to the true value of the parameters. Therefore, it does not seem necessary to specify proper priors to ensure the convergence of the algorithm. By the other hand, the algorithm could easily accomodate the specification of informative conjugate priors.

To simplify the exposition the following notation will be used below,

$$c_i = \sum_{t \in \Upsilon_i} 1 \qquad \Gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\delta \\ 0 & -\gamma & 1 \end{pmatrix}$$

2.1 Sampling Σ

This section describes a Metropolis-Hasting step to generate Σ. The acceptance probability can be seen to depend only on $\sigma_{\varepsilon 13}$. Hence, the rest of parameters in Σ are not generated when the proposed value is rejected.

The conditional posterior for Σ is an inverted Wishart with the restriction that both $\sigma_{\varepsilon 11}$ and $\sigma_{\varepsilon 33}$ are equal to one. The parameters of this inverted Wishart are $df = \sum_{i=1}^N (c_i) + 4$ and $K = \sum_{i=1}^N \sum_{t \in \Upsilon_i} E_{it} E_{it}^T$.

Cowles et al. (1996) outline the algorithm to sample from an IW with the restriction of one element in the diagonal being equal to one. They implicitly make use of the following theorem, which can be found in Bauwens et al. (1999, pages 305-306).

Theorem 1 Let Σ be distributed as an $IW(d; n; G)$, and be partitioned as $\Sigma = (\Sigma_{ij})$; $i,j=1,2$, Σ_{11} being a scalar. Define $\Sigma_{22\cdot 1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, then

$$1) \quad \Sigma_{22\cdot 1}|\Sigma_{11} \sim IW(d-1; n-1; G_{22\cdot 1})$$
$$2) \quad \Sigma_{12}|(\Sigma_{22\cdot 1}; \Sigma_{11}) \sim N\left(\frac{\Sigma_{11}}{G_{11}}G_{12}; \frac{(\Sigma_{11})^2}{G_{11}}\Sigma_{22\cdot 1}\right)$$

Define $\Sigma_{22\cdot 1}$ as

$$\begin{pmatrix} \sigma^{11}_{22\cdot 1} & \sigma^{12}_{22\cdot 1} \\ \sigma^{12}_{22\cdot 1} & \sigma^{22}_{22\cdot 1} \end{pmatrix} = \begin{pmatrix} 1 & \sigma_{e12} \\ \sigma_{e12} & \sigma_{e22} \end{pmatrix} - \begin{pmatrix} \sigma_{e13} \\ \sigma_{e23} \end{pmatrix} \begin{pmatrix} \sigma_{e13} & \sigma_{e23} \end{pmatrix}$$

The following proposition gives the marginal distribution of σ_{e13}.

Proposition 2 The marginal distribution of σ_{e13} conditional on $(\sigma_{e11} = \sigma_{e33} = 1)$, is proportional to

$$g(\sigma_{e13}) = \frac{1}{(1-(\sigma_{e13})^2)^{(df-3)/2}} \exp\left\{-\frac{1}{2}\frac{1}{1-(\sigma_{e13})^2}k^{11}_{22\cdot 1}\right\}$$
$$\times \frac{1}{((1-(\sigma_{e13})^2)/k_{33})^{1/2}} \exp\left\{-\frac{1}{2}\frac{(\sigma_{e13}-k_{13}/k_{33})^2}{(1-(\sigma_{e13})^2)/k_{33}}\right\} I_{-1<\sigma_{e13}<1}; \qquad (2)$$

Proof. Theorem 1 gives the joint conditional distribution of $\Sigma_{22\cdot 1}$ and $(\sigma_{e13}; \sigma_{e23})^T$ given $\sigma_{e33} = 1$, which is the product of the marginal of $\Sigma_{22\cdot 1}$ times the conditional of $(\sigma_{e13}; \sigma_{e23})^T$ given $\Sigma_{22\cdot 1}$. The restriction $\sigma_{e11} = 1$ implies only one restriction on the joint distribution of $\Sigma_{22\cdot 1}$ and $(\sigma_{e13}; \sigma_{e23})^T$. This restriction is that $\sigma^{11}_{22\cdot 1} + (\sigma_{e13})^2 = 1$. Thus, integrating out σ_{e23}; $\sigma^{12}_{22\cdot 1}$ and $\sigma^{22}_{22\cdot 1}$ we obtain that the unrestricted marginal distribution of $(\sigma^{11}_{22\cdot 1}; \sigma_{e13})$ is proportional to,

$$\frac{1}{(\sigma^{11}_{22\cdot 1})^{(df-3)/2}} \exp\left\{-\frac{1}{2}\frac{1}{\sigma^{11}_{22\cdot 1}}k^{11}_{22\cdot 1}\right\} \frac{1}{(\sigma^{11}_{22\cdot 1}/k_{33})^{1/2}}$$
$$\times \exp\left\{-\frac{1}{2}\frac{(\sigma_{e13}-k_{13}/k_{33})^2}{\sigma^{11}_{22\cdot 1}/k_{33}}\right\} I_{\sigma^{11}_{22\cdot 1}>0}$$

The derivation of this marginal distribution has used the property according to which the submatrices centered along the diagonal of an inverted Wishart matrix also follow an inverted Wishart distribution (Press 1982, p. 118). The restriction $\sigma^{11}_{22\cdot 1} + (\sigma_{e13})^2 = 1$ can be imposed by calculating the joint density function of $(\sigma^{11}_{22\cdot 1} + (\sigma_{e13})^2; \sigma_{e13})$. Since the Jacobian of the transformation is 1, the desired density is proportional to expression (2). ■

A Metropolis step could be employed to generate σ_{e13}, using a normal proposal density centered in the previous value of σ_{e13} in the chain. σ_{e13} could also be generated using a proposal density that approximates $g(\sigma_{e13})$. Since the unrestricted marginal distribution of σ_{e13} is a student-t, one possible approximation is a student-t centered on k_{13}/k_{33} with $(df-2)$ degrees of freedom and truncated to $(-1; 1)$.

The distribution of $(\sigma_{e12}; \sigma_{e22}; \sigma_{e23})$ conditional on σ_{e13} can be sampled directly. Once σ_{e13} has been generated, $\Sigma_{22\cdot 1}$ should be sampled from an $IW(2; df-1; K_{22\cdot 1})$ with the restriction $\sigma^{11}_{22\cdot 1} = 1 - (\sigma_{e13})^2$. By theorem 1, this distribution can be sampled by sampling $\omega = \sigma^{22}_{22\cdot 1} - (\sigma^{12}_{22\cdot 1})^2/\sigma^{11}_{22\cdot 1}$ from a $IW\left(1; df-2; k^{22}_{22\cdot 1} - (k^{12}_{22\cdot 1})^2/k^{11}_{22\cdot 1}\right)$ and $\sigma^{12}_{22\cdot 1} = \sigma_{e12} - \sigma_{e23}\sigma_{e13}$ from a $N\left(\sigma^{11}_{22\cdot 1}k^{12}_{22\cdot 1}/k^{11}_{22\cdot 1}; (\sigma^{11}_{22\cdot 1})^2\omega/k^{11}_{22\cdot 1}\right)$. The sampled value for $\sigma^{22}_{22\cdot 1}$ is $\omega + (\sigma^{12}_{22\cdot 1})^2/\sigma^{11}_{22\cdot 1}$. Finally, σ_{e23} follows a

$$N\left(\frac{k_{23}}{k_{33}} + \frac{(\sigma^{12}_{22\cdot 1}/k_{33})}{(\sigma^{11}_{22\cdot 1}/k_{33})}\left(\sigma_{e13} - \frac{k_{13}}{k_{33}}\right); \frac{\sigma^{22}_{22\cdot 1}}{k_{33}} - \frac{(\sigma^{12}_{22\cdot 1}/k_{33})^2}{(\sigma^{11}_{22\cdot 1}/k_{33})}\right) \qquad (3)$$

Once σ_{e13} and σ_{e23} have been sampled, σ_{e12} and σ_{e22} will be fixed to $\sigma_{e12} = \sigma^{12}_{22\cdot 1} + \sigma_{e13}\sigma_{e23}$ and $\sigma_{e22} = \sigma^{22}_{22\cdot 1} + (\sigma_{e23})^2$.

Let σ^{k}_{e13} be the kth value of σ_{e13} in the chain. The following algorithm summarizes the procedure to sample the (k+1) value of Σ in the chain.

Algorithm 3 Step 1. Sample a candidate v from a Normal distribution centered on σ^{k}_{e13} and with variance s_1.

Step 2. If $|v| < 1$ go to step 3. Otherwise fix $\Sigma^{k+1} = \Sigma^{k}$.

Step 3. Accept v as a value for σ^{k+1}_{e13} with probability

$$\min \left\{ \frac{g\left(\sigma^{k+1}_{e13}\right)}{g\left(\sigma^{k}_{e13}\right)}; 1 \right\}$$

If v is accepted fix $\sigma^{11}_{22\cdot 1} = 1 - \left(\sigma^{k+1}_{e13}\right)^2$ and go to step 4. Otherwise fix $\Sigma^{k+1} = \Sigma^{k}$.

Step 4. Sample η from a

$$IW\left(1; df - 2; k^{22}_{22\cdot 1} - \left(k^{12}_{22\cdot 1}\right)^2 / k^{11}_{22\cdot 1}\right)$$

Step 5. Generate $\sigma^{12}_{22\cdot 1}$ from a

$$N\left(\sigma^{11}_{22\cdot 1} k^{12}_{22\cdot 1} / k^{11}_{22\cdot 1}; \left(\sigma^{11}_{22\cdot 1}\right)^2 \eta / k^{11}_{22\cdot 1}\right)$$

and fix $\sigma^{22}_{22\cdot 1}$ equal to $\eta + \left(\sigma^{12}_{22\cdot 1}\right)^2 / \sigma^{11}_{22\cdot 1}$.

Step 6. Sample σ_{e23} from distribution (3).

Step 7. Fix σ_{e12} equal to $\sigma^{12}_{22\cdot 1} + \sigma^{k+1}_{e13}\sigma_{e23}$ and σ_{e22} equal to $\sigma^{22}_{22\cdot 1} + (\sigma_{e23})^2$.

2.2 Sampling β

The fact that w_{it} cannot be imputed whenever it is not observed, together with the simultaneous realization of w_{it} and h^{*}_{it}, means that the posterior distribution for β is different from the common Normal distribution obtained in similar models (e.g. Chib and Hamilton 2000).

Proposition 4 The conditional posterior for β up to a constant of proportionality is

$$(|\Delta|)^{\sum_{i=1}^{N} c_i} \exp\left\{ -\frac{1}{2}(\beta_1 - m)^T (\Psi_1)^{-1} (\beta_1 - m) \right\} \exp\left\{ -\frac{1}{2}(\beta - \beta^1)^T (\Psi)^{-1} (\beta - \beta^1) \right\} \quad (4)$$

where $m = \left(\sum_{i=1}^{N} \sum_{t \notin \Upsilon_i} X_{1it}(X_{1it})^T\right)^{-1} \sum_{i=1}^{N} \sum_{t \notin \Upsilon_i} X_{1it}(p^{*}_{it} - u_{1i})$, $\Psi_1 = \left(\sum_{i=1}^{N} \sum_{t \notin \Upsilon_i} X_{1it}(X_{1it})^T\right)^{-1}$, $\Psi = \left(\sum_{i=1}^{N} \sum_{t \in \Upsilon_i} \left(X_{it}\Sigma^{-1}X_{it}^T\right)\right)^{-1}$, $\beta^1 = \Psi\left(\sum_{i=1}^{N} \sum_{t \in \Upsilon_i} \left(X_{it}\Sigma^{-1}(Z_{it} - U_i)\right)\right)$.

Proof. The likelihood of β given $G_1; G_3; G_4; G_5$ and G_6 is proportional to

$$(|\Delta|)^{\sum_{i=1}^{N} c_i} \exp\left\{ -\frac{1}{2} \sum_{i=1} \sum_{t \in \Upsilon_i} \left(Z_{it} - (X_{it}^T\beta + U_i)\right)^T \Sigma^{-1} \left(Z_{it} - (X_{it}^T\beta + U_i)\right) \right\}$$

$$\times \exp\left\{ -\frac{1}{2} \sum_{i=1} \sum_{t \notin \Upsilon_i} \left(p^{*}_{it} - \left(X_{1it}^T\beta_1 + u_{1i}\right)\right)^2 \right\}$$

This expression can be shown to be proportional to density (4) using standard algebraic transformations. ■

The effect of simultaneity on the posterior distribution is captured in the following proposition.

Proposition 5 Let μ_e and Ψ_e be the elements corresponding to (δ, γ) in μ and Ψ, respectively. The conditional posterior of $e = (\delta, \gamma)^T$ marginally on the rest of elements in β is

$$\pi(\delta, \gamma) \propto (|\Delta|)^{\sum_{i=1}^{N} c_i} \exp\left\{ -\frac{1}{2}(e - \mu_e)^T (\Psi_e)^{-1}(e - \mu_e)\right\} \tag{5}$$

The conditional distribution of $(\beta_1, \beta_2, \beta_3)$ given (δ, γ) is not normal due to the non-observability of some w_{it}. However, as the following proposition shows, both the marginals and conditionals of this distribution are normal, making it possible to sample from it.

Proposition 6 Define $\Psi_{con} = \Psi_{\beta\beta} - \Psi_{\beta e}(\Psi_{ee})^{-1}\Psi_{e\beta}$ where $\Psi_{\beta\beta}$ is the submatrix of Ψ which corresponds to the variance-covariance matrix of $(\beta_1, \beta_2, \beta_3)$, and $\Psi_{\beta e}$ is the submatrix of Ψ with the covariances between $(\beta_1, \beta_2, \beta_3)$ and (δ, γ). Let μ_1 and μ_2 be the elements in μ corresponding to β_1 and (β_2, β_3), respectively. Let Ψ_{con} be partioned as $\Psi_{con} = \left[\Psi_{con}^{ij}\right]$, i,j=1,2, with Ψ_{con}^{11} being a squared matrix with the same number of rows as β_1. Then, the distribution of β_1 conditional on (δ, γ) is a normal distribution with mean $\left((\Psi_1)^{-1} + \left[\Psi_{con}^{11}\right]^{-1}\right)^{-1}\left((\Psi_1)^{-1} m + \left[\Psi_{con}^{11}\right]^{-1}\mu_1\right)$ and variance-covariance matrix $\left((\Psi_1)^{-1} + \left[\Psi_{con}^{11}\right]^{-1}\right)^{-1}$. The conditional distribution of (β_2, β_3) given $(\beta_1, \delta, \gamma)$ is normal with mean $\mu_2 + \Psi_{con}^{21}\left[\Psi_{con}^{11}\right]^{-1}(\beta_1 - \mu_1)$ and variance-covariance matrix $\Psi_{con}^{22} - \Psi_{con}^{21}\left[\Psi_{con}^{11}\right]^{-1}\Psi_{con}^{12}$.

Parameters (δ, γ) can be generated using a Metropolis-within-Gibbs step. A student-t centered on a consistent estimator of the mode of distribution (5) can be used as a proposal density. The instrumental variables three stages least squares (see Judge et al. 1985, pp. 599-601) is adapted to the situation in which both U_i and Σ are known. Such approximation of the mode converges to the true value as the number of observations in the sample tends to infinity. The estimated variance-covariance matrix of this estimator can be used as the variance-covariance matrix of the proposal density. Alternatively, it can be set to be equal to the inverse of the negative of the Hessian of the logarithm of expression (5) evaluated at the approximated mode. The negative of this Hessian can be obtained analytically and is equal to

$$\Psi_e^{-1} + \frac{1}{(1 - \delta\gamma)^2}\begin{pmatrix} \gamma^2 & 1 \\ 1 & \delta^2 \end{pmatrix}\sum_{i=1} c_i$$

Let $\eta(\delta, \gamma)$ be the described proposal density and $\beta^k = \left(\beta_1^k, \beta_2^k, \delta^k, \beta_3^k, \gamma^k\right)$ the kth value of β in the chain. The (k+1)th value of β is generated as follows.

Algorithm 7 1) Generate a candidate $v = (v_1, v_2)$ from the proposal density $\eta(\delta, \gamma)$.

2) Accept candidate v as a new value for (δ, γ) with probability

$$\min\left\{1, \frac{\pi\left(\delta^{k+1}, \gamma^{k+1}\right)\eta\left(\delta^k, \gamma^k\right)}{\pi\left(\delta^k, \gamma^k\right)\eta\left(\delta^{k+1}, \gamma^{k+1}\right)}\right\}$$

If v is not accepted fix $\beta^{k+1} = \beta^k$. Otherwise fix $\left(\delta^{k+1}, \gamma^{k+1}\right) = v$ and go to step 3.

3) Sample β_1 conditional on $(\delta, \gamma) = v$ from the distribution described in proposition 7.

4) Sample (β_2, β_3) conditional on $(\beta_1, \delta, \gamma)$.

2.3 Sampling $\{U_i\}_{i=1}^N$

The conditional posterior distribution of U_i is of the same type as that of $(\beta_1, \beta_2, \beta_3)$. The joint distribution is still not normal even though the marginals and conditionals are normal.

Proposition 8 The distribution of U_i is proportional to

$$\exp\left\{-\frac{1}{2}(U_i - a_i)^T V_i^{-1}(U_i - a_i)\right\}\exp\left\{-\frac{1}{2}\frac{1}{v_i}(u_{1i} - b_i)^2\right\}$$

where $a_i = \left(c_i\Sigma^{-1} + D^{-1}\right)^{-1}\Sigma^{-1}\left(\sum_{t\in Y_i}\left(Z_{it} - X_{it}^T\beta\right)\right)$, $V_i = \left(c_i\Sigma^{-1} + D^{-1}\right)^{-1}$, $\frac{1}{v_i} = T - c_i$, $b_i = \sum_{t\in Y_i}\frac{(p^*_{it} - X^T_{1it}\beta_1)}{T - c_i}$ if $T - c_i \neq 0$, $b_i = 0$ if $T - c_i = 0$

Let $a_i = (a_{i1}, a_{i2}, a_{i3})^T = \left(a_{i1}, a_{ir}^T\right)^T$ and $V_i = \begin{pmatrix} v_{i11} & V_{i1r} \\ V_{i1r}^T & V_{irr}\end{pmatrix}$, where v_{i11} is a scalar.

The following algorithm shows how to sample from this distribution.

Algorithm 9 1) Sample u_{1i} from a normal distribution with mean $(b_i(T - c_i) + a_{i1}/v_{i11})/(T - c_i + 1/v_{i11})$ and variance $1/(T - c_i + 1/v_{i11})$.

2) Sample (u_{2i}, u_{3i}) given u_{1i} from a normal with mean $a_{ir} + V_{i1r}^T(u_{i1} - a_{i1})/v_{i11}$ and variance $V_{irr} - V_{i1r}^T V_{i1r}/v_{i11}$.

2.4 Sampling $\left\{\{h^*_{it} : t \in \tau_i\}_{i=1}^N; \{l_i\}_{i=2}^J\right\}$.

The algorithm in Cowles (1996) can be applied to sample G_4 just noting that the distribution of h^*_{it} is given by the reduced form of system (1).

Define $\Sigma^* = \Gamma^{-1}\Sigma\left(\Gamma^{-1}\right)^T$ and denote the elements in Σ^* as $\left(\sigma^*_{ij}\right)$, $i, j = 1, 2, 3$. Let $\Sigma_{pw} = \begin{pmatrix}\sigma^*_{e11} & \sigma^*_{e12} \\ \sigma^*_{e12} & \sigma^*_{e22}\end{pmatrix}$, and $\Sigma_{h;pw} = (\sigma^*_{e13}, \sigma^*_{e23})$. The conditional posterior distribution of h^*_{it} given $h_{it} = j, \{l_i\}_{i=1}^{J+1}$ is a normal truncated to $(l_j, l_{j+1}]$ with mean μ_{it}

$$|\Gamma|^{-1}\left((X_{3it}^T\beta_3 + u_{3i}) + \Sigma_{h;pw}(\Sigma_{pw})^{-1}\begin{pmatrix} p^*_{it} - X^T_{1it}\beta_1 - u_{1i} \\ w_{it} - |\Gamma|^{-1}\left((X^T_{2it}\beta_2 + u_{2i}) + \delta(X^T_{3it}\beta_3 + u_{3i})\right)\end{pmatrix}\right)$$

and variance $\sigma = 1 - \Sigma_{h;pw}(\Sigma_{pw})^{-1}(\Sigma_{h;pw})^T$.

Let h^{*k}_{it}, $\{l^k_i\}_{i=2}^J$ denote the kth value of these parameters in the chain. Define C^k_{it} as the interval $\left(l^k_j; l^k_{j+1}\right]$ with j being the value taken by h^*_{it}. The (k+1)th value is generated as follows.

Algorithm 10 1) Sample a candidate $v = (v_2, \ldots, v_J)$ from (J-1) independent normal distributions centered on $\{l^k_i\}_{i=2}^J$ and with variances $s_2, \ldots, s_J$.

2) If $0 < l_2 < \cdots < l_J$ then go to step 3. Otherwise, fix $\{h^{*k+1}_{it} : t \in \tau_i\}_{i=1}^N$, $\{l^{k+1}_i\}_{i=2}^J$ equal to $\{h^{*k}_{it} : t \in \tau_i\}_{i=1}^N$, $\{l^k_i\}_{i=2}^J$.

3) Accept v as a value for $\{l^{k+1}_i\}_{i=2}^J$ with probability

$$\min\left\{\prod_{i=1}\prod_{t\in\tau_i}\frac{\Pr\left(N(\mu_{it}; \sigma) \in C^{k+1}_{it}\right)}{\Pr\left(N(\mu_{it}; \sigma) \in C^k_{it}\right)}; 1\right\}$$

if v is accepted go to step 4. Otherwise, fix $\{h^{*k+1}_{it} : t \in \tau_i\}_{i=1}^N$, $\{l^{k+1}_i\}_{i=2}^J$ equal to $\{h^{*k}_{it} : t \in \tau_i\}_{i=1}^N$, $\{l^k_i\}_{i=2}^J$

4) Sample h^*_{it} from a normal with mean μ_{it} and variance σ truncated to C^{k+1}_{it}, for $i = 1, \ldots, N$ and $t \in \tau_i$.

The values for the variances $s_2, \ldots, s_J$ are chosen to obtain a reasonable acceptance rate. Note that the proposal density does not incorporate the restriction $0 < \gamma_2 < \ldots < \gamma_J$. This specification is preferred because the acceptance probability is simplified and new candidates never violated the restriction in the estimation presented in section 3.

2.5 Sampling $\{p^*_{it} : t = 1, \ldots, T\}_{i=1}^N$.

Let p_{it} be a binary variable which takes the value 1 when $p^*_{it} > 0$ and 0 otherwise.

If $p_{it} = 0$, p^*_{it} follows a normal distribution truncated to the interval $(-\infty; 0)$ and having mean and variance equal to $\left(X^T_{1it}\beta_1 + u_{1i}\right)$ and 1, respectively.

Let $\Sigma_{wh} = \begin{pmatrix} \sigma^*_{22} & \sigma^*_{23} \\ \sigma^*_{23} & \sigma^*_{33} \end{pmatrix}$, and $\Sigma_{p;wh} = (\sigma^*_{e12}; \sigma^*_{e13})$. Then, if $p_{it} = 1$ the conditional posterior distribution of p^*_{it} is a normal truncated to $(0; \infty)$ with mean

$$X^T_{1it}\beta_1 + u_{1i} + \Sigma_{p;wh}(\Sigma_{wh})^{-1}\begin{pmatrix} w_{it} - |\Delta|^{-1}((X^T_{2it}\beta_2 + u_{2i}) + \delta(X^T_{3it}\beta_3 + u_{3i})) \\ h^*_{it} - |\Delta|^{-1}((X^T_{3it}\beta_3 + u_{3i}) + \gamma(X^T_{2it}\beta_2 + u_{2i})) \end{pmatrix}$$

and variance $1 - \Sigma_{p;wh}(\Sigma_{wh})^{-1}(\Sigma_{p;wh})^T$.

2.6 Sampling D

The conditional posterior of D is an inverted wishart distribution.

$$D \sim IW\left(3; N + 4; \sum_{i=1}^{N} U_i U_i^T\right)$$

3 Estimation with a simulated dataset.

The model was estimated using a simulated dataset with 1200 individuals in 7 periods. This size resembles that of the widely used British Household Panel Survey. The exogenous regressors were simulated from a standard normal distribution. The estimations are based on 5000 iterations after having discarded 3000. The algorithm was implemented in the Gauss language. The following tables indicate that the algorithm converged to the true values.

4 Conclusions.

This paper has outlined a Bayesian MCMC algorithm to estimate a simultaneous equation model with a dependent categorical variable and selectivity. Due to simultaneity and selectivity, the conditional posteriors of β, and U_i do not belong to standard families of distributions. However, this paper shows how these parameters can be generated using standard distributions.

Acknowlegment: I thank P. Contoyannis, A. Jones and N. Rice for helpful comments and discussions. Of course, all errors and omissions are of my own responsability. I am also grateful to the Centre for Health Economics for financial support.

References:

Blundell, R.W., and Smith, R.J. (1993), "Simultaneous Microeconometric Models with Censored or Qualitative Dependent Variables," in Handbook of Statistics (Vol. 11), eds. G.S. Maddala, C.R. Rao and H.D. Vinod, Amsterdam: Elsevier, pp. 117-143.

Chib, S., and Hamilton, B. (2000), "Bayesian Analysis of Cross-Section and Clustered Data Selection Models," Journal of Econometrics, 97, 25-50.

Cowles, K. (1996), "Accelerating Monte Carlo Markov chain convergence for cumulative-link generalized linear models," Statistics and Computing 6, 101-111.

Cowles, K., Carlin, B. P., and Connett, J. E. (1996) "Bayesian Tobit Modellig of Longitudinal Ordinal Clinical Trial Compliance Data with Nonignorable Missingness," Journal of the American Statistical Association, 91, 86-98.

Gelfand, A. E., and Smith, A. F. M. (1990) "Sampling based approaches to calculating marginal densities," Journal of the American Statistical Association, 85, 398-409.

Haveman, R., Wolfe, B., Kreider, B., and Stone, M. (1994) "Market work, wages, and men's health," Journal of Health Economics, 13, 163-182.

Judge, G., Gri¢ths, W.E., and Carter Hill, R. (1985) The theory and practice of econometrics, Wiley.

Press, S. J. (1982) Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference, Krieger.

Tanner, T. A., and Wong, W. H. (1987) "The Calculation of Posterior Distributions by Data Augmentation," Journal of the American Statistical Association, 82, 528-550.

TABLES.

Table 1. True Values, Initial Values, Posterior Means and Credible Intervals for Slope Parameters.

	$!_1$	$!_2$	$!_3$	$!_4$	$!_5$	$!_6$	±	$!_7$	°	l_1	l_2
TV	1	2	0.5	-0.2	-1	0.8	0.8	-3	0.5	2	4
IV	0.29	0.58	0.14	-2.49	-1.45	0.4	1.8	-0.15	0.21	1.12	1.41
PM	1.04	1.99	0.48	-0.24	-0.98	0.78	0.79	-3.02	0.48	1.99	3.93
CI-	0.97	1.88	0.43	-0.32	-1.04	0.71	0.76	-3.2	0.44	1.87	3.71
CI+	1.11	2.08	0.52	-0.17	-0.92	0.85	0.83	-2.86	0.52	2.11	4.12

NOTE: $¯_1 = (!_1; !_2; !_3)$; $¯_2 = (!_4; !_5)$; $¯_3 = (w_6; w_7)$; TV indicates true value,IV initial value, PM posterior mean,and (CI-,CI+) is a 95% credibleinterval.

Table 2. True Values, Initial Values, Posterior Means and Credible Intervals for Variance and Covariance Parameters.

	$¾_{u11}$	$¾_{u12}$	$¾_{u13}$	$¾_{u22}$	$¾_{u23}$	$¾_{u33}$	$¾_{e12}$	$¾_{e13}$	$¾_{e22}$	$¾_{e23}$
TV	1	1	0.5	3	1	2	1	-0.3	2.5	0.2
IV	0.04	-0.014	0.07	3.33	-0.27	0.26	-0.48	0.02	14.46	-1.45
PM	1.03	0.987	0.42	3.05	1.15	2.17	1.027	-0.39	2.55	0.12
CI-	0.85	0.8	0.26	2.69	0.89	1.76	0.86	-0.52	2.35	-0.04
CI+	1.21	1.17	0.58	3.45	1.41	2.62	1.14	-0.25	2.77	0.28