Working Paper 91-27
November 1991

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

ROBUST BAYESIAN INFERENCE IN EMPIRICAL REGRESSION MODELS*

Jacek Osiewalski** and Mark F.J. Steel***

Abstract

Broadening the stochastic assumptions on the error terms of regression models was prompted by the analysis of linear multivariate *t* models in Zellner (1976). We consider a possible non-linear regression model under *any* multivariate elliptical data density, and examine Bayesian posterior and productive results. The latter are shown to be robust with respect to the specific choice of a sampling density within this elliptical class. In particular, sufficient conditions for such model robustness are that we single out a precision factor τ^2 on which we can specify an improper prior density. Apart from the posterior distribution of this nuisance parameter τ^2, the entire analysis will then be completely unaffected by departures from Normality. Similar results hold in finite mixtures of such elliptical densities, which can be used to average out specification uncertainty.

Key words: Multivariate elliptical data densities, model robustness, improper priors, finite mixtures.

* This paper was written while the first author was a visiting fellow at the Center of Operations Research and Econometrics (CORE), Louvain-la-Neuve, Belgium. The second author held a research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW), and benefited from the hospitality of the Universidad Carlos III de Madrid. Useful discussions with Julián de la Horra, Jean-Pierre Florens, Michel Mouchart, Theo Nijman and Jean-Marie Rolin are gratefully acknowledged, though the usual disclaimer applies. Two anonymous referees provided us with insightful comments that improved the presentation.

** Academy of Economics. Kraków, Poland.
*** Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain.

1. INTRODUCTION

The Bayesian analysis of regression models with dependent non-Normal error terms has received considerable attention, especially since the seminal paper of Zellner (1976), who considered linear multivariate Student t regression models. This assumption was extended to scale mixtures of Normal distributions in Jammalamadaka *et al.* (1987) and in Chib *et al.* (1988) whereas Osiewalski (1991) and Chib *et al.* (1990a) generalize, in addition, to nonlinear models. Here we shall examine a further generalization to the *entire class of multivariate elliptical or ellipsoidal densities,* as it was defined in e.g. Kelker (1970), Cambanis *et al.* (1981) or Dickey and Chen (1985).

In this paper, we show that any multivariate elliptical regression model, combined with an improper reference prior on a "nuisance" scalar precision parameter τ^2, will lead to exactly the same posterior and predictive analyses as in the Normal case. Thus, in this sense, inference is fully *robust* with respect to changes in the specification of the sampling process within this wide class of elliptical densities. Remark that this property differs from robustness against extreme observations, as used e.g. in Ramsay and Novick (1980), who defined a concept of "L robustness". The latter relates to the sensitivity of the likelihood to the *data*, and is based on the influence function. Instead, we arrive at robustness of posterior and predictive results with respect to the *sampling model* itself, within a broad class of models that includes, e.g., multivariate Student or Pearson II models. Thus, we focus on "model robustness" [see Berger (1985, p. 248)], and in particular, on what Box and Tiao (1973, p. 152) call "inference robustness". Classical counterparts of these findings were derived by Zellner (1976) for the Student t case and by Girón *et al.* (1989) for scale mixtures of Normals, which is a subclass of the elliptical family [see, e.g., Kelker (1970)].

These robustness results are derived for multivariate elliptical distributions, and do not generally hold under independent non-Normal error terms. If we assume that the errors are independently and identically distributed according to some elliptical process other than the Normal, no such robustness occurs. The results in Box and Tiao (1973, Ch. 3), West (1984) and Bagchi and Guttman (1988) provide some evidence in this respect. However, if we start from a multivariate elliptical framework, where independence can only be accommodated under Normality [see Kelker (1970, Lemma 5)], the usual improper reference prior on τ^2 does the trick. Only posterior results for τ^2 are affected by departures from Normality, as in Zellner (1976). Given the nuisance character of this scale factor, however, these results are not explicitly stated here. Predictive inference and posterior inference on the parameter θ, which defines the location and shape of the ellipsoids, can then be conducted exactly as in the usual Normal case. Any remaining parameters, introduced to index the way the density function changes over ellipsoids, will only be updated through their possible prior dependence on θ.

A finite mixture of elliptical data densities is then considered for cases in which we want to avoid a single specification. The mixing will be preserved in posterior and predictive analyses, which allows broadening the class of data densities, without really affecting the complexity of the ensuing analysis. It is just like mixing Normal distributions defined over different ellipsoids, where the mixing parameter λ will be revised by the sample, albeit in a rather moderate way.

Section 2 introduces the Bayesian model, on the basis of which we derive posterior and predictive results in Sections 3 and 4, respectively. Finite mixtures of data densities are examined in Section 5, whereas a final section summarizes some conclusions.

2. THE BAYESIAN MODEL

2.1. The Elliptical Sampling Model

A general form of elliptical, also known as ellipsoidal, distributions will be assumed for the sampling process. The observation vector $y \in \mathbb{R}^n$ has an $n-$variate continuous elliptical distribution, given a set of exogenous variables X and a sufficient parameterization, say ω, if and only if its data density is

$$p(y \mid X, \omega) = \mid \tilde{V}(X, \tilde{\eta}) \mid^{-\frac{1}{2}} g_{n,\nu}\left[(y - h(X,\beta))'(\tilde{V}(X,\tilde{\eta}))^{-1}(y - h(X,\beta))\right]. \quad (2.1)$$

In (2.1) $g_{n,\nu}(\cdot)$ is a nonnegative function, which for any n and ν has to fulfil the condition

$$\int_0^\infty u^{\frac{n}{2}-1} g_{n,\nu}(u)du = \Gamma(\frac{n}{2})\pi^{-\frac{n}{2}}. \quad (2.2)$$

It can be shown [see Cambanis *et al.* (1981), Dickey and Chen (1985), Kelker (1970)] that (2.2) is both necessary and sufficient to make (2.1) a proper, normalized density function.

The location vector in (2.1) is the, possibly nonlinear, but known, function $h(X, \beta)$, and the scale matrix is $\tilde{V}(X, \tilde{\eta})$, where $\tilde{V}$ is positive definite symmetric (PDS) and a known matrix function of X and $\tilde{\eta}$. Therefore, $\beta \in B \subseteq \mathbb{R}^k$ and $\tilde{\eta} \in \tilde{H} \subseteq \mathbb{R}^q$ serve to define the isodensity ellipsoids of y. The labelling function $g_{n,\nu}$ that determines the density value for each of these ellipsoids [see e.g. Leamer (1978, p. 150)] is indexed by n and $\nu \in N \subseteq \mathbb{R}^\ell$, which may contain parameters other than β and $\tilde{\eta}$, introduced specifically for the purpose of describing $g_{n,\nu}$. A well-known example is found in the multivariate Student t distribution, where $\nu \in \mathbb{R}_+$ and

$$g_{n,\nu}(\cdot) = \frac{\Gamma(\frac{n+\nu}{2})}{\Gamma(\frac{\nu}{2})} (\nu\pi)^{-\frac{n}{2}}(1 + \frac{\cdot}{\nu})^{-\frac{n+\nu}{2}}.$$

Indeed, from (2.1) we will then obtain a Student t data density with ν degrees of freedom, location vector $h(X, \beta)$ and precision matrix $\tilde{V}(X, \tilde{\eta})^{-1}$, denoted by

$$p(y \mid X, \omega) = f_S^n(y \mid \nu,\ h(X, \beta),\ \tilde{V}(X, \tilde{\eta})^{-1}). \tag{2.3}$$

A generalization of (2.3), where the dimension of ν is extended, can be found in Dickey and Chen (1985, p. 173). However, in some cases N will be empty and $g_{n,\nu}(\cdot)$ will only depend on n, the dimension of y. If, in particular, we choose

$$g_{n,\nu}(\cdot) = g_n(\cdot) = (2\pi)^{-\frac{n}{2}} \exp\,(-\frac{\cdot}{2}),$$

our data density in (2.1) will be of the Normal form with mean $h(X, \beta)$ and covariance matrix $\tilde{V}(X, \tilde{\eta})$:

$$p(y \mid X, \omega) = f_N^n(y \mid h(X, \beta),\ \tilde{V}(X, \tilde{\eta})), \tag{2.4}$$

where $(\beta, \tilde{\eta})$ is now a sufficient parameterization.

Another type of multivariate elliptical distribution is the Pearson Type II distribution, as described in, e.g., Johnson (1987) where the entire probability mass is located inside a finite ellipsoid, i.e., a case with truncated tails.[1] The labelling function becomes

$$g_{n,\nu}(\cdot) = \frac{\Gamma(\frac{n+\nu}{2})}{\Gamma(\frac{\nu}{2})} \pi^{-\frac{n}{2}}(1 - \cdot)^{\frac{\nu}{2}-1}$$

[1]The bounded support assumption prompts Spanos (1990b) to suggest this type of distribution for modelling stock share returns.

on the support $(y - h(X,\beta))'(\tilde{V}(X,\tilde{\eta}))^{-1}(y - h(X,\beta)) \leq 1$, and with $\nu \in I\!R_+$, whereas $g_{n,\nu}(\cdot) = 0$ elsewhere. The ensuing density will be denoted as

$$p(y \mid X, \omega) = f^n_{PII}(y \mid \nu,\ h(X,\beta),\ \tilde{V}(X,\tilde{\eta})^{-1}). \tag{2.5}$$

Finally, the bivariate Laplace and generalized Laplace (or Bessel) distributions have received some attention in the literature [see, e.g., McGraw and Wagner (1968)].

The explanatory variables in X have a sampling distribution whose sufficient parameterization is denoted by λ. If we assume that the joint prior on ω and λ is a product of $p(\omega)$ and $p(\lambda)$, both σ-finite, we can ignore the process of X for the purpose of conducting inference with (2.1). These assumptions, in fact, amount to operating a Bayesian cut [see e.g. Florens and Mouchart (1985) and Florens *et al.* (1990)].

Provided second order moments exist, the covariance matrix for any elliptical density in (2.1) will be proportional to $\tilde{V}(X,\tilde{\eta})$. By choosing a diagonal $\tilde{V}(X,\tilde{\eta})$ we thus obtain zero correlations, but from Kelker (1970) we know that independence then only holds under Normality, i.e. (2.4). Non-Normal multivariate elliptical densities can combine zero correlation with dependence and a referee suggested this may be useful for modelling ARCH-type behaviour [see Engle (1982)]. In fact Spanos (1990a) examines this connection in some detail and finds that,e.g., the Student t distribution can indeed be used to treat the issue of dynamic heteroskedasticity in a natural fashion. Clearly, in the latter context the dimension n of y relates to time. In other contexts, it might be more appropriate to let n be the dimension of some vector observed at a particular point in time, and possibly consider many (independent) observations from (2.1), as e.g., in van Praag and Wesselman (1989) or in Spanos (1990b). However, for repeated independent sampling from (2.1) our robustness results do not hold. See, e.g., West (1984) who assumes $n = 1$.

Finally, we shall find it useful to reparameterize $\tilde{\eta}$ into (η, τ^2) such that

$$\tilde{V}(X,\tilde{\eta}) = \frac{1}{\tau^2}\, V(X,\eta), \tag{2.6}$$

where $\tau^2 \in I\!R_+$ is a scalar precision parameter and $V(X,\eta)$ is a normalized (e.g., through imposing tr $V(X,\eta) = n$) scale matrix with $\eta \in H$. For notational convenience, we now define $\theta = (\beta,\eta)$ which contains all the information about the location and shape of the ellipsoids. We also partition ν into (ν_1,ν_2) where $\nu_1 = f(\theta)$ and $\nu_2 \in N_2 \subseteq N$ only serves to describe the tail behaviour.

2.2. Prior Densities

We now face the task of completing the Bayesian model by assigning a prior distribution to $\omega = (\theta, \tau^2, \nu_2)$. Typically, θ will be the parameter of interest, although in some cases we may want to conduct inference on ν_2 (but then we need to specify a particular $g_{n,\nu}(\cdot)$). Accordingly, we leave the specification of the prior density of (θ, ν_2) completely free at this stage. We shall see in Section 3 that if we choose the ("usual") improper prior structure

$$p(\theta, \tau^2, \nu_2) = p(\tau^2)p(\theta, \nu_2) = \frac{c}{\tau^2}\, p(\theta, \nu_2), \tag{2.7}$$

where c is a positive constant and $p(\theta, \nu_2)$ is functionally independent of τ^2, the analysis will simplify greatly. More in particular, the actual form of $g_{n,\nu}(\cdot)$ becomes completely irrelevant, so that both posterior and predictive analyses are *fully robust* with respect to any departures from Normality in the wide class of multivariate elliptical densities.

The prior (2.7) implicitly excludes certain forms of the labelling function $g_{n,\nu}(\cdot)$ and the scale matrix $\tilde{V}(X, \tilde{\eta})$. In particular, τ^2 and (θ, ν_2) are variation free so that τ^2 is not functionally related to ν and, therefore, does not index $g_{n,\nu}(\cdot)$.[2] Also, (2.7) allows neither functional dependence between τ^2 and η nor concentrating all prior probability mass at one value for τ^2. This excludes, e.g., $\tilde{V}(X, \tilde{\eta}) = I_n$ [as in Hill (1969)] which could be reparameterized as (2.6) by assuming either $V(X, \eta) = f(\eta) I_n$ and $\tau^2 = f(\eta)$ for any positive function $f(\cdot)$ or $V(X, \eta) = I_n$ and a Dirac prior at $\tau^2 = 1$. Clearly, informative prior densities that are natural conjugate for the Normal case are excluded by (2.7). Such proper prior structures are discussed in the elliptical context by Osiewalski and Steel (1991).

[2] All this parameter does is to influence the scale of the ellipsoids; under the first equality sign in (2.7) the way the density function changes over ellipsoids no longer depends on τ^2. The fact that the interpretation of τ^2 does not vary over the elliptical class in (2.1) allows assigning a specific prior density to τ^2 in the second part of (2.7) without actually choosing a particular model.

3. POSTERIOR INFERENCE

Assuming (2.6) and combining the general class of elliptical data densities in (2.1) with the improper prior family in (2.7), we obtain the joint density

$$p(y,\omega \mid X) = c\, p(\theta,\nu_2)\ (\tau^2)^{\frac{n}{2}-1} \mid V(X,\eta) \mid^{-\frac{1}{2}}\ g_{n,\nu}[\tau^2 d(y,X,\theta)], \qquad (3.1)$$

where we have defined

$$d(y,X,\theta) = (y - h(X,\beta))'\ V(X,\eta)^{-1}(y - h(X,\beta)).$$

Let us now consider the transformation from (y,θ,τ^2,ν_2) to (y,θ,r^2,ν_2), where

$$r^2 = \tau^2 d(y,X,\theta), \qquad (3.2)$$

leading to

$$p(y,\theta,r^2,\nu_2 \mid X) = c\, p(\theta,\nu_2)\ \mid V(X,\eta)\mid^{-\frac{1}{2}}\ d(y,X,\theta)^{-\frac{n}{2}}(r^2)^{\frac{n}{2}-1} g_{n,\nu}(r^2). \qquad (3.3)$$

The function $g_{n,\nu}(\cdot)$ is not affected by the transformation in (3.2), since it does not involve τ^2, so that property (2.2) can directly be applied to integrate out r^2 in (3.3). This leaves us with

$$p(y,\theta,\nu_2 \mid X) = c\, \Gamma\left(\frac{n}{2}\right)\ \pi^{-\frac{n}{2}}\ p(\theta,\nu_2)\ \mid V(X,\eta)\mid^{-\frac{1}{2}} d(y,X,\theta)^{-\frac{n}{2}}, \qquad (3.4)$$

which no longer depends on the form of $g_{n,\nu}(\cdot)$. The joint (improper) density of our parameters of interest and y is thus *completely robust* with respect to *any* departures from Normality in the class of elliptical data densities (2.1) when τ^2 is treated by assuming the improper prior (2.7). In addition, (3.4) is invariant with respect to rescaling $V(X,\eta)$ by any positive scalar function $z(X,\eta)$.

Let us now assume that the prior $p(\theta,\nu_2)$ is defined as a product of a proper $p(\nu_2 \mid \theta)$ and a σ-finite $p(\theta)$, which makes (3.4) integrable in ν_2. If the resulting density $p(y,\theta \mid X)$ is also integrable in θ over $\Theta \subseteq B \times H$, we are sure that the posterior of (θ,ν_2) is well defined as

$$p(\theta,\nu_2 \mid y,X) \propto p(\theta,\nu_2)\ \mid V(X,\eta)\mid^{-\frac{1}{2}}\ d(y,X,\theta)^{-\frac{n}{2}}, \qquad (3.5)$$

from which we can easily derive the posterior for the location and shape parameters θ.

Theorem 1: For *any* elliptical data density (2.1), assuming (2.6) and using an improper prior (2.7), which is integrable in ν_2, we obtain the *same* posterior of θ:

$$p(\theta \mid y, X) \propto p(\theta) \ \mid V(X,\eta) \mid^{-\frac{1}{2}} \ d(y, X, \theta)^{-\frac{n}{2}}, \tag{3.6}$$

where $p(\theta) = \int_{N_2} p(\theta, \nu_2) d\nu_2$ and we have assumed that (3.6) is integrable in θ over Θ. ∎

Of course, (3.6) is exactly the posterior one obtains for the Normal data density (2.4), and may look even more familiar if we consider the simple linear case:

Corollary 1: In the special linear case of Theorem 1 where $h(X, \beta) = X\beta$ and $\Theta = \mathbb{R}^k \times H$, the posterior densities of θ are given by

$$p(\beta \mid \eta, y, X) = K(\eta)^{-1} \ p(\beta, \eta) \ f_S^k(\beta \mid n-k, \ \hat{\beta}, \ s^{-2} X' V(X, \eta)^{-1} X) \tag{3.7}$$

and

$$p(\eta \mid y, X) \propto K(\eta) \ \mid V(X,\eta) \mid^{-\frac{1}{2}} \mid X'V(X,\eta)^{-1} X \mid^{-\frac{1}{2}} \ (s^2)^{-\frac{1}{2}(n-k)}, \tag{3.8}$$

where

$$\hat{\beta} = (X'V(X,\eta)^{-1}X)^{-1} \ X'V(X,\eta)^{-1} y$$
$$s^2 = \frac{1}{n-k} \ (y - X\hat{\beta})' \ V(X,\eta)^{-1} \ (y - X\hat{\beta})$$

and $K(\eta)$, the inverse of the normalizing constant of (3.7), absorbs the prior information on η. ∎

Implicitly, we have also made the assumption that X is of full column rank in Corollary 1, which implies $n \geq k$ in this linear case. If we specify a uniform prior on β, i.e. $p(\beta, \eta) \propto p(\eta)$, we simply have a Student t conditional posterior of β, which is proper if $n > k$. Moments of (3.7) then exist up to (not including) order $n - k$. Adding some prior information will typically lead to the existence of higher order moments. In particular, if $p(\beta, \eta)$ contains a Student t kernel for β with ν_0 degrees of freedom, the conditional posterior in (3.7) will be of a 2-0 poly-t form [see Drèze (1977) and Richard and Tompa (1980)], allowing for posterior moments up to order $\nu_0 + n$.

The invariance results obtained here are a direct consequence of the fact that, after integrating out τ^2 under (2.7), we have

$$p(y \mid X, \theta, \nu_2) = p(y \mid X, \theta) \ \propto \ d(y, X, \theta)^{-\frac{n}{2}}, \tag{3.9}$$

irrespective of the form of $g_{n,\nu}(\cdot)$.[3] Therefore, we address a particular case of Hill (1969), who proposed specifying a spherical model without considering a scale parameter. In the framework of (2.1) he does *not* impose (2.6) and introduces sphericity by assuming $\tilde{V}(X, \tilde{\eta}) = I_n$ directly. The more "traditional" approach, in e.g. Zellner (1976), Jammalamadaka *et al.* (1987), Chib *et al.* (1988) and Osiewalski (1991), implicitly starts from the deeper level of parameterization used here and amounts to assuming (2.6). In that case sphericity is induced by taking $V(X, \eta) = I_n$. Hill's (1969) specification of general spherical errors is thus made at a level of parameterization comparable to the one in (3.9). By not imposing (2.6), with functionally independent η and τ^2 and without a Dirac prior of τ^2, Hill's approach is slightly more general, but at the cost of not obtaining the robustness that follows from (3.9). Nevertheless, Hill (1969) does introduce a scale factor in his discussion of Normality. At that level, our results imply that it is *not* the Normality assumption but the use of Jeffreys' prior on this scale factor [as in (2.7)] that accounts for finding the "usual" posterior results. Therefore, provided one is willing to accept (2.6), Normality does not seem to be quite as restrictive as suggested by Hill.

The joint model in (3.4) clearly illustrates a case of marginal underidentification. Florens *et al.* (1990) discuss this issue in detail and show that it can even occur under proper priors. The parameter ω is minimal sufficient in the complete Bayesian model (3.1), but after marginalizing with respect to τ^2 in (3.4), (θ, ν_2) are no longer minimal sufficient and ν_2 is not identified. This means that *given* θ the sample contains no information regarding ν_2, so that conditionally upon θ ν_2 is not updated through the observations. Thus, under the conditions of Theorem 1, we then have $p(\nu_2 \mid \theta, y, X) = p(\nu_2 \mid \theta)$. The *marginal* prior on ν_2, however, will generally be updated [see also Drèze and Richard (1983, p.522)], since the marginal posterior can be written as

$$p(\nu_2 \mid y, X) \propto \int_{\Theta} p(\nu_2 \mid \theta)\, p(\theta \mid y, X)\, d\theta$$

where $p(\theta \mid y, X)$ was defined in (3.6). Thus, if $p(\nu_2 \mid \theta)$ does not depend on θ (i.e.

[3]One of the referees has suggested to look for a semiparametric presentation. In principle, we could drop the finitely dimensional index ν and treat the labelling function g itself as an infinitely dimensional parameter in the space G_n of all nonnegative functions satisfying (2.2). We would then interpret (2.1) as $p(y \mid X, \theta, \tau^2, g)$, and under (2.6) and the prior structure $p(\theta, \tau^2, g) = p(\tau^2)\, p(\theta, g) = c\, \tau^{-2} p(\theta, g)$, we would obtain $p(y \mid X, \theta, g)$ which no longer depends on g, just as (3.9) does not depend on ν_2. The crucial problem would then be to construct probabilities on G_n, which can easily give rise to very subtle and complicated problems, as explained in Diaconis and Friedman (1986).

independence in probability if $p(\theta)$ is proper and functional independence if it is not) the sample cannot revise the marginal prior of ν_2 either and we state:

Theorem 2: Under the conditions of Theorem 1, the prior structure for (θ, ν_2)

$$p(\theta, \nu_2) = p(\theta)\ p(\nu_2) \tag{3.10}$$

will prevent updating of the marginal prior information on ν_2, i.e.

$$p(\nu_2 \mid y, X) = p(\nu_2). \tag{3.11}$$

■

The lack of dependence in (3.10), which is taken to be integrable in ν_2, will, for any elliptical sampling model (2.1) under (2.6), lead to posterior independence of θ and ν_2, provided we express our prior ignorance about τ^2 by the class of improper densities in (2.7), and if the joint posterior exists, which is assured if (3.6) is integrable in θ. This can be seen directly from (3.5), and, given the fact that the sample can only update ν_2 through θ, this posterior independence will make sure that our marginal opinions regarding ν_2 will not be revised through the observations. Of course, inference on ν_2 only makes sense given a choice of a particular $g_{n,\nu}(\cdot)$. Chib *et al.* (1990a) analyse the subclass of (2.1) where the elliptical densities can be described as scale mixtures of Normals. A prominent member of this subclass is the Student t model in (2.3), in which case Theorem 2 exactly reduces to their Corollary 4, stating a set of sufficient conditions for the impossibility to update the prior of the degrees-of-freedom parameter.

4. PREDICTIVE ANALYSIS

Alternatively, we can focus on the predictive properties of Bayesian models involving elliptical data densities as in (2.1) and improper priors as in (2.7), maintaining also (2.6).

For this purpose, we partition the n dimensional vector y as follows

$$y = \begin{pmatrix} y_{(1)} \\ y_{(2)} \end{pmatrix},$$

where $y_{(i)} \in \mathbb{R}^{n_i}$ $(i = 1, 2)$; $n = n_1 + n_2$, and we are interested in forecasting $y_{(2)}$, given $y_{(1)}$ and X. Conformably, we partition

$$h(X, \beta) = \begin{bmatrix} h_{(1)}(X, \beta) \\ h_{(2)}(X, \beta) \end{bmatrix} \equiv \begin{pmatrix} h_{(1)} \\ h_{(2)} \end{pmatrix}$$

and

$$V(X,\eta) = \begin{bmatrix} V_{11}(X,\eta) & V_{12}(X,\eta) \\ V_{21}(X,\eta) & V_{22}(X,\eta) \end{bmatrix} \equiv \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where the defining equalities are just used to economize on notation. From (3.4) it is immediately clear that the form of $g_{n,\nu}(\cdot)$ will not affect the predictive analysis either, and we obtain directly

$$\begin{aligned} p(y_{(1)}, y_{(2)}, \theta, \nu_2 \mid X) = c\,\Gamma(\frac{n_1}{2})\ \pi^{-\frac{n_1}{2}}\ p(\theta, \nu_2)\ \mid V_{11} \mid^{-\frac{1}{2}}\ a(y_{(1)}, X, \theta)^{-\frac{n_1}{2}} \\ f_S^{n_2}(y_{(2)} \mid n_1,\ h_{(2)} + V_{21} V_{11}^{-1}(y_{(1)} - h_{(1)}),\ \frac{n_1}{a(y_{(1)}, X, \theta)}\ V_{22.1}^{-1}), \end{aligned} \tag{4.1}$$

with $a(y_{(1)}, X, \theta) = (y_{(1)} - h_{(1)})'\ V_{11}^{-1}(y_{(1)} - h_{(1)})$ and $V_{22.1} = V_{22} - V_{21} V_{11}^{-1} V_{12}$. Given our assumption of integrability of the joint prior in ν_2, it is trivial to integrate it out, as in Section 3. The posterior of θ given the first subsample $y_{(1)}$ will be of exactly the same form as (3.6) in Theorem 1, but with $y_{(1)}$ instead of y throughout:

$$p(\theta \mid y_{(1)}, X) \propto p(\theta) \mid V_{11} \mid^{-\frac{1}{2}}\ a(y_{(1)}, X, \theta)^{-\frac{n_1}{2}}, \tag{4.2}$$

provided (4.2) is integrable in θ over Θ. The predictive density thus becomes the Student density in (4.1) of $y_{(2)}$, given $y_{(1)}$, X and θ, weighted by this posterior on the basis of $y_{(1)}$:

$$p(y_{(2)} \mid y_{(1)}, X) = \int_{\Theta} p(y_{(2)} \mid y_{(1)}, X, \theta)\ p(\theta \mid y_{(1)}, X)\ d\theta. \tag{4.3}$$

As was to be expected from (3.4), the general elliptical character of the data density does not induce any difference in our predictive analysis with respect to the Normal framework. Remark that integrating out r^2 in (3.3), under the prior (2.7), always leads to a density of y, given θ and X, proportional to $d(y, X, \theta)^{-\frac{n}{2}}$, as was stressed in Section 3. This, of course, implies the Student density of $y_{(2)}$, given $y_{(1)}, \theta$ and X, but this Student t form will generally be lost when we integrate out θ in the predictive density as in (4.3).

A predictive analysis on the basis of (4.3) can be called for when e.g. the observations on $y_{(2)}$ are missing, whereas both $y_{(1)}$ and the entire X matrix are observed. However, in actual practice, it is often the case that only a submatrix of X, say X_1, is jointly observed with $y_{(1)}$, so that the posterior information available for forecasting is only based on $y_{(1)}$ and X_1. We then set out to predict $y_{(2)}$, given the observed $(y_{(1)}, X_1)$ and a set of exogenously given values for the remaining part of X, say X_2. Given our maintained assumption of independence between X and ω, it is sufficient to assume

$$h_{(1)} = h_{(1)}(X_1, \beta) \quad \text{and} \quad V_{11} = V_{11}(X_1, \eta) \tag{4.4}$$

in order to have posterior independence between θ and X_2.

Theorem 3: Under the conditions of Theorem 1 and (4.4) any elliptical sampling model (2.1) will allow conditional forecasting based on the predictive density

$$\begin{aligned} p(y_{(2)} \mid y_{(1)}, X) = \int_\Theta & f_S^{n_2}(y_{(2)} \mid n_1,\ h_{(2)} + V_{21}V_{11}^{-1}(y_{(1)} - h_{(1)}), \\ & \frac{n_1}{a(y_{(1)}, X_1, \theta)} V_{22.1}^{-1})\ p(\theta \mid y_{(1)}, X_1)\ d\theta, \end{aligned} \tag{4.5}$$

and $p(\theta \mid y_{(1)}, X_1)$ is obtained from (4.2) but now with (4.4) holding. ∎

The improper prior on τ^2 in (2.7) and the existence of the posterior thus lead to perfect predictive robustness which can be used in practice under assumption (4.4).

In the simplest linear case with a uniform prior on $\theta = \beta$, we can write:

Corollary 2: If $h(X, \beta) = X\beta,\ V(X, \eta) = V$ is assumed known and $\Theta = \mathbb{R}^k$, then under a uniform prior on $\theta = \beta$ the predictive (4.5) in Theorem 3 reduces to the Student density

$$p(y_{(2)} \mid y_{(1)}, X) = f_S^{n_2}(y_{(2)} \mid n_1 - k,\ X_2\hat{\beta}_1 + V_{21}V_{11}^{-1}(y_{(1)} - X_1\hat{\beta}_1),\ s_1^{-2}W^{-1}) \tag{4.6}$$

with

$$\begin{aligned} \hat{\beta}_1 &= (X_1'V_{11}^{-1}X_1)^{-1}X_1'V_{11}^{-1}y_{(1)} \\ s_1^2 &= \frac{1}{n_1 - k}\ (y_{(1)} - X_1\hat{\beta}_1)'\ V_{11}^{-1}(y_{(1)} - X_1\hat{\beta}_1) \\ W &= (X_2 - V_{21}V_{11}^{-1}X_1)(X_1'V_{11}^{-1}X_1)^{-1}(X_2 - V_{21}V_{11}^{-1}X_1)' + V_{22.1} \end{aligned}$$

and $V_{22.1}$ defined as in (4.1). ∎

A uniform prior of β has to be used for obtaining the Student predictive in (4.6), since we have left the class of prior densities that are natural conjugate for the Normal case (2.4) by assuming that $p(\theta, \nu_2)$ does not depend on τ^2 in (2.7).

5. FINITE MIXTURES OF ELLIPTICAL DATA DENSITIES

Although the class of sampling models described in (2.1) and (2.6) can already cover many cases used in practical applications, it still forces us to choose one particular functional form for the location vector and the covariance structure. If we wish to consider various alternatives, we can use finite mixtures of data densities as in (2.1), with (2.6) holding. In case we would allow for repeated sampling from such a finite mixture, the assumption of symmetry, inherent to elliptical densities, can be circumvented. However, Bernardo and Girón (1988) show that this seriously complicates the analysis. Therefore, we restrict ourselves here to the case of one vector observation which can come from any of the elements in the mixture.

Finite mixtures of conjugate prior densities were used to approximate more general classes of priors in Dalal and Hall (1983) and Diaconis and Ylvisaker (1985), but here we introduce the mixing in the sampling model instead. This, of course, widens the family of data densities we can accommodate, and, in principle, $p(\theta, \nu)$ can also involve prior mixtures in our framework, although the latter point will not be elaborated here. We feel it is important to allow for a large enough class of sampling models, since the likelihood is (too) often felt to have some "external validity" [see Berger (1985, p. 249)], and therefore not questioned, whereas we "agree to disagree" on the formulation of the prior. In the terminology of Poirier (1988, p. 130) the "window" entertained should be large enough to interest a "sizeable audience of like-minded researchers". For practical purposes, we only consider finite mixtures. Assessment methods for such mixtures are found in Dickey and Chen (1985, Section 5), based on elicited quantiles. We are not really treating these mixtures in a model selection context, as we see no reason to choose any particular elliptical submodel (pretest). We would rather find it natural to average out our uncertainty over models [see also Poirier (1991) and Chib *et al.* (1990b)].

If we suitably extend $\theta = (\beta, \eta)$ and ν_2 to parameterize a finite number of densities as in (2.1), each of which has the same scalar precision parameter τ^2, we still have to introduce a mixing parameter λ. Let us, more in detail, analyse the case where we mix only two elliptical densities, implying that λ is scalar. The relevant sampling model becomes

$$p(y \mid X, \omega, \lambda) = \lambda \mid \frac{1}{\tau^2} V(X, \eta) \mid^{-\frac{1}{2}} g_{n,\nu}[(y - h(X, \beta))'(\frac{1}{\tau^2} V(X, \eta))^{-1}(y - h(X, \beta))]$$
$$+ (1 - \lambda) \mid \frac{1}{\tau^2} W(X, \eta) \mid^{-\frac{1}{2}} k_{n,\nu}[(y - m(X, \beta))'(\frac{1}{\tau^2} W(X, \eta))^{-1}(y - m(X, \beta))],$$
$$0 \le \lambda \le 1, \tag{5.1}$$

where both $g_{n,\nu}(\cdot)$ and $k_{n,\nu}(\cdot)$ satisfy condition (2.2), and $m(\cdot)$ and $W(\cdot)$ are known functions in $\mathbb{R}^n$ and the space of all $n \times n$ PDS matrices, respectively. The nuisance parameter τ^2 does not index either of the functions $g_{n,\nu}(\cdot)$ and $k_{n,\nu}(\cdot)$, since we assume the improper prior structure, integrable in ν_2 over N_2:

$$p(\omega, \lambda) = \frac{c}{\tau^2} \, p(\theta, \nu_2, \lambda). \tag{5.2}$$

As in Section 3, this results in a joint density of $(y, \theta, \nu_2, \lambda \mid X)$ that no longer involves the functions $g_{n,\nu}(\cdot)$ or $k_{n,\nu}(\cdot)$. We obtain

$$p(y, \theta, \nu_2, \lambda \mid X) = c \, \Gamma(\frac{n}{2}) \, \pi^{-\frac{n}{2}} \, p(\theta, \nu_2, \lambda) \, [\lambda a_\theta + (1 - \lambda) b_\theta],$$

where we have defined

$$a_\theta = \mid V(X, \eta) \mid^{-\frac{1}{2}} d(y, X, \theta)^{-\frac{n}{2}}$$
$$b_\theta = \mid W(X, \eta) \mid^{-\frac{1}{2}} [(y - m(X, \beta))' W(X, \eta)^{-1} (y - m(X, \beta))]^{-\frac{n}{2}}.$$

Under the prior in (5.2), mixing *any* elliptical data densities with common τ^2 has the same consequences for both posterior [on (θ, λ)] and predictive inference as the mixing of Normals. In particular, if the joint density of $(y, \theta, \lambda \mid X)$ is integrable in (θ, λ), the posterior of (θ, λ) will be

$$p(\theta, \lambda \mid y, X) \propto p(\theta, \lambda) \, [\lambda a_\theta + (1 - \lambda) b_\theta], \tag{5.3}$$

whereas the prior density

$$p(\theta, \lambda) = \int_{N_2} p(\theta, \nu_2, \lambda) \, d\nu_2$$

must be at least integrable in those elements of θ that appear in only one of the mixed densities in (5.1), due to the summation character of mixtures. Note that the assumption of a common τ^2 is not restrictive. If instead we assume that τ^2 is multiplied by any positive scalar function of (X, η) that varies over models, the density $p(y, \theta, \nu_2, \lambda \mid X)$ remains unaffected under (5.2). The condition that η is not functionally related to τ^2, implicit in (5.2), is what really matters. The posterior density in (5.3) is a generalization of (3.6), which it reduces to for $\lambda = 1$. For nondegenerate λ, however, the mixing in the data density (5.1) is carried over to the posterior. A convenient choice for the prior of λ may be a beta density, independent of θ, i.e.

$$p(\theta, \lambda) = p(\theta)\ f_B(\lambda \mid p, q) \tag{5.4}$$

with $0 \leq \lambda \leq 1$ and $p,\ q > 0$.

From (5.3) we then obtain the conditional posterior of λ as a mixture of beta densities

$$p(\lambda \mid \theta, y, X) = (pa_\theta + qb_\theta)^{-1}\ [pa_\theta\ f_B(\lambda \mid p+1, q) + qb_\theta\ f_B(\lambda \mid p, q+1)]. \tag{5.5}$$

The marginal posterior density of θ will be given by

$$p(\theta \mid y, X) \propto p(\theta)\ (pa_\theta + qb_\theta),$$

which can be written as the following mixture of the "individual" posteriors, each calculated as in (3.6) on the basis of one of the elliptical models in (5.1):

$$p(\theta \mid y, X) = (\alpha K_a + (1-\alpha)K_b)^{-1}\ [\alpha K_a p_a(\theta \mid y, X) + (1-\alpha)K_b p_b(\theta \mid y, X)], \tag{5.6}$$

where $\alpha = E(\lambda) = \frac{p}{p+q}$, and

$$p_a(\theta \mid y, X) = K_a^{-1}\ p(\theta)\ a_\theta$$
$$p_b(\theta \mid y, X) = K_b^{-1}\ p(\theta)\ b_\theta.$$

The marginal posterior of θ depends on the prior of λ only through its mean. If, as in Chib *et al.* (1990b), we interpret λ as the prior probability of the first model given λ, then $\alpha = E(\lambda)$ is the marginal prior probability that this model generates the observation. Marginal posterior model probabilities are then given by the weights in (5.6), namely $\frac{\alpha K_a}{\alpha K_a + (1-\alpha)K_b}$ and $\frac{(1-\alpha)K_b}{\alpha K_a + (1-\alpha)K_b}$.[4] The latter and, more generally, all results on (y, ω) given X are only affected by the prior mean of λ. Of course, (5.6)

[4]These are *not* the posterior means of λ. In the extreme case $a_\theta = 0$ the posterior mean of λ is $E(\lambda \mid \theta, y, X) = E(\lambda \mid y, X) = \frac{p}{p+q+1}$, whereas the posterior probability of the corresponding model becomes 0. Only if p and q tend to zero will the posterior mean of λ tend to the relevant model probability [see Chib *et al.* (1990b)].

reduces to (3.6) for $q = 0$, in which case (5.4) groups all the prior mass at the point $\lambda = 1$.

From the posterior density in (5.3) it becomes apparent that, unless the functional forms of $h(\cdot)$ and $m(\cdot)$ or those of $V(\cdot)$ and $W(\cdot)$ differ,[5] the mixing in (5.1) will not affect the inference at all. Indeed, then the posterior of λ in (5.5) will reduce to the beta density in the prior (5.4) as $a_\theta = b_\theta$, and the posterior of θ will be the same as (3.6) in Section 3.

Let us now generalize the main results of this section to mixtures of $\ell > 2$ proper elliptical densities. We shall retain the improper prior as in (5.2) for the common nuisance parameter τ^2, but λ will now be of dimension ℓ, and we shall, therefore, generalize the beta prior in (5.4) to a Dirichlet prior on λ, with the parameter vector $S\alpha = S(\alpha_1 \ldots \alpha_\ell)'$, $\alpha_i > 0$, $\forall\, i$; $\sum_{i=1}^{\ell} \alpha_i = 1, S > 0$:

$$p(\lambda \mid \theta) = p(\lambda) = f_D^\ell(\lambda \mid S\alpha), \tag{5.7}$$

where λ is restrained to the set $\{\lambda \in \mathbb{R}^\ell : \lambda_i > 0,\ \forall\, i;\ \sum_{i=1}^{\ell} \lambda_i = 1\}$. Analogously to a_θ and b_θ in the case $\ell = 2$, we define c_θ^i for the i^{th} density in the mixture, and we denote by e^i the ℓ-dimensional vector with one in the i^{th} position and zeros elsewhere. Then we can state:

Theorem 4: Finite mixtures of ℓ elliptical densities, i.e. an obvious extension of (5.1), with common nuisance parameter τ^2 on which the improper prior (5.2) is defined, will, under (5.7), lead to

$$p(\lambda \mid \theta, y, X) = \left(\sum_{i=1}^{\ell} \alpha_i\, c_\theta^i\right)^{-1} \left[\sum_{i=1}^{\ell} \alpha_i\, c_\theta^i\, f_D^\ell(\lambda \mid S\alpha + e^i)\right] \tag{5.8}$$

and

$$p(\theta \mid y, X) = \left(\sum_{i=1}^{\ell} \alpha_i\, K_i\right)^{-1} \left[\sum_{i=1}^{\ell} \alpha_i\, K_i\, p_i(\theta \mid y, X)\right], \tag{5.9}$$

where $p_i(\theta \mid y, X) = K_i^{-1} p(\theta)\, c_\theta^i$, $\forall\, i$, provided all these posterior densities are well defined. ■

[5] If $h(\cdot) = m(\cdot)$, then $V(\cdot)$ and $W(\cdot)$ should differ by more than just a multiplicative scalar. If both are proportional and $h(\cdot) = m(\cdot)$, then $a_\theta = b_\theta$ and we are still in the elliptical class. A special case of this would be the scale contaminated Normal distribution mentioned in Johnson (1987, p. 123).

Since the posterior results in (5.8) and (5.9) are also finite mixtures, their analysis is not more difficult than with a single elliptical sampling density. Just like in the previous section, prediction can also be based on mixed sampling models, now using the posterior densities for both λ and θ. Again, we end up with a mixture, as formally stated in the final theorem.

Theorem 5: Under the conditions of Theorem 4, we can base our predictions for a finitely mixed elliptical model on the predictive density

$$p(y_{(2)} \mid y_{(1)}, X) = \left(\sum_{i=1}^{\ell} \alpha_i \; L_i\right)^{-1} \left[\sum_{i=1}^{\ell} \alpha_i \; L_i \; p_i(y_{(2)} \mid y_{(1)}, X)\right], \tag{5.10}$$

which is itself a mixture of

$$p_i(y_{(2)} \mid y_{(1)}, X) = \int_{\Theta} p_i(y_{(2)} \mid y_{(1)}, X, \theta) \; p_i(\theta \mid y_{(1)}, X) \; d\theta,$$

where $p_i(y_{(2)} \mid y_{(1)}, X, \theta)$ is the Student t density in (4.1) now corresponding to the i^{th} data density in the mixture, and

$$p_i(\theta \mid y_{(1)}, X) = L_i^{-1} \; p(\theta) \mid V_{11}^i \mid^{-\frac{1}{2}} a_i(y_{(1)}, X, \theta)^{-\frac{n_1}{2}},$$

as in (4.2), where each L_i must be finite, and indices i refer to the i^{th} data density throughout. ∎

As in Section 4, if we wish to use posterior densities for θ, computed after observing $y_{(1)}$ and only part of X, namely X_1, we need a bit more. Imposing condition (4.4) on every data density that is used in the sampling model will be sufficient.

We suggest treating specification uncertainty by such finite mixtures of elliptical densities, since the mixing will be preserved in both posterior and predictive analyses. We thus have a way of considerably broadening the class of data densities, without really adding to the complexity of the analysis.

6. CONCLUDING REMARKS

Under certain conditions, it was shown that Bayesian posterior and predictive analysis is perfectly robust with respect to the choice of a sampling density within the entire class of elliptical densities. Sufficient conditions are that we can single out a scale factor τ^2 on which we can specify an improper prior density.

Once the scale factor is then integrated out, the tails of the sampling density do not matter anymore, only the location and shape of the ellipsoids, parameterized by θ, are relevant. The posterior of θ will then be given by the simple expression in Theorem 1, which is the same as in the Normal case. The only purpose of the parameter ν_2 is to describe the tails of the data density. Thus, if the latter become irrelevant, then, clearly, the sample can not directly revise our opinion about ν_2. It can only do so through revising θ if there is prior dependence between θ and ν_2. This is the object of Theorem 2.

Our conclusions are similar for prediction: given an improper prior on the nuisance parameter τ^2, everything is just like in the Normal regression model. Theorem 3 summarizes these findings.

If the choice of a single elliptical data density is found to be too restrictive, we can make use of finite mixtures of elliptical densities to average out over specification uncertainty. These mixtures are then carried over to posterior and predictive results, without leading to an increase in complexity (see Theorems 4 and 5, respectively). Note that the contenders have to correspond to different ellipsoids, e.g. through different functional form or choice of regressors. Mixing e.g. a Normal and a Cauchy defined over the same ellipsoid will, of course, give the same results as with a single Normal data density.

The findings in this paper generalize and explain several results that have appeared in the literature,[6] and give remarkably weak sufficient conditions for robustness with respect to the data density within the multivariate elliptical family.

[6]The results from Sections 1 through 4 can be related to previous work in this area; in particular, our paper extends the framework of scale mixtures of Normal densities, found in Jammalamadaka *et al.* (1987), Chib *et al.* (1988), Osiewalski (1991) and Chib *et al.* (1990a), to general elliptical densities. It also broadens the linear regression model, used in the first two of the above references, to a possibly nonlinear one. Taking into account that only a diffuse prior for τ^2 was considered in the present paper, we can establish the following correspondences. Within the class of scale mixtures of Normals, Proposition 1 of Jammalamadaka *et al.* (1987) is a special case of our Corollary 2 for $V(X,\eta) = I_n$, whereas Theorem 3 generalizes Proposition 1 of Chib *et al.* (1988), who assumed linearity and a uniform prior on β. Both Theorems 1 and 3 extend results obtained under scale mixtures of Normals in Osiewalski (1991) to general elliptical densities, and Theorem 2 generalizes Theorem 2 in Chib *et al.* (1990a) in the same way.

REFERENCES

Bagchi, P. and I. Guttman, 1988, Bayesian regression analysis under non-normal errors, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* 3 (Clarendon Press, Oxford), 543-552.

Berger, J.O., 1985, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York).

Bernardo, J.M. and F.J. Girón, 1988, A Bayesian analysis of simple mixture problems, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* 3 (Clarendon Press, Oxford), 67-78.

Box, G.E.P. and G.C.Tiao, 1973, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, Reading).

Cambanis, S., S. Huang and G. Simons, 1981, On the theory of elliptically contoured distributions, *Journal of Multivariate Analysis* 11, 368-385.

Chib, S., J. Osiewalski and M.F.J. Steel, 1990a, Posterior inference on degrees of freedom in multivariate-t regression models, CentER Discussion Paper 9043, Tilburg University, Tilburg.

Chib, S., J. Osiewalski and M.F.J. Steel, 1990b, Regression models under competing covariance structures: A Bayesian perspective, CentER Discussion Paper 9063, Tilburg University, Tilburg.

Chib, S., R.C. Tiwari and S.R. Jammalamadaka, 1988, Bayes prediction in regressions with elliptical errors, *Journal of Econometrics* 38, 349-360.

Dalal, S. and W.J. Hall, 1983, Approximating priors by mixtures of natural conjugate priors, *Journal of the Royal Statistical Society* B 45, 278-286.

Diaconis, P. and D. Friedman, 1986, On the consistency of Bayes estimates, *The Annals of Statistics* 14, 1-26.

Diaconis, P. and D.Ylvisaker, 1985, Quantifying prior opinion, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* 2 (North-Holland, Amsterdam), 133-156.

Dickey, P. and C.-H. Chen, 1985, Direct subjective-probability modelling using ellipsoidal distributions, in: J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds., *Bayesian Statistics* 2 (North-Holland, Amsterdam),157-182.

Drèze, J.H., 1977, Bayesian regression analysis using poly-t densities, *Journal of Econometrics* 6, 329-354.

Drèze, J.H. and J.-F. Richard, 1983, Bayesian analysis of simultaneous equation systems, Ch. 3 in: Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics* (North-Holland, Amsterdam).

Engle, R.F., 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* 50, 987-1007.

Florens, J.P. and M. Mouchart, 1985, Conditioning in dynamic models, *Journal of Time Series Analysis* 6, 15-34.

Florens, J.P., M. Mouchart and J.M. Rolin, 1990, *Elements of Bayesian Statistics* (M. Dekker, New York).

Girón, F.J., E. Caro and M.L. Martinez, 1989, Una generalización del teorema de Gauss-Markov, Proceedings of the XVIII National Reunion of Statistics, Operations Research and Informatics, Santiago de Compostelo.

Hill, B.M., 1969, Foundations for the theory of least squares, *Journal of the Royal Statistical Society* B 31, 89-97.

Jammalamadaka, S.R., R.C.Tiwari and S. Chib, 1987, Bayes prediction in the linear model with spherically symmetric errors, *Economics Letters* 24, 39-44.

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

Kelker, D., 1970, Distribution theory of spherical distributions and a location-scale generalization, *Sankhyā* A 32, 419-430.

Leamer, E.E., 1978, *Specification Searches* (Wiley, New York).

McGraw, D.K. and J.F. Wagner, 1968, Elliptically symmetric distributions, *IEEE Transactions on Information Theory* 14, 110-120.

Osiewalski, J., 1991, A note on Bayesian inference in a regression model with elliptical errors, *Journal of Econometrics* 48, 183-193.

Osiewalski, J. and M.F.J. Steel, 1991, Bayesian marginal equivalence of elliptical regression models, CentER Discussion Paper 9119, Tilburg University, Tilburg.

Poirier, D.J., 1988, Frequentist and subjectivist perspectives on the problems of model building in economics, *Economic Perspectives* 2, 121-144.

Poirier, D.J., 1991, A Bayesian view of nominal money and real output through a new classical macroeconomic window, *Journal of Business and Economic Statistics*, forthcoming.

Ramsay, J.O. and M.R. Novick, 1980, PLU robust Bayesian decision theory: point estimation, *Journal of the American Statistical Association* 75, 901-907.

Richard, J.-F. and H. Tompa, 1980, On the evaluation of poly-t density functions, *Journal of Econometrics* 12, 331-351.

Spanos, A., 1990a, Modeling dynamic heteroskedasticity using thick-tailed distributions: The Student's t and related linear models, mimeo, VPI and State University, Blackburg.

Spanos, A., 1990b, A parametric approach to heteroskedasticity: The Student's t and elliptical linear regression models, mimeo, VPI and State University, Blackburg.

van Praag, B.M.S. and B.M. Wesselman, 1989, Elliptical multivariate analysis, *Journal of Econometrics* 41, 189-203.

West, M., 1984, Outlier models and prior distributions in Bayesian linear regression, *Journal of the Royal Statistical Society* B 46, 431-439.

Zellner, A., 1976, Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, *Journal of the American Statistical Association* 71, 400-405.