CentER for Economic Research

Discussion paper

TILBURG UNIVERSITY

No. 9063

REGRESSION MODELS UNDER COMPETING COVARIANCE MATRICES: A BAYESIAN PERSPECTIVE

by Siddhartha Chib,
Jacek Osiewalski and Mark Steel

November 1990

ISSN 0924-7815

REGRESSION MODELS UNDER COMPETING COVARIANCE MATRICES: A BAYESIAN PERSPECTIVE

Siddhartha Chib
Washington University
St. Louis, MO 63108, USA

Jacek Osiewalski
Academy of Economics
Kraków, Poland

Mark F.J. Steel
Tilburg University
Tilburg, The Netherlands

Keywords: elliptical distributions; posterior odds;
model mixtures; AR(1) and MA(1) errors.

Abstract: *This paper develops Bayesian approaches to linear elliptical regression models that differ in the covariance structure. A pretest method based on posterior model probabilities is compared with a pooling approach in which the data density is defined as a mixture of elliptical densities with weights that are unknown parameters. All calculations are simple, and prior inputs may be kept to a minimum in an important reference case. An example from the econometrics literature is presented as an illustration of the ideas.*

Acknowledgements: The first two authors wish to acknowledge the hospitality of the Center for Economic Research (CentER), Tilburg University, which provided a congenial environment for writing this paper. The third author has benefited from a research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

1. INTRODUCTION

In recent years, Bayesian researchers have devoted a great deal of attention to the problem of model selection in regression (cf., Gaver and Geisel, 1974, Geweke, 1988, Poirier, 1988, Smith, 1977, Zellner and Siow, 1980, and Zellner, 1984). Usually, the focus has been on selecting the most adequate regression model from a collection of models which differ in their mean, for a given covariance structure of the data. In this paper, as in Poirier (1988), we examine the opposite situation in which the mean is fixed, and the covariances vary, a problem often tackled by the applied modeler, through classical tests for autocorrelation, heteroskedasticity, etc.

The approaches taken here are Bayesian in nature. First we develop the conventional Bayesian pretest approach in which inferences about the parameters and future observations are based on a single model selected through, perhaps, the highest posterior probability criterion. We emphasize, however, an alternative approach in which the competing models are pooled in terms of a finite mixture model, similar to Griffiths and Dao (1980). The selection of a single model is unnecessary in this approach, and information from all the models is combined, an attractive feature for the parameters that are common to all the models. The methods we develop are easy to implement, especially in a convenient reference case that, in addition, requires minimal prior inputs. Finally, we note that all results are derived for general elliptical data densities and finite mixtures of them.

Some comments about the notation that is used. The density of the k-variate Student-t distribution with degrees of freedom ν, location vector μ, and precision matrix Ω, is denoted by $f_S^k(. \mid \nu,\mu,\Omega)$. Similarly, the densities of the Dirichlet distribution with parameter vector γ, and the Beta distribution with parameters a and b, are denoted by $f_D(. \mid \gamma)$ and $f_B(. \mid a,b)$, respectively.

2. THE BAYESIAN MODEL

Consider the m competing linear regression models

$$M_i\colon y = X\beta + \varepsilon \qquad i\colon 1 \to m \tag{1}$$

where the error vector ε has a n-variate elliptical distribution with location vector 0, and dispersion matrix $\sigma^2 V_i$, with σ^2 a common scale factor, and $V_i = V_i(\eta_i)$ a model specific PDS matrix function of η_i, a vector of dimension l_i. Note that the m models share the same location vector $X\beta$ where X is a n×k full column rank matrix that is either not random or independent of the parameters β, σ^2 and η_i. Dynamic models with lagged values of y as regressors are entirely covered by our framework, as mentioned at the end of Subsection 3.1.

Under these assumptions the data density corresponding to the ith model M_i is

$$p(y \mid X,\beta,\sigma^2,\eta_i,M_i) = (\sigma^2)^{-\frac{n}{2}} \, |V_i|^{-\frac{1}{2}} \, g_i[(y-X\beta)'\sigma^{-2}V_i^{-1}(y-X\beta)] \tag{2}$$

where $g_i[.]$, $i=1,...,m$, is a nonnegative function fulfilling the condition (cf. Dickey and Chen, 1985)

$$\int_{\mathbb{R}_+} u^{\frac{n}{2}-1}\, g_i(u)\, du = \Gamma\!\left(\frac{n}{2}\right)\pi^{-\frac{n}{2}} . \tag{3}$$

It should be noted that the model errors in (1) are assigned a very general distribution that gives rise to e.g. the multivariate Normal, Student-t and Pearson type II distributions (see Johnson, 1987).

Suppose that the prior density of the parameters is given by

$$p(\beta,\sigma^2,\eta_i \mid M_i) = c_1\sigma^{-2}\, p(\beta)p(\eta_i \mid M_i), \tag{4}$$

a product of the Jeffreys' type improper prior on σ^2, a prior on the common regression coefficients β, and a possibly model specific prior on η_i, where c_1 is an arbitrary positive constant. As shown in Osiewalski and Steel (1990), this prior structure assures that the joint density of (y, β, η_i) is the same as that obtained under the usual Normality assumption in (1), and is given by

$$p(y,\beta,\eta_i|X,M_i) = \int_{R_+} p(y,\beta,\sigma^2,\eta_i|X,M_i)d\sigma^2 =$$

$$= c_1\ \Gamma\left(\frac{n-k}{2}\right)\ \pi^{-\frac{n-k}{2}}\ p(\beta)p(\eta_i|M_i)\ h_i(\eta_i)\quad f_s^k\left(\beta\,|\,n-k, \hat{\beta}_i, \frac{n-k}{SSE_i}X'V_i^{-1}X\right) \tag{5}$$

defining $h_i(\eta_i) = |V_i|^{-\frac{1}{2}}\,|X'V_i^{-1}X|^{-\frac{1}{2}}\,(SSE_i)^{-\frac{n-k}{2}}$, and where $\hat{\beta}_i = (X'V_i^{-1}X)^{-1}\,X'V_i^{-1}y$ is the generalized least squares estimate and $SSE_i = (y-X\hat{\beta}_i)'\ V_i^{-1}(y-X\hat{\beta}_i)$.

3. POSTERIOR ODDS ANALYSIS AND INFERENCE WITH A SINGLE MODEL

In this section, we derive the posterior probability of model M_i under two different priors on the regression coefficients. If we assign prior probability $p(M_i)$ to the ith model, then the posterior probability of M_i is given by

$$p(M_i \mid y,X) = \frac{p(M_i)\; p(y \mid X,M_i)}{\sum_{j=1}^{m} p(M_j)\; p(y \mid X,M_j)} \tag{6}$$

where $p(y \mid X,M_i)$, $i=1,\ldots,m$, denotes the predictive density.

The Bayesian counterpart of the conventional pretest procedure is to first select a particular model by employing (6) and then conducting inference with the chosen model. In this approach the model choice that minimizes posterior expected loss is suggested. If losses of incorrect decisions are identical then this is equivalent to the criterion of highest posterior model probability.

3.1 Uniform Prior on β

We shall consider in detail the reference case with an improper uniform prior on β in (4). The resulting model probabilities are notable in that they are easy to calculate and prior elicitation only has to be done for the η_i's. In practice, their priors can be chosen to be diffuse. It should be emphasized that although the prior densities on the common parameters β and σ^2 can be improper, the priors on the model-specific parameters η_i have to be proper. Otherwise, posterior probabilities of the models, given in Proposition 1 below, are not well defined due to a dependence on arbitrary constants we may put in the priors of the η_i's.

Let us assume the prior

$$p(\beta,\sigma^2,\eta_i \mid M_i) = p(\beta)p(\sigma^2)p(\eta_i \mid M_i) = c\sigma^{-2}p(\eta_i \mid M_i) \quad , \tag{7}$$

$\beta \in \mathbf{R}^k$, $\sigma^2 \in \mathbf{R}_+$, $c>0$ and $\int p(\eta_i \mid M_i)d\eta_i = 1$, where the integral is taken over the support of η_i, $i=1,\ldots,m$.

Combining the data density in (2), with the prior in (7), and assigning prior probability, $p(M_i)$, to the ith model, $i=1,\dots,m$, we obtain the following result.

Proposition 1: Under (2) and (7) the posterior probability of model i is given by (6) where $p(y \mid X,M_i)$ is the (improper) predictive density given by

$$p(y \mid X,M_i) = c\Gamma(\frac{n-k}{2})\pi^{-\frac{n-k}{2}}\int h_i(\eta_i)p(\eta_i \mid M_i)d\eta_i = c\Gamma(\frac{n-k}{2})\pi^{-\frac{n-k}{2}}K_i \qquad (8)$$

provided the value of the integral $K_i < \infty$, $i=1,\dots,m$.

The proof of Proposition 1 is straightforward in the Normal case, and as mentioned in Section 2, the result carries over to the more general elliptical model in (2).

After choosing a particular model, posterior and predictive inferences are conducted with the retained model on the basis of the standard formulas. For example, if M_i is selected, then the posterior of β is given by

$$p(\beta \mid y,X,M_i) = \int f_s^k(\beta \mid n-k, \hat{\beta}_i, \frac{n-k}{SSE_i}X'V_i^{-1}X)p(\eta_i \mid y,X,M_i)d\eta_i \qquad (9)$$

where the weighting function is the posterior of η_i

$$p(\eta_i \mid y,X,M_i) = K_i^{-1}h_i(\eta_i)p(\eta_i \mid M_i) \ . \qquad (10)$$

Note that $p(M_i \mid y,X)$ can also be expressed as $p(M_i \mid y,X) = p(M_i)K_i \,/ \sum_{j=1}^{m} p(M_j)K_j$ and the Bayes factor B_{rs} of M_r against M_s is K_r/K_s. It is not difficult to see that these Bayes factors are invariant with respect to affine transformations of the data, from

$y \to \tilde{y} = sy+Xq$ ($s \in \mathbf{R}$, $q \in \mathbf{R}^k$).

Also, calculating the Bayes factors, and posterior and predictive densities under a uniform prior on β as in (7) will only require numerical integration of dimension l_i, $i=1,...,m$, which will typically be very small (as in the example in Section 5).

A special case of Proposition 1 provides a direct link with some classical testing results, as have appeared in King (1983, 1987-88). If we calculate (8) under a Dirac prior measure for η_i, namely $p(\eta_i \mid M_i) = I(\eta_i = \eta_i^*)$, $i=1,...,m$, then the resulting Bayes factor becomes

$$B_{rs} = \frac{h_r(\eta_r^*)}{h_s(\eta_s^*)} \ . \tag{11}$$

It can be shown that this is exactly the quantity arising from the use of the Neyman-Pearson lemma for constructing a Most Powerful Invariant test in King (1983, p.40).

The expression in (11) can alternatively be interpreted as the conditional Bayes factor given $\eta_r = \eta_r^*$, and $\eta_s = \eta_s^*$. The theory of maximal invariants, which allows King (1983) to eliminate (β, σ^2) cannot be followed for η_i. Therefore, the sampling-theory analysis has to be conducted for specific values of η_i that restate the model choice in terms of simple hypotheses.

In the more general setting of dynamic linear regression models, Inder (1990) proposes a test for autocorrelation which also conditions on the OLS estimate for the coefficient of the lagged dependent variable. In our framework, the latter coefficient can analytically be integrated out jointly with the coefficients of the exogenous variables, leading to similar predictive densities as in (8). The entire analysis of the static case discussed here directly carries over to dynamic models, without any additional complications.

3.2 Student prior on β

If we use the prior structure in (4) with an independent Student-t density on β, say,

$$p(\beta,\sigma^2,\eta_i \mid M_i) = c_1\sigma^{-2}f_s^k(\beta \mid e,b,A)p(\eta_i \mid M_i) \ , \tag{12}$$

with $\sigma^2 \in \mathbf{R}_+$, $c_1 > 0$ and $\int p(\eta_i \mid M_i) = 1$, we obtain the following posterior results

$$p(\beta \mid \eta_i,y,X,M_i) = H_i(\eta_i)f_s^k(\beta \mid e,b,A)f_s^k(\beta \mid n-k,\hat{\beta}_i,\frac{n-k}{SSE_i}X'V_i^{-1}X) \ , \tag{13}$$

a 2-0 poly-t density (see Drèze, 1977), which can be marginalized with respect to the density

$$p(\eta_i \mid y,X,M_i) = L_i^{-1}H_i(\eta_i)h_i(\eta_i)p(\eta_i \mid M_i) \ . \tag{14}$$

Proposition 2: Under (2) and (12) the posterior probability of model M_i is given by (6) where the predictive densities now take the form

$$p(y \mid X,M_j) = c_1\Gamma(\frac{n-k}{2})\pi^{-\frac{n-k}{2}}L_j, \ j=1,...,m \ , \tag{15}$$

provided all L_j are finite.

The Bayes factor B_{rs} is now L_r/L_s, which is again invariant to affine transformations from $y \rightarrow \tilde{y} = sy+Xq$ ($s \in \mathbf{R}$, $q \in \mathbf{R}^k$) provided we also transform the hyperparameters as follows $\tilde{b} = sb + q$, $\tilde{A} = s^{-2}A$.

The price to pay for using an independent Student-t prior on β in (12) is that the required numerical integrations are now of dimension $l_i + 1$, using the properties of 2-0 poly-t densities (see Richard and Tompa, 1980).

4. MIXTURES OF DATA DENSITIES

In this section, we avoid selecting a particular model and develop a Bayesian pooling approach that does not require the direct specification of prior probabilities for each of the models. Specifically, we consider a mixture of sampling densities, and depart from tradition (cf. Griffiths and Dao, 1980), by letting the weights of the mixture, or prior probabilities of models, be random quantities.

Consider now the data density

$$p(y \mid X, \beta, \sigma^2, \eta, \lambda) = \sum_{i=1}^{m} \lambda_i p(y \mid X, \beta, \sigma^2, \eta_i, M_i) \tag{16}$$

$\lambda_i > 0$, $i = 1,...m$, $\sum_{j=1}^{m} \lambda_j = 1$, and $\eta = (\eta_i,\ i = 1,...m)$, $\lambda = (\lambda_i,\ i = 1,...,m)$, which is a finite mixture of the elliptical densities in (2). The mixing parameter λ_i can be interpreted as $p(M_i \mid \lambda)$, the prior probability of M_i conditional on λ. Sometimes, it is also possible to interpret each λ_i as representing the proportion of the ith subpopulation in an aggregate population. In fact, the stochastic nature of λ leads to a hierarchical structure on the prior model probabilities through $p(M_i) = \int p(M_i \mid \lambda)\ p(\lambda)\ d\lambda = E(\lambda_i)$.

We consider in detail the reference case with independent improper prior

$$p(\beta, \sigma^2, \eta, \lambda) = c_1 \sigma^{-2} p(\beta) p(\eta) p(\lambda) \ , \tag{17}$$

where $p(\eta) = \prod_{j=1}^{m} p(\eta_j)$, and each $p(\eta_j)$ is proper. The prior independence between η_j's reflects our assumption that all the parameters in η are model-specific. Furthermore, we assume that prior information on η_i is not affected by conditioning on any of the models, i.e. $p(\eta_i \mid M_j) = p(\eta_i \mid M_i) = p(\eta_i)$, $j = 1,...,m$. This ensures that the same priors on the η_i's are used both in Section 3 and here. Also, we require the existence of the prior mean vector of $\boldsymbol{\lambda}$, say $\alpha = (\alpha_i,\ i=1,...,m)$. Remark that, by its very definition, α_i is the unconditional prior probability of model i, $p(M_i)$. A natural choice for the prior on $\boldsymbol{\lambda}$ would be a Dirichlet density with parameter vector $C\alpha$, where the scalar C reflects the strength of our beliefs.

As in Section 3, irrespective of the particular elliptical densities chosen in (16), the results after integrating out σ^2 are given by (see Osiewalski and Steel, 1990)

$$\begin{aligned} p(y,\boldsymbol{\beta},\boldsymbol{\eta},\boldsymbol{\lambda} \mid X) &= c_1\Gamma(\frac{n-k}{2})\pi^{-\frac{n-k}{2}}p(\boldsymbol{\beta})p(\boldsymbol{\eta})p(\boldsymbol{\lambda}) \\ &\sum_{i=1}^{m}\lambda_i h_i(\eta_i) f_s^k(\boldsymbol{\beta} \mid n-k, \hat{\boldsymbol{\beta}}_i, \frac{n-k}{SSE_i}X'V_i^{-1}X), \end{aligned} \tag{18}$$

with $\hat{\boldsymbol{\beta}}_i$ and SSE_i defined as previously. Integrating out $\boldsymbol{\lambda}$ we get

$$\begin{aligned} p(y,\boldsymbol{\beta},\boldsymbol{\eta} \mid X) &= c_1\Gamma(\frac{n-k}{2})\pi^{-\frac{n-k}{2}}p(\boldsymbol{\beta})p(\boldsymbol{\eta}) \\ &\sum_{i=1}^{m}\alpha_i h_i(\eta_i) f_s^k(\boldsymbol{\beta} \mid n-k, \hat{\boldsymbol{\beta}}_i, \frac{n-k}{SSE_i}X'V_i^{-1}X). \end{aligned} \tag{19}$$

From (19) it follows that the joint density of y, $\boldsymbol{\beta}$, and $\boldsymbol{\eta}$ is a finite mixture of the densities $p(y,\boldsymbol{\beta},\boldsymbol{\eta} \mid X,M_i)=p(y,\boldsymbol{\beta},\eta_i \mid X,M_i)\prod_{j\neq i}p(\eta_j)$, with $p(y,\boldsymbol{\beta},\eta_i \mid X,M_i)$ as in (5) and the unconditional prior probabilities α_i as weights. More importantly, the uncertainty regarding $\boldsymbol{\lambda}$ is completely irrelevant, in the sense of Lindley (1990, pp. 54-55), for the purpose of prediction, and posterior inference on $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$. Thus, the extension to the "large world" where $\boldsymbol{\lambda}$ is stochastic does not affect the "small world". As a practical matter, this implies that elicitation of the mean prior model probabilities is sufficient, unless $\boldsymbol{\lambda}$ itself is of interest. In other words, if we use a Dirichlet prior on $\boldsymbol{\lambda}$ with parameter $C\alpha$, the value of C does not matter.

Taking $p(\boldsymbol{\beta})$ to be uniform over $\mathbf{R}^k$ in the prior structure

$$p(\beta,\sigma^2,\eta,\lambda) = c\sigma^{-2}p(\eta)p(\lambda) \tag{20}$$

$\beta \in \mathbf{R}^k$, $\sigma^2 \in \mathbf{R}_+$, $c>0$, $\int p(\eta)d\eta = 1$, we can state the following proposition:

Proposition 3: Under (16) and (20) the marginal posterior densities are given by the following finite mixtures

$$p(\beta \mid y,X) = \sum_{j=1}^{m} w_j \, p(\beta \mid y,X,M_j) \tag{21}$$

$$p(\eta_i \mid y,X) = w_i \, p(\eta_i \mid y,X,M_i) + (1-w_i)p(\eta_i) \tag{22}$$

where $w_i = p(M_i \mid y,X) = \alpha_i K_i \,/\, \sum_{j=1}^{m} \alpha_j K_j$ and the mixands are the model-specific posterior densities given in (9) and (10), respectively.

Remark that the weights used to mix the posterior densities in (21) and (22) are exactly the posterior model probabilities given in Proposition 1. The model-specific character of η_i implies that sample information will only enter through M_i. Finally, note that extending our results to other prior distributions for β, as e.g. in Subsection 3.2, is straightforward, but we shall not treat this issue here.

In the case that λ itself is of interest to the model user, the complete specification of its prior density becomes relevant. If we assume in (20) that

$$p(\lambda) = f_D(\lambda \mid C\alpha) \tag{23}$$

the marginal posterior density of λ_i takes the convenient form of a mixture of two Beta densities where the weights are $p(M_i \mid y,X)$ and $1-p(M_i \mid y,X)$,

$$p(\lambda_i \mid y, X) = w_i f_B(\lambda_i \mid C\alpha_i + 1, C(1-\alpha_i)) + (1-w_i) f_B(\lambda_i \mid C\alpha_i, C(1-\alpha_i)+1) \; . \tag{24}$$

It is immediately obvious from (24) that for large C the prior on λ will undergo almost no revision, whereas if C becomes very small the posterior of λ will tend to a Dirac distribution putting point mass on the posterior model probabilities $p(M_i \mid y,X) = w_i$.

5. AN EXAMPLE: AR(1) VERSUS MA(1) ERRORS

To illustrate the ideas developed in the previous sections, we now consider a problem that is extensively discussed in the classical literature (cf. King, 1983, 1987-88, King and McAleer, 1987, Dastoor and Fisher, 1988, and Burke, Godfrey, and Tremayne, 1990), but has hitherto not been analysed in the Bayesian framework.

The problem is concerned with testing whether the errors in the regression model follow a first order autoregressive process, AR(1), as opposed to a moving average process of the same order, MA(1). Interest in this issue appears to have been stimulated by the finding that a significant Durbin-Watson (DW), and more generally Lagrange Multiplier, statistic can imply the presence of either process (cf. Breusch, 1978, and Godfrey, 1978).

We consider the model and data used in Chow (1983, pp. 53-55) given by

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t \tag{25}$$

where y, X_2 and X_3 are the logarithms of the relative price of automobiles, the automobile stock per capita, and the real disposable income per capita, respectively. The data are for the United States for the period 1921-1953, with n=33.

For the model in (25) we let the errors be elliptically distributed, as in (1) and

(2), and specialize $V_i = V_i(\eta_i)$ in the error dispersion matrix, $\sigma^2 V_i$, to take one of the three forms

$$
\begin{aligned}
V_1 &= [(1-\eta_1)^2 I_n + \eta_1 A - \eta_1^2 B]^{-1} \\
V_2 &= (1+\eta_2)^2 I_n - \eta_2 A \qquad\qquad (26) \\
V_3 &= I_n,
\end{aligned}
$$

where $0<\eta_1$, $\eta_2<1$, B = diag(1,0,...,0,1) and A is a tridiagonal matrix whose main diagonal elements are 2 and whose off diagonal elements are -1. Further, we can also obtain that $|V_1| = (1-\eta_1^2)^{-1}$ and $|V_2| = (1-\eta_2^{2n+2})/(1-\eta_2^2)$. It should be noted that the dispersion structure described by V_1 arises through an AR(1) process, given by $\varepsilon_t = \eta_1 \varepsilon_{t-1} + u_t$, while that described by V_2 arises from the MA(1) process, $\varepsilon_t = u_t + \eta_2 u_{t-1}$, $t=1,\dots,n$, where the n+1 dimensional vector $(u_0,u_1,\dots,u_n)'$ is jointly spherically distributed with location vector zero, and dispersion matrix $\sigma^2 I_{n+1}$. In the case of AR(1) the initial element ε_0 is implicitly defined as $\varepsilon_0 = u_0 / \sqrt{1-\eta_1^2}$.

Using results from Subsection 3.1, we obtain the posterior model probabilities and moments given in Table 1, which constitutes a reference case. The prior in (7) is used with both η_1 and η_2 uniformly distibuted on the unit interval. Prior odds of the models are taken to be unity.

Note from Table 1 that prior model probabilities are strongly revised by the data, in favour of the AR(1) specification in M_1. Thus, we expect the posterior moments of β resulting from mixing models as in Section 4 (Proposition 3) to be similar to those of the favoured model. For comparison, Table 2 reports the findings when mixing all three models under the same prior specification as in Table 1, i.e. taking α_i = 1/3 for all i. Revision through the data for η_i can only occur using M_i, since η_i's are model-specific. Therefore, overall posterior results for η_1 are close to the ones conditional upon M_1, whereas those for η_2 are very close to the prior.

Table 1: Posterior results for individual models.

	M_1 AR(1)	M_2 MA(1)	M_3 white noise
$p(M_i)$ $p(M_i \mid y,X)$	0.333 0.932 mean (s.dev)	0.333 0.067 mean (s.dev)	0.333 8.59e-4 mean (s.dev)
$p(\beta \mid y,X,M_i)$	-1.351 (1.900) -0.955 (0.145) 1.282 (0.299)	-2.938 (1.036) -0.896 (0.115) 1.510 (0.162)	-3.222 (0.813) -0.902 (0.091) 1.556 (0.125)
$p(\eta_i \mid M_i)$ $p(\eta_i \mid y,X,M_i)$	0.500 (0.289) 0.722 (0.161)	0.500 (0.289) 0.529 (0.153)	

Although the mixing parameter λ has no clear interpretation in this example, we report its posterior results in Table 3 for the sake of completeness. If we assume $p(\lambda)= f_D(\lambda \mid C\alpha)$, the posterior distribution of λ collapses to the posterior model probabilities given in Table 1 for C tending to zero, whereas for large values of C we essentially reproduce the prior.

Table 2: Posterior results for mixture of models.

	mean (s.dev)
$p(\beta \mid y,X)$	-1.460 (1.854) -0.951 (0.143) 1.298 (0.292)
$p(\eta_1)$ $p(\eta_1 \mid y,X)$	0.500 (0.289) 0.707 (0.182)
$p(\eta_2)$ $p(\eta_2 \mid y,X)$	0.500 (0.289) 0.502 (0.282)

Table 3. Prior and posterior moments of λ.

		C=1.0e-6 mean (s.dev)	C=1 mean (s.dev)	C=1.0e+6 mean (s.dev)
$p(\lambda)$	M_1	0.333 (0.471)	0.333 (0.333)	0.333 (4.71e-4)
	M_2	0.333 (0.471)	0.333 (0.333)	0.333 (4.71e-4)
	M_3	0.333 (0.471)	0.333 (0.333)	0.333 (4.71e-4)
$p(\lambda \mid y,X)$	M_1	0.932 (5.67e-4)	0.633 (0.269)	0.333 (4.71e-4)
	M_2	0.067 (4.22e-4)	0.200 (0.219)	0.333 (4.71e-4)
	M_3	8.59e-4(4.08e-4)	0.167 (0.215)	0.333 (4.71e-4)

6. SUMMARY

In this paper we have considered from the Bayesian perspective the problem of linear elliptical regression models that differ in the covariance structure. We develop two approaches, a pretest method that involves choosing a model based on posterior model probabilities, and a pooling approach in which all models are retained for inference. In the second approach, the data density is defined as a mixture of elliptical densities with weights that are unknown parameters; the stochastic nature of these weights is shown to be irrelevant for prediction and posterior inference on the regression parameters.

All calculations are surprisingly easy and prior elicitation can be kept to a minimum by using a convenient reference case. An example of considerable practical interest to econometricians is presented to illustrate our findings. Many other cases of relevance in applied econometrics, though not explicitly discussed here, are covered by our framework.

References

Breusch, T.S. (1978). Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, 17, 334-355.

Burke, S.P., Godfrey, L.G. and Tremayne, A.R. (1990). Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model: An Alternative Procedure, *Review of Economic Studies*, 57, 135-145.

Chow, G.C. (1983). *Econometrics*. New-York: McGraw-Hill.

Dastoor, N.K. and Fisher, G. (1988). On Point Optimal Cox Tests, *Econometric Theory*, 4, 97-107.

Dickey, J.M. and Chen, C.H. (1985). Direct Subjective Probability Modelling Using Ellipsoidal Distributions, in *Bayesian Statistics 2* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith). Amsterdam: North Holland.

Drèze, J.H. (1977). Bayesian Regression Analysis Using Poly-t Densities, *J. of Econometrics*, 6, 329-354.

Gaver, K.M. and Geisel, M.S. (1974). Discriminating Among Alternative Models: Bayesian and non-Bayesian Methods, in *Frontiers of Econometrics* (ed. P. Zarembka). New-York: Academic Press.

Geweke, J. (1988). Exact Inference in Models with Autoregressive Conditional Heteroscedasticity, in *Dynamic Econometric Modeling* (eds. E. Berndt, H. White, W. Barnett). Cambridge: Cambridge University Press.

Griffiths, W. and Dao, D. (1980). A Note on a Bayesian Estimator in an Autocorrelated Error Model, *J. of Econometrics*, 12, 389-392.

Inder, B.A. (1990). A New Test for Autocorrelation in the Disturbances of the Dynamic Linear Regression Model, *International Economic Review*, 31, 341-354.

Johnson, M.E. (1987). *Multivariate Statistical Simulation*. New-York: Wiley.

King, M.L. (1983). Testing for Autoregressive Against Moving Average Errors in the Linear Regression Model, *J. of Econometrics*, 21, 35-51.

King, M.L. (1987-88). Towards a Theory of Point Optimal Testing (with discussion), *Econometric Reviews*, 6, 169-255.

King, M.L. and McAleer, M. (1987). Further Results on Testing AR(1) Against MA(1) Disturbances in the Linear Regression Model, *Review of Economic Studies*, 54, 649-663.

Lindley, D. V. (1990). The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics, *Statistical Science*, 5, 44-89.

Osiewalski, J. and Steel, M.F.J. (1990). Robust Bayesian Inference in Elliptical Regression Models, CentER Discussion Paper # 9032, Tilburg University.

Poirier, D. J. (1988). Bayesian Diagnostic Testing in the General Linear Normal Regression Model, in *Bayesian Statistics 3* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith). Oxford: University Press.

Richard, J.F. and Tompa, H. (1980). On the Evaluation of Poly-t Density Functions, *J. of Econometrics*, 12, 335-351.

Smith, A.F.M. (1977). Bayesian Statistics and Efficient Information Processing Constrained by Probability Models, in *Decision Making and Change in Human Affairs* (eds. H. Jungerman, and G. de Zeeuw). Dordrecht: D. Reidel.

Zellner, A. (1984). *Basic Issues in Econometrics.* Chicago: University of Chicago Press.

Zellner, A. and Siow, A. (1980). Posterior Odds Ratios for Selected Regression Hypotheses, in *Bayesian Statistics* (eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, A.F.M. Smith). Valencia: University Press.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
8938	G. Marini	Monetary Shocks and the Nominal Interest Rate
8939	W. Güth and E. van Damme	Equilibrium Selection in the Spence Signaling Game
8940	G. Marini and P. Scaramozzino	Monopolistic Competition, Expected Inflation and Contract Length
8941	J.K. Dagsvik	The Generalized Extreme Value Random Utility Model for Continuous Choice
8942	M.F.J. Steel	Weak Exogenity in Misspecified Sequential Models
8943	A. Roell	Dual Capacity Trading and the Quality of the Market
8944	C. Hsiao	Identification and Estimation of Dichotomous Latent Variables Models Using Panel Data
8945	R.P. Gilles	Equilibrium in a Pure Exchange Economy with an Arbitrary Communication Structure
8946	W.B. MacLeod and J.M. Malcomson	Efficient Specific Investments, Incomplete Contracts, and the Role of Market Alternatives
8947	A. van Soest and A. Kapteyn	The Impact of Minimum Wage Regulations on Employment and the Wage Rate Distribution
8948	P. Kooreman and B. Melenberg	Maximum Score Estimation in the Ordered Response Model
8949	C. Dang	The D_3-Triangulation for Simplicial Deformation Algorithms for Computing Solutions of Nonlinear Equations
8950	M. Cripps	Dealer Behaviour and Price Volatility in Asset Markets
8951	T. Wansbeek and A. Kapteyn	Simple Estimators for Dynamic Panel Data Models with Errors in Variables
8952	Y. Dai, G. van der Laan, D. Talman and Y. Yamamoto	A Simplicial Algorithm for the Nonlinear Stationary Point Problem on an Unbounded Polyhedron
8953	F. van der Ploeg	Risk Aversion, Intertemporal Substitution and Consumption: The CARA-LQ Problem
8954	A. Kapteyn, S. van de Geer, H. van de Stadt and T. Wansbeek	Interdependent Preferences: An Econometric Analysis

No.	Author(s)	Title
8955	L. Zou	Ownership Structure and Efficiency: An Incentive Mechanism Approach
8956	P.Kooreman and A. Kapteyn	On the Empirical Implementation of Some Game Theoretic Models of Household Labor Supply
8957	E. van Damme	Signaling and Forward Induction in a Market Entry Context
9001	A. van Soest, P. Kooreman and A. Kapteyn	Coherency and Regularity of Demand Systems with Equality and Inequality Constraints
9002	J.R. Magnus and B. Pesaran	Forecasting, Misspecification and Unit Roots: The Case of AR(1) Versus ARMA(1,1)
9003	J. Driffill and C. Schultz	Wage Setting and Stabilization Policy in a Game with Renegotiation
9004	M. McAleer, M.H. Pesaran and A. Bera	Alternative Approaches to Testing Non-Nested Models with Autocorrelated Disturbances: An Application to Models of U.S. Unemployment
9005	Th. ten Raa and M.F.J. Steel	A Stochastic Analysis of an Input-Output Model: Comment
9006	M. McAleer and C.R. McKenzie	Keynesian and New Classical Models of Unemployment Revisited
9007	J. Osiewalski and M.F.J. Steel	Semi-Conjugate Prior Densities in Multivariate t Regression Models
9008	G.W. Imbens	Duration Models with Time-Varying Coefficients
9009	G.W. Imbens	An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling
9010	P. Deschamps	Expectations and Intertemporal Separability in an Empirical Model of Consumption and Investment under Uncertainty
9011	W. Güth and E. van Damme	Gorby Games - A Game Theoretic Analysis of Disarmament Campaigns and the Defense Efficiency-Hypothesis
9012	A. Horsley and A. Wrobel	The Existence of an Equilibrium Density for Marginal Cost Prices, and the Solution to the Shifting-Peak Problem
9013	A. Horsley and A. Wrobel	The Closedness of the Free-Disposal Hull of a Production Set

No.	Author(s)	Title
9014	A. Horsley and A. Wrobel	The Continuity of the Equilibrium Price Density: The Case of Symmetric Joint Costs, and a Solution to the Shifting-Pattern Problem
9015	A. van den Elzen, G. van der Laan and D. Talman	An Adjustment Process for an Exchange Economy with Linear Production Technologies
9016	P. Deschamps	On Fractional Demand Systems and Budget Share Positivity
9017	B.J. Christensen and N.M. Kiefer	The Exact Likelihood Function for an Empirical Job Search Model
9018	M. Verbeek and Th. Nijman	Testing for Selectivity Bias in Panel Data Models
9019	J.R. Magnus and B. Pesaran	Evaluation of Moments of Ratios of Quadratic Forms in Normal Variables and Related Statistics
9020	A. Robson	Status, the Distribution of Wealth, Social and Private Attitudes to Risk
9021	J.R. Magnus and B. Pesaran	Evaluation of Moments of Quadratic Forms in Normal Variables
9022	K. Kamiya and D. Talman	Linear Stationary Point Problems
9023	W. Emons	Good Times, Bad Times, and Vertical Upstream Integration
9024	C. Dang	The D_2-Triangulation for Simplicial Homotopy Algorithms for Computing Solutions of Nonlinear Equations
9025	K. Kamiya and D. Talman	Variable Dimension Simplicial Algorithm for Balanced Games
9026	P. Skott	Efficiency Wages, Mark-Up Pricing and Effective Demand
9027	C. Dang and D. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms for Computing Solutions of Nonlinear Equations
9028	J. Bai, A.J. Jakeman and M. McAleer	Discrimination Between Nested Two- and Three-Parameter Distributions: An Application to Models of Air Pollution
9029	Th. van de Klundert	Crowding out and the Wealth of Nations
9030	Th. van de Klundert and R. Gradus	Optimal Government Debt under Distortionary Taxation

No.	Author(s)	Title
9031	A. Weber	The Credibility of Monetary Target Announcements: An Empirical Evaluation
9032	J. Osiewalski and M. Steel	Robust Bayesian Inference in Elliptical Regression Models
9033	C. R. Wichers	The Linear-Algebraic Structure of Least Squares
9034	C. de Vries	On the Relation between GARCH and Stable Processes
9035	M.R. Baye, D.W. Jansen and Q. Li	Aggregation and the "Random Objective" Justification for Disturbances in Complete Demand Systems
9036	J. Driffill	The Term Structure of Interest Rates: Structural Stability and Macroeconomic Policy Changes in the UK
9037	F. van der Ploeg	Budgetary Aspects of Economic and Monetary Integration in Europe
9038	A. Robson	Existence of Nash Equilibrium in Mixed Strategies for Games where Payoffs Need not Be Continuous in Pure Strategies
9039	A. Robson	An "Informationally Robust Equilibrium" for Two-Person Nonzero-Sum Games
9040	M.R. Baye, G. Tian and J. Zhou	The Existence of Pure-Strategy Nash Equilibrium in Games with Payoffs that are not Quasiconcave
9041	M. Burnovsky and I. Zang	"Costless" Indirect Regulation of Monopolies with Substantial Entry Cost
9042	P.J. Deschamps	Joint Tests for Regularity and Autocorrelation in Allocation Systems
9043	S. Chib, J. Osiewalski and M. Steel	Posterior Inference on the Degrees of Freedom Parameter in Multivariate-t Regression Models
9044	H.A. Keuzenkamp	The Probability Approach in Economic Methodology: On the Relation between Haavelmo's Legacy and the Methodology of Economics
9045	I.M. Bomze and E.E.C. van Damme	A Dynamical Characterization of Evolutionarily Stable States
9046	E. van Damme	On Dominance Solvable Games and Equilibrium Selection Theories
9047	J. Driffill	Changes in Regime and the Term Structure: A Note

No.	Author(s)	Title
9048	A.J.J. Talman	General Equilibrium Programming
9049	H.A. Keuzenkamp and F. van der Ploeg	Saving, Investment, Government Finance and the Current Account: The Dutch Experience
9050	C. Dang and A.J.J. Talman	The D_1-Triangulation in Simplicial Variable Dimension Algorithms on the Unit Simplex for Computing Fixed Points
9051	M. Baye, D. Kovenock and C. de Vries	The All-Pay Auction with Complete Information
9052	H. Carlsson and E. van Damme	Global Games and Equilibrium Selection
9053	M. Baye and D. Kovenock	How to Sell a Pickup Truck: "Beat-or-Pay" Advertisements as Facilitating Devices
9054	Th. van de Klundert	The Ultimate Consequences of the New Growth Theory; An Introduction to the Views of M. Fitzgerald Scott
9055	P. Kooreman	Nonparametric Bounds on the Regression Coefficients when an Explanatory Variable is Categorized
9056	R. Bartels and D.G. Fiebig	Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads
9057	M.R. Veall and K.F. Zimmermann	Evaluating Pseudo-R^2's for Binary Probit Models
9058	R. Bartels and D.G. Fiebig	More on the Grouped Heteroskedasticity Model
9059	F. van der Ploeg	Channels of International Policy Transmission
9060	H. Bester	The Role of Collateral in a Model of Debt Renegotiation
9061	F. van der Ploeg	Macroeconomic Policy Coordination during the Various Phases of Economic and Monetary Integration in Europe
9062	E. Bennett and E. van Damme	Demand Commitment Bargaining: - The Case of Apex Games
9063	S. Chib, J. Osiewalski and M. Steel	Regression Models under Competing Covariance Matrices: A Bayesian Perspective

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLANDS