ELSEVIER Journal of Econometrics 78 (1997) 217–227

JOURNAL OF Econometrics

Learning about the across-regime correlation in switching regression models

Gary Koop, Dale J. Poirier*

Department of Economics, University of Toronto, 150 St. George St., Toronto, Ont., M5S 3G7, Canada

Received February 1995; final version received March 1996

Abstract

Vijverberg (1993) considers the extended Roy/switching regression model of selectivity, focusing attention on the nonidentified correlation between the regime disturbances and describing how the positive definiteness of the covariance matrix implies that it is possible to learn about this covariance. In this paper, we show that this learning derives from prior dependence between identified and nonidentified parameters. Even though beliefs about the nonidentified covariance are updated, we show under reasonable a priori independence assumptions that beliefs about the partial correlation between disturbances, control of the switching index, are not updated. Empirical illustrations show how an exact Bayesian analysis can be carried out using Gibbs sampling and related techniques.

Key words: Bayesian; Gibbs; identification; Roy model
JEL classification: C11

1. Bayesian analysis of the Roy model

The extended Roy/switching regression model involves two continuous dependent variables determined by fixed explanatory variables in two sectors (regimes). Individuals choose between the sectors. References to numerous applications are given in Poirier and Ruud (1981) and in Vijverberg (1993). Omitting observation subscripts, the basic model is

$$I^* = Z\gamma + u, \qquad Y_1^* = X_1\beta_1 + \varepsilon_1, \qquad Y_2^* = X_2\beta_2 + \varepsilon_2, \tag{1}$$

* Correspondence address: 2746 Gough Street, Unit #3, San Francisco, CA 94123–4405, USA
Both authors acknowledge the generous research support of the Social Science Research Council of Canada.

PII S0304-4076(96)00009-7

where I^* measures, say, the difference between the utility experienced by a worker in sector 1 versus sector 2, Y_i^* is the wage offer (productivity) in sector i, Y_i is the wage received in sector i, and the observability conditions are

$$Y_1 = \begin{cases} Y_1^* & \text{iff } I^* \geqslant 0, \\ 0 & \text{otherwise} \end{cases} \qquad Y_2 = \begin{cases} Y_2^* & \text{iff } I^* < 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The vectors X_i and β_i are $K_i \times 1$ $(i = 1, 2)$. The random disturbance $\varepsilon = [u, \varepsilon_1, \varepsilon_2]' \sim \mathrm{N}_3(0_3, \Omega)$, where the covariance matrix

$$\Omega = \begin{bmatrix} 1 & \rho_{1u}\sigma_1 & \rho_{2u}\sigma_2 \\ \rho_{1u}\sigma_1 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{2u}\sigma_2 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{3}$$

is positive definite. The 1 in the upper left corner of Ω is the standard normalization. Given $-1 \leqslant \rho_{iu} \leqslant 1$ $(i = 1, 2)$, the positive definiteness of Ω implies $\underline{\rho}_{12} \leqslant \rho_{12} \leqslant \bar{\rho}_{12}$, where $\underline{\rho}_{12} \equiv \rho_{1u}\rho_{2u} - [(1 - \rho_{1u}^2)(1 - \rho_{2u}^2)]^{1/2}$ and $\bar{\rho}_{12} \equiv \rho_{1u}\rho_{2u} + [(1 - \rho_{1u}^2)(1 - \rho_u^2)]^{1/2}$.

Like Vijverberg (1993), let us focus attention on the parameter ρ_{12}, and treat the remaining parameters as nuisance parameters. Thus, the problem under consideration involves both unobserved variables through (2) and nuisance parameters. In such cases careful discussion of the likelihood principle is warranted (see Berger and Wolpert, 1988, especially Chap. 3). Since the likelihood function does *not* depend explicitly on ρ_{12}, ρ_{12} is *not identified*. However, apparently because the support for the parameter space implies $\underline{\rho}_{12} \leqslant \rho_{12} \leqslant \bar{\rho}_{12}$, where the bounds can be calculated from identified parameters appearing in the likelihood function, Vijverberg (1993, p. 75) pronounces that the *likelihood principle is violated*. This strikes us as unsubstantiated. The fact that a different trivariate distribution than the trivariate normal assumed here could conceivably produce the same observed likelihood is also irrelevant to the point.

Rather than engage in a semantical debate over the likelihood principle for this nonstandard problem, we instead move to the more important part of Vijverberg's study, namely its general implications for Bayesian analyses. As Vijverberg (1993, p. 75) observes, the prior and posterior distributions for ρ_{12} need not be identical. This fact, however, does not imply that the likelihood principle is violated. Instead what it demonstrates is a widely appreciated fact: in regular problems, unless an nonidentified parameter like ρ_{12} is a priori independent of the identified parameters, then the prior and posterior distributions for ρ_{12} need not be identical. The explanation is intuitively simple: data information on the identified parameters merely 'spills-over' to the nonidentified parameter if the two groups are a priori dependent. This point is worth expanding upon.

Let $K = K_1 + K_2 + 2$ and $\Delta \equiv \Re^{k-2} \times \Re^+ \times \Re^+$. Partition the $(K+3) \times 1$ vector θ as $\theta = [\delta', \alpha', \rho_{12}]'$, with $\delta = [\beta'_1, \beta'_2, \sigma_1^2, \sigma_2^2]' \in \Delta$ and $\alpha = [\rho_{1u}, \rho_{2u}]' \in A(\rho_{12})$, where

$$A(\rho_{12}) \equiv \left\{ \begin{bmatrix} \rho_{1u} \\ \rho_{2u} \end{bmatrix} \mid [\rho_{1u}, \rho_{2u}] \begin{bmatrix} 1 & -\rho_{12} \\ -\rho_{12} & 1 \end{bmatrix} \begin{bmatrix} \rho_{1u} \\ \rho_{2u} \end{bmatrix} < 1 - \rho_{12}^2 \right\} \tag{4}$$

is the support for α. Note that $[\rho_{1u}, \rho_{2u}]' \in A(\rho_{12})$ iff $\underline{\rho}_{12} < \rho_{12} < \bar{\rho}_{12}$. Consider the prior density $f(\delta, \alpha, \rho_{12}) = f(\delta | \alpha, \rho_{12}) f(\alpha | \rho_{12}) f(\rho_{12})$. Given data on the sign of the index I^* and on the observables indicated by (2), the marginal posterior density of ρ_{12} is

$$f(\rho_{12} | \text{Data}) = \frac{f(\rho_{12}) f(\text{Data} | \rho_{12})}{f(\text{Data})}, \qquad -1 < \rho_{12} < 1, \tag{5}$$

where the likelihood $L(\delta, \alpha; \text{Data})$ does *not* depend on ρ_{12}, and

$$f(\text{Data} | \alpha, \rho_{12}) = \int_{\Delta} f(\delta | \alpha, \rho_{12})\, L(\delta, \alpha; \text{Data})\, \mathrm{d}\delta, \tag{6}$$

$$f(\text{Data} | \rho_{12}) = \int_{A(\rho_{12})} f(\alpha | \rho_{12}) f(\text{Data} | \alpha, \rho_{12})\, \mathrm{d}\alpha, \tag{7}$$

$$f(\text{Data}) = \int_{-1}^{1} f(\rho_{12}) f(\text{Data} | \rho_{12})\, \mathrm{d}\rho_{12}. \tag{8}$$

Provided $f(\text{Data} | \rho_{12}) \neq f(\text{Data})$, (5) implies that prior beliefs on ρ_{12} are updated. The vehicles for this learning are the prior densities $f(\delta | \alpha, \rho_{12})$ and $f(\alpha | \rho_{12})$, and the support $A(\rho_{12})$. Note, however, that $f(\rho_{12} | \delta, \alpha, \text{Data}) = f(\rho_{12} | \delta, \alpha)$, i.e. conditional on δ and α, the data do *not* modify prior beliefs about ρ_{12}.

Vijverberg (1993, p. 79) employs a uniform prior over the interval $[\underline{\rho}_{12}, \bar{\rho}_{12}]$ for ρ_{12} given ρ_{1u} and ρ_{2u}, and a uniform prior over the unit square for ρ_{1u} and ρ_{2u}. This is equivalent to a uniform prior for ρ_{1u}, ρ_{2u} and ρ_{12} over the region for which Ω is positive definite. His proposed 'posterior distribution' (Vijverberg, 1993, p. 79, eq. (19)), however, is *at best* an approximate posterior distribution (Vijverberg, 1993, p. 79, fn. 11). Instead of employing $f(\text{Data} | \rho_{12})$ in (7), Vijverberg approximates it with the asymptotic sampling distribution for the maximum likelihood estimators of ρ_{1u} and ρ_{2u}. Even if Vijverberg had used the exact finite sampling distribution for these estimators, this would *not* be the correct Bayesian approach and would be subject to marginalization paradoxes (e.g. see Press and Zellner, 1978, p. 312).

The use of this asymptotic approximation also eliminates one of the major advantages of the Bayesian approach over the classical approach – the ability to perform finite sample inference. It is possible to perform an exact Bayesian analysis using Gibbs sampling with data augmentation using a computationally convenient prior (e.g. see McCulloch et al., 1994). Our experience indicates that the convergence properties of the Gibbs sampler are excellent provided a proper prior is used.

Insight into the critical issues involved in Vijverberg (1993) can be obtained by reparameterizing the problem and replacing ρ_{12} by the partial correlation $\rho_{12\cdot u} = [\rho_{12} - \rho_{1u}\rho_{2u}][(1-\rho_{1u}^2)(1-\rho_{2u}^2)]^{-1/2}$. Since $|\Omega| = (1-\rho_{1u}^2)(1-\rho_{2u}^2)$ $(1-\rho_{12\cdot u}^2)$, it follows that Ω is positive definite for ρ_{1u}, ρ_{2u} and $\rho_{12\cdot u}$ over $(-1,1)\times(-1,1)\times(-1,1)$, yielding a prior support for ρ_{1u}, ρ_{2u} and $\rho_{12\cdot u}$ that is a Cartesian product *not* involving these parameters (unlike when parameterizing in terms of ρ_{1u}, ρ_{2u} and ρ_{12}). If a priori independence indicated by $f(\delta|\alpha, \rho_{12\cdot u}) = f(\delta|\alpha)$ and $f(\alpha|\rho_{12\cdot u}) = f(\alpha)$ is assumed, then $f(\text{Data}|\rho_{12\cdot u})$ and $f(\text{Data}|\alpha, \rho_{12\cdot u})$ analogous to (7) and (6) do *not* depend on $\rho_{12\cdot u}$, and so no learning about the partial correlation $\rho_{12\cdot u}$ occurs. Intuitively, this makes sense: the updating of prior beliefs on ρ_{12} occurs as a result of data information on ρ_{1u} and ρ_{2u}. After taking into account the linear effect of u on ε_1 and ε_2, nothing is learned about the remaining correlation between ε_1 and ε_2, i.e. $\rho_{12\cdot u}$.

Vijverberg (1993, pp. 75–78) notes numerous examples motivating interest in the parameter ρ_{12}. Such interest arises whenever quantities involving both sectors (regimes), e.g. $E(Y_1^*|\varepsilon_2 \geqslant 0)$, $E(Y_1^*|\varepsilon_2 > 0, I^* \geqslant 0)$, $E(Y_1{}^*|\varepsilon_2 > 0, I^* \geqslant 0) - E(Y_1^*|I^* \geqslant 0)$, or $P(I^* \geqslant 0| Y_1^* > Y_2^*)$, are of interest. Rewriting such expressions in terms of the partial correlation, and assuming the prior independence noted above, demonstrates that posterior inferences about such quantities depend in part on information arising solely from the prior distribution.

2. The Gibbs sampler: Theoretical derivation

Vijverberg (1993) carries out a pseudo-Bayesian analysis, using Monte Carlo integration to evaluate the properties of a normal approximation for the posterior. Since the key parameters of interest are correlations, the accuracy of this approximation is open to doubt. Fortunately, it is relatively easy to do an exact Bayesian analysis using Gibbs sampling, providing X_1 and X_2 are observed for all time periods.[1] In practice, X_1 is often equal to X_2 (e.g. if the explanatory variables contain individual characteristics) so that the present analysis is widely relevant.

[1] If the X_i is not observed when Y_i^* is not observed ($i = 1, 2$), then the Gibbs sampler as described here is not appropriate. Another step would have to be added to the Gibbs sampler to draw the unobserved elements of X_i. However, to do this we would have to make assumptions about the process generating X_i.

To describe the Gibbs sampler, consider the definitions in (1) and (2), assume that we have n i.i.d. observations, and reorder the data so that we can write:

$$Y_1^* = \begin{bmatrix} Y_1^{\mathrm{O}} \\ Y_1^{\mathrm{M}} \end{bmatrix}, \qquad Y_2^* = \begin{bmatrix} Y_2^{\mathrm{M}} \\ Y_2^{\mathrm{O}} \end{bmatrix}, \tag{9}$$

where the 'O' superscript indicates data which are observed and 'M' indicates missing. Y_1^{O} and Y_2^{O} are of length n_1 and n_2, respectively. Hence,

$$I = \begin{bmatrix} 1_{n_1} \\ 0_{n_2} \end{bmatrix}. \tag{10}$$

We work with the error covariance matrix (3) reparameterized as

$$\Omega = \begin{bmatrix} \sigma_u^2 & \sigma_{1u} & \sigma_{2u} \\ \sigma_{1u} & \sigma_1^2 & \sigma_{12} \\ \sigma_{2u} & \sigma_{12} & \sigma_2^2 \end{bmatrix}. \tag{11}$$

Interest centres on the posterior distribution conditional on the observed data, i.e. $p(\theta | Y_1^{\mathrm{O}}, Y_2^{\mathrm{O}}, I)$. This posterior is difficult to work with directly. However, if we augment the posterior with latent data Y_1^{M}, Y_2^{M} and I^* then a Gibbs sampler can be set up. That is, instead of working with the unwieldy density $p(\theta | Y_1^{\mathrm{O}}, Y_2^{\mathrm{O}}, I)$, we work with $p(\theta, Y_1^{\mathrm{M}}, Y_2^{\mathrm{M}}, I^* | Y_1^{\mathrm{O}}, Y_2^{\mathrm{O}}, I)$. A Gibbs sampler using the full conditionals of this latter density yields a series of draws of θ which can be treated as though they come from the former density.

Conditional on the latent data, the posterior for θ looks as though it takes the familiar normal-Wishart form of a multiple equation regression model. That is, conditional on the latent data and Ω, the regression parameters have a normal distribution and (conditional on the latent data) Ω^{-1} has a Wishart distribution. The fact that we normalize the model by setting $\sigma_u^2 = 1$, causes some minor problems with drawing from the latter distribution. However, a similar problem arises in the Bayesian analysis of the multinomial Probit model, and McCulloch, et al. (1994) describe a way around it. We use an identical method.

We are interested in imposing a prior that says $\sigma_u^2 = 1$ with probability one. We cannot impose this prior in an easy way through a Wishart prior for Ω^{-1}, since the distribution of a Wishart conditional on one element is no longer Wishart. Instead, we reparameterize Ω. If we let $\varepsilon_i = [\varepsilon'_{1i}, \varepsilon'_{2i}]'$, then the joint distribution of $[u_i, \varepsilon'_i]'$, can be written as the marginal of u_i time the conditional distribution of ε_i given u_i, where $i = 1, 2, \ldots, n$ indexes observations. The former distribution is $N(0, \sigma_u^2)$ and the latter is $N(0, \Omega_\varepsilon - \delta\delta'/\sigma_u^2)$, where $E[\varepsilon_i \varepsilon'_i] = \Omega_\varepsilon$ and $E[\varepsilon_i u_i] = \delta$. If we define $\Phi = \Omega_\varepsilon - \delta\delta'/\sigma_u^2$, then there is a one-to-one correspondence between Ω and $[\sigma_u^2, \delta, \Phi]$. Hence, we can put a prior on

$\{\Omega|\sigma_u^2 = 1\}$ by setting $\sigma_u^2 = 1$ and putting priors over δ and Φ. We follow McCulloch et al. (1994) who choose

$$\delta \sim \mathrm{N}(\delta_0, B_0^{-1}), \qquad \Phi^{-1} \sim \mathrm{W}(\nu_0, C_0), \tag{12}$$

where $\mathrm{W}(a, A)$ is the Wishart distribution with degrees of freedom a and mean aA^{-1}. Note that we can write Ω as

$$\Omega = \begin{bmatrix} 1 & \delta' \\ \delta & \Phi + \delta\delta' \end{bmatrix}. \tag{13}$$

This prior does *not* imply that $\rho_{12\cdot u}$ is independent of the other parameters. However, if we set $\delta_0 = 0$ approximate independence occurs.

In this parameterization, the full conditionals of the augmented posterior are:

(i) $\mathrm{p}(Y_1^{\mathrm{M}}| Y_1^{\mathrm{O}}, Y_2^{\mathrm{O}}, I, Y_2^{\mathrm{M}}, I^*)$ takes the form of a multivariate normal density. Since the observations are assumed to be independent of one another, we can draw from univariate normals. The form of this conditional can be easily seen by noting that, conditional on θ:

$$[I_i^*, Y_{1i}^*, Y_{2i}^*]' \sim \mathrm{N}_3(\mu_i, \Omega), \tag{14}$$

where

$$\mu_i = [Z_i\gamma, X_i\beta_1, X_i\beta_2]'. \tag{15}$$

Using the properties of the multivariate normal, it can be seen that the relevant conditional reduces to $\mathrm{p}(Y_{1i}^{\mathrm{M}}|I_i^*, Y_{2i}^*, \theta)$ which is $\mathrm{N}(\mu_{1i}, \omega_{1i})$, $i = 1, 2, \ldots, n_1$. The mean and variance of this distribution are given by

$$\mu_{1i} = X_i\beta_1 + (I_i^* - Z_i\gamma)\left[\frac{\sigma_2^2\sigma_{1u} - \sigma_{12}\sigma_{2u}}{\sigma_u^2\sigma_2^2 - \sigma_{2u}^2}\right] + (Y_{2i}^{\mathrm{O}} - X_i\beta_2) \times \left[\frac{\sigma_u^2\sigma_{12} - \sigma_{2u}\sigma_{1u}}{\sigma_u^2\sigma_2^2 - \sigma_{2u}^2}\right], \tag{16}$$

and

$$\omega_{1i} = \sigma_1^2 - \frac{\sigma_{1u}^2\sigma_2^2 - 2\sigma_{12}\sigma_{2u}\sigma_{1u} + \sigma_u^2\sigma_{12}^2}{\sigma_u^2\sigma_2^2 - \sigma_{2u}^2}. \tag{17}$$

(ii) For analogous reasons, $\mathrm{p}(Y_{2i}^{\mathrm{M}}|I_i^*, Y_{1i}^*, \theta)$ is $\mathrm{N}(\mu_{2i}, \omega_{2i})$, $i = n_1 + 1, n_2 + 1, \ldots, n$. The mean and variance of this distribution are given by

$$\mu_{2i} = X_i\beta_2 + (I_i^* - Z_i\gamma)\left[\frac{\sigma_1^2\sigma_{2u} - \sigma_{12}\sigma_{1u}}{\sigma_u^2\sigma_1^2 - \sigma_{1u}^2}\right] + (Y_{1i}^{\mathrm{O}} - X_i\beta_1) \times \left[\frac{\sigma_u^2\sigma_{12} - \sigma_{2u}\sigma_{1u}}{\sigma_u^2\sigma_1^2 - \sigma_{1u}^2}\right], \tag{18}$$

and

$$\omega_{2i} = \sigma_2^2 - \frac{\sigma_{2u}^2\sigma_1^2 - 2\sigma_{12}\sigma_{2u}\sigma_{1u} + \sigma_u^2\sigma_{12}^2}{\sigma_u^2\sigma_1^2 - \sigma_{1u}^2}. \tag{19}$$

(iii) The conditional distribution of I_i* is complicated by the fact that we know the sign (e.g. if $I_i = 1$ we know I_i^* is positive). With this minor complication, the conditional we use in the Gibbs sampler can be derived in an identical way as the previous one. In particular, for $i = 1, 2, \ldots, n_1$, $I_i^* | \theta$, Y_{1i}^{O}, Y_{2i}^{M}, $I_i = 1$ is $\mathrm{N}(\mu_{Ii}, \omega_{Ii})$ truncated to be nonnegative. For $i = n_1 + 1$, $n_2 +, \ldots, n$, we obtain the same density except that it is truncated to be negative. The mean and variance are given by:

$$\mu_{Ii} = Z_i\gamma + (Y_{1i}^* - X_i\beta_1)\left[\frac{\sigma_2^2\sigma_{1u} - \sigma_{12}\sigma_{2u}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}\right] + (Y_{2i}^* - X_i\beta_2) \times \left[\frac{\sigma_1^2\sigma_{1u} - \sigma_{12}\sigma_{2u}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}\right], \tag{20}$$

and

$$\omega_{Ii} = \sigma_u^2 - \frac{\sigma_{1u}^2\sigma_2^2 - 2\sigma_{12}\sigma_{2u}\sigma_{1u} + \sigma_1^2\sigma_{2u}^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}. \tag{21}$$

(iv) Assuming a flat prior for the regression parameters ($\beta = [\beta'_1, \beta'_2, \gamma']'$), the conditional is just the normal distribution with mean $[W'(\Omega^{-1} \otimes I)W]^{-1} W'(\Omega^{-1} \otimes I)y$ and variance $[W'(\Omega^{-1} \otimes I)W]^{-1}$, where W is a block diagonal matrix containing Z, X_1 and X_2 on the main diagonal.

(v) $\mathrm{p}(\delta | Y_1^*, Y_2^*, I^*, \beta, \Phi)$ is $\mathrm{N}(\mu_\delta, V_\delta)$, where

$$\mu_\delta = V_\delta(\Phi^{-1}\varepsilon' u + B_0\delta_0), \tag{22}$$

$$V_\delta = (u'u\Phi^{-1} + B_0)^{-1}. \tag{23}$$

Note that the conditional on β and the latent data, $u = [u_1, u_2, \ldots, u_n]'$ and $\varepsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]'$ are known.

(vi) $\mathrm{p}(\Phi^{-1} | Y_1^*, Y_2^*, I^*, \beta, \delta)$ is $W(\nu_0 + N, C + [\varepsilon - u\delta']' [\varepsilon - u\delta'])$.

3. The Gibbs sampler: Practical experience

To illustrate some of the ideas discussed above, we use the Gibbs sampler on artificially generated data from an extended Roy model with no regressors and $N = 100$. The covariance matrix is set to the identity matrix for Dataset 1, while for Dataset 2 we choose σ_{12} such that $\rho_{12 \cdot u} = -0.95$.

Two different choices for prior hyperparameters are considered, loosely based on McCulloch et al. (1994). In Experiment 1, we choose prior hyperparameters $\delta_0 = 0_2$, $B_0^{-1} = I/(v_0 + 1)$, $v_0 = 7$, $C = v_0 I$. Note that this prior implies that ρ_{12} is *approximately independent* of the other parameters. In Experiment 2, all prior hyperparameters are as in Experiment 1, except we set $\delta_0 = [1, 1]'$. This latter change induces prior correlation between ρ_{12} and the other parameters.

Figs. 1 and 2 plot prior and posterior densities for ρ_{12} for Experiment 1 using Datasets 1 and 2, respectively. Fig. 3 presents the same information for Experiment 2 using Dataset 1. In all cases, we ran the Gibbs sampler for 20 000 passes after discarding an initial 1000 passes. The Gibbs sampler exhibited very good numerical properties, and the final results passed standard checks of convergence.

Fig. 1 illustrates the lack of identifiability of ρ_{12}: given a prior where ρ_{12} is *approximately independent* of the other parameters, the prior and posteriors of this parameter are approximately the same. Fig. 2 re-emphasizes this point. Note that the data for this figure was simulated assuming $\rho_{12.u} = -0.95$. Clearly, the actual choice of $\rho_{12\cdot u}$ in the simulated data is irrelevant, since Figs 1 and 2 are more or less identical. However, Fig. 3 indicates how prior correlation between parameters can result in learning about ρ_{12}. That is, the prior and posterior can differ widely if ρ_{12} is a priori correlated with the other parameters.

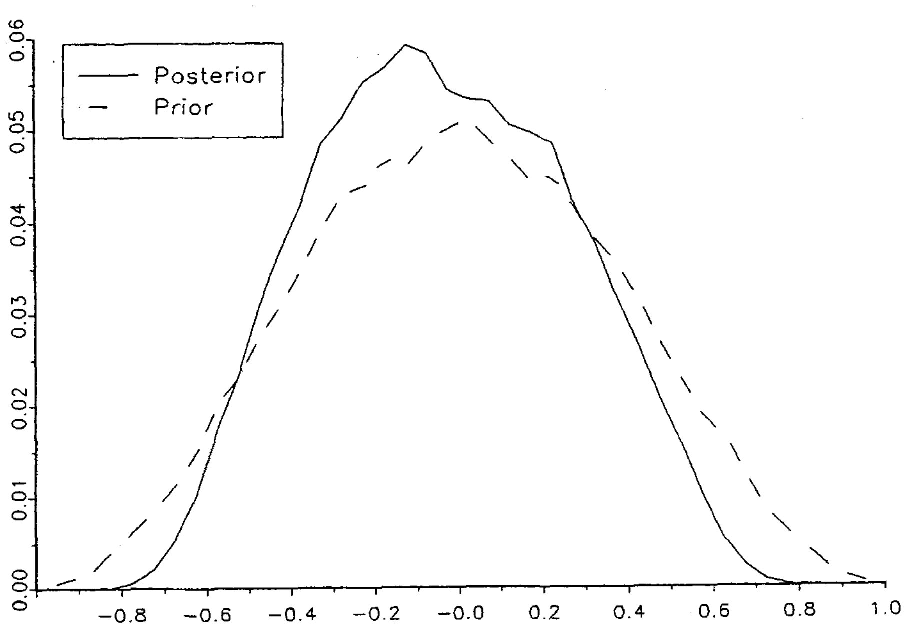

Fig. 1. Prior and posterior densities of ρ_{12} in experiment 1, dataset 1.

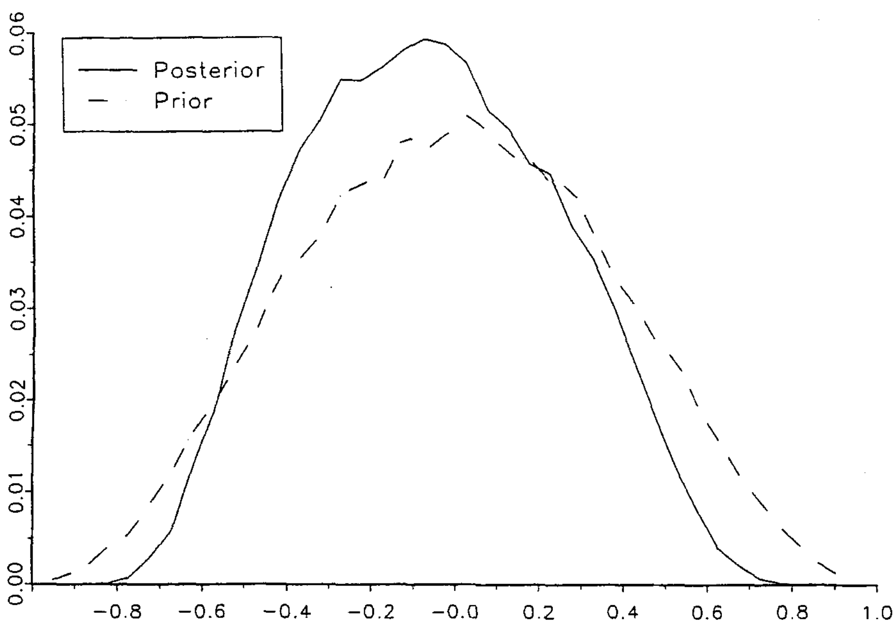

Fig. 2. Prior and posterior densities of ρ_{12} in experiment 1, dataset 2.

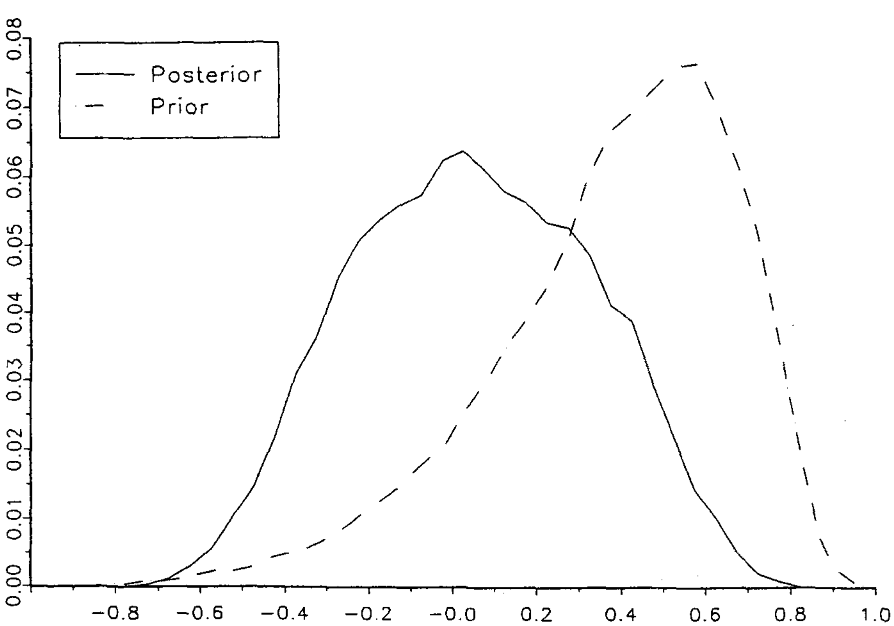

Fig. 3. Prior and posterior densities of ρ_{12} in experiment 2, dataset 1.

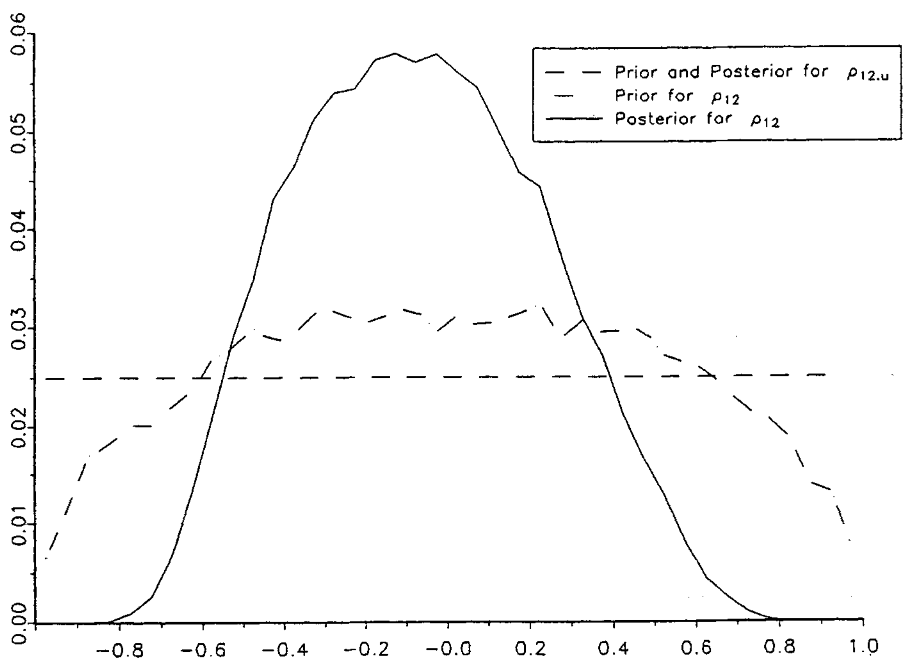

Fig. 4. Posterior densities of ρ_{12} and $\rho_{12\cdot u}$, Vijverberg's prior, dataset 1.

Fig. 4 graphs the prior densities of $\rho_{12\cdot u}$ and ρ_{12} using Vijverberg's prior described previously, and the corresponding posterior density for ρ_{12} for Dataset 1. Vijverberg's prior is flat over $(-1, 1)$ for $\rho_{12\cdot u}$, but the restrictions on ρ_{12} to ensure positive definiteness of Ω cause the prior for this parameter to contain some information. The prior dependence of ρ_{12} on the other parameters of Ω through the positive-definiteness restriction, ensures that posterior learning occurs for ρ_{12}. The shape of the posterior for ρ_{12} in Fig. 4 reflects this learning. However, we do not recommend the use of Vijverberg's prior since the computational demands required are much greater. With this prior, the convenient form for the conditional for Ω in the Gibbs sampler is lost. Fig. 4 is calculated using a Metropolis-within-Gibbs algorithm (see Tierney, 1994)[2]. Of course for Vijverberg's prior, the posterior for $\rho_{12\cdot u}$ is identical to its prior.

[2] The density used for candidate draws for Ω is uniform over $(-1, 1)\times(-1, 1)\times(-1, 1)$ for ρ_{1u}, ρ_{2u}, and ρ_{12} and is gamma for σ_i^{-2} $(i = 1, 2)$ with parameters for the gamma based on output from the McCulloch–Rossi prior.

4. Conclusions

In conclusion, the nonidentified covariance of the Roy/switching regression model is often of great empirical interest and Vijverberg's discussion of how the positive definiteness of the covariance matrix can be used to learn about this key parameter is important. However, it is also important to understand how this learning occurs and what precisely is the role of prior assumptions in this process. The present paper demonstrates that it is prior dependence between identified and nonidentified parameters that drive the learning. We would also like to stress that the approximate Bayesian analysis carried out by Vijverberg could yield misleading results in small samples. Given the ease with which one can carry out an exact Bayesian analysis using Gibbs sampling, we strongly recommend that empirical researchers use such an approach.

References

Berger, J.O. and R.L. Wolpert, 1988, The likelihood principle, 2nd edn. (Institute for Mathematical Statistics, Hayward).

McCulloch, R., N. Polson and P. Rossi, 1994, A Bayesian analysis of the multinomial probit model with fully identified parameters, Manuscript.

Poirier, D.J. and P.A. Ruud, 1981, On the appropriateness of endogenous switching, Journal of Econometrics 16, 249–256.

Press, S.J. and A. Zellner, 1978, Posterior distribution for the multiple correlation coefficient with fixed regressors, Journal of Econometrics 8, 307–321.

Tierney, L., 1994, Markov chains for exploring posterior distributions (with discussion), Annals of Statistics 22, 1701–1762.

Vijverberg, W.P.M., 1993, Measuring the unidentified parameter of the extended Roy model of selectivity, Journal of Econometrics 57, 69–89.