# BAYESIAN MARGINAL EQUIVALENCE OF ELLIPTICAL REGRESSION MODELS

Jacek Osiewalski and Mark F.J. Steel

92 – 04

Universidad Carlos III de Madrid

# BAYESIAN MARGINAL EQUIVALENCE OF ELLIPTICAL REGRESSION MODELS[*]

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

Abstract_____

The use of proper prior densities in regression models with multivariate non-Normal elliptical error distributions is examined when the scale matrix is known up to a precision factor $\tau$, treated as a nuisance parameter. Marginally equivalent models preserve the convenient predictive and posterior results on the parameter of interest ß obtained in the reference case of the Normal model and its conditionally natural conjugate gamma prior. Prior densities inducing this property are derived for two special cases of non-Normal elliptical densities representing very different patterns of tail behavior. In a linear framework, so-called semi-conjugate prior structures are defined as leading to marginal equivalence to a Normal data density with a fully natural conjugate prior.

Key words: multivariate elliptical data densities, proper priors, model robustness, Student t density.

[**] Academy of Economics, Kraków, Poland.

[***] Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Spain and Department of Econometrics, Tilburg University, The Netherlands.

# 1. Introduction

Part of the literature in Bayesian econometrics has been directed towards broadening the distributional assumptions on the error terms of the multiple regression model. Zellner (1976) considered multivariate Student t errors and concluded that inference still remains relatively simple with diffuse priors. Jammalamadaka **et al.** (1987), Chib **et al.** (1988) and Osiewalski (1991) considered an error vector distributed as a scale mixture of Normals and stated that under certain improper prior assumptions both prediction and posterior inference is unaffected by such departures from Normality.

Under **improper** priors, these results were generalized to any multivariate elliptical data density in Osiewalski and Steel (1992). They showed that it suffices to single out a scalar precision factor $\tau$ on which we specify a Jeffreys' type prior to obtain full robustness within the **entire** family of multivariate elliptical sampling models. This robustness property holds for both predictive and for posterior results on the parameters other than $\tau$. If we, however, insist on using **proper** prior structures, the results of Zellner (1976) already suggest that such robustness no longer holds in general. The focus of this paper is on obtaining proper priors under which this robustness occurs with respect to **specific** non-Normal multivariate elliptical data densities.
The concept of robustness used here is defined with respect to the sampling model, and is referred to as "model robustness" in Berger (1985, p. 248) or "inference robustness" in Box and Tiao (1973, p. 152). Thus, it differs from robustness against extreme observations [see Ramsay and Novick (1980)] as well as from robustness with respect to the specification of the prior [see Berger (1985, p. 195-247) and Berger (1990) for a recent survey].

We consider two parametric families of sampling densities P and $P_*$ with the parameter of **interest** $\beta$ in common and different sets of **nuisance** parameters. Bayesian models from both families are called **marginally equivalent** if prior densities are such that they lead to the same posterior inference on $\beta$ and the same predictive inference. In particular, we take P to be the class of multivariate **elliptical** data densities with location parameter $\beta$ and nuisance parameter $\delta$ involving $\tau$. For $P_*$ we choose the usual Normal sampling model with the same location vector and with nuisance precision factor $\varphi$. A convenient reference prior $p_*(\varphi|\beta)$ is the conditional natural conjugate gamma density, and we

examine which (proper) priors on $\delta$, given $\beta$, make a non-Normal member of P marginally equivalent to the Normal model with gamma prior. For two leading examples, the **Student t** case and the **Pearson II** sampling density, such conditional priors of $\delta$ are derived in closed form.

In the **linear** regression model, an even more convenient prior structure is the particular Student-gamma form which is natural conjugate for both $\beta$ and $\varphi$ under Normality. For alternative linear elliptical sampling processes, any prior on $(\beta, \delta)$ that induces marginal equivalence with this Normal-natural conjugate model will be called **semi-conjugate**. Mimicking the behaviour of this most popular reference model is seen to imply some potentially severe restrictions.

Appendix A lists three different representations of a multivariate Student t density as a scale mixture. For convenience, the probability density functions used in the course of the paper, are grouped in appendix B.

## 2. Marginally equivalent Bayesian models

Consider a parametric family $P = \{p(y|\beta, \delta) : \beta \in B, \delta \in \Delta\}$ of probability densities for a vector observation y, where $\beta$ is a parameter of primary interest and $\delta$ is a nuisance parameter. Suppose also that there is another parametric family $P_* = \{p_*(y|\beta, \varphi) : \beta \in B, \varphi \in \Phi\}$ for y in which $\beta$ plays the same role as in P (e.g., $\beta$ is a location parameter for both P and $P_*$), but $\varphi$ need not be linked to $\delta$ (even the dimensions can differ). Now consider the two Bayesian models, i.e. joint densities for observations and parameters,

$$p(y, \beta, \delta) = p(y|\beta, \delta)p(\beta, \delta), \tag{2.1}$$

$$p_*(y, \beta, \varphi) = p_*(y|\beta, \varphi)p_*(\beta, \varphi), \tag{2.2}$$

where $p(\beta, \delta) = p(\beta)p(\delta|\beta)$ and $p_*(\beta, \varphi) = p_*(\beta)p_*(\varphi|\beta)$ are prior densities for the parameters of P and $P_*$, respectively.

If $\beta$ has the same interpretation in both parametric families, we can regard it as having a reality independent of the choice of P or $P_*$, and thus we will naturally require that the marginal prior density of $\beta$ does not depend on the choice of the sampling family, i.e.

$$p(\beta) = p_*(\beta). \tag{2.3}$$

Suppose that $p_*(y,\beta,\varphi)$ has a particularly convenient form. The issue is then whether we can use this convenient model for Bayesian marginal posterior inference about $\beta$ and for predictive inference when y comes from a density in P rather than a density in $P_*$. For this to be valid, the Bayesian models in (2.1) and (2.2) must be **marginally equivalent** for y and $\beta$, i.e. integrating out $\delta$ in (2.1) and $\varphi$ in (2.2) should lead to the same density of y and $\beta$

$$p(y,\beta) = p_*(y,\beta). \tag{2.4}$$

Under (2.3), (2.4) reduces to the requirement that the marginalized likelihoods

$$p(y|\beta) = \int_\Delta p(y|\beta,\delta)p(\delta|\beta)d\delta \tag{2.5}$$

and

$$p_*(y|\beta) = \int_\Phi p_*(y|\beta,\varphi)p_*(\varphi|\beta)d\varphi \tag{2.6}$$

be identical, i.e. $p(y|\beta) = p_*(y|\beta)$.

In this paper we assume that P is the class of n-variate non-Normal elliptical data densities with location vector $h(X,\beta)$ and scale $\tau^{-1}V$, redefining $\delta$ as $\delta = (\tau,\nu)$

$$p(y|\beta,\delta) = f_g^n(y|h(X,\beta),\tau^{-1}V) = |V|^{-\frac{1}{2}}\tau^{\frac{n}{2}}g\{[y-h(X,\beta)]'\tau V^{-1}[y-h(X,\beta)]\},$$
$$\tag{2.7}$$

where $g(.)$ is a known nonnegative function indexed by $\nu$ such that $u^{\frac{n}{2}-1}g(u)$ is integrable in u over $R_+$. The latter requirement is shown in e.g. Kelker (1970), Dickey and Chen (1985) and Fang **et al.** (1990) to be necessary and sufficient for properness of (2.7). $g(.)$ essentially controls tail behaviour. Non-normality of (2.7) means that $g(u)$ is not exponential in $-u$, $u \in R_+$. We also assume throughout the paper that the reference family $P_*$ consists of the Normal densities

$$p_*(y|\beta,\varphi) = f_N^n(y|h(X,\beta),\varphi^{-1}V) =$$

$$= (2\pi)^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \varphi^{\frac{n}{2}} \exp\{- \frac{\varphi}{2} [y-h(X,\beta)]'V^{-1}[y-h(X,\beta)]\}. \qquad (2.8)$$

Thus, in the case considered in the paper, $\tau$ and $\varphi$ are positive scalar parameters (precision parameters) and $\nu \in N$ which may be empty. In both (2.7) and (2.8), V is a known $n \times n$ PDS matrix, and h is a known vector function of the matrix X and of $\beta \in B \subseteq R^k$. In line with the existing literature, prompted by Zellner (1976), we focus in particular upon the regression context, but, of course, h does not have to involve exogenous variables. In any case, it should be stressed that deviations from Normality in the class P necessarily imply that we treat y as a single vector observation. For independent sampling from non-Normal elliptical distributions, the posterior and predictive analysis becomes far more complicated and marginal equivalence seems excluded. The regression models condition on X which is independent of all the parameters in the implied conditional models. A convenient concept that ensures the latter condition is a Bayesian cut [see Florens and Mouchart (1985)]. As conditioning on X will be maintained throughout the analysis, it will not be explicited in the notation. Remark that, in (2.7) and (2.8), the location of ellipsoids is entirely determined by $\beta$ (given X), which has an unambiguous interpretation, irrespective of the parametric family we choose.

Definition 1: any elliptical sampling model from P in (2.7) together with a prior on the nuisance parameter $p_g(\delta|\beta)$ is **marginally equivalent** to a Normal model from $P_*$ in (2.8) with the prior $p_*(\varphi|\beta)$ if, under (2.3), the marginalized likelihoods $p(y|\beta)$ and $p_*(y|\beta)$ in (2.5) and (2.6) coincide. □

An important example of marginal equivalence of Bayesian models is given by Osiewalski and Steel (1992) who show that the data density (2.7) and the improper prior structure

$$p(\beta,\delta) = p(\tau)p(\beta,\nu) = c\tau^{-1}p(\beta,\nu), \quad \tau \in R_+, \quad \beta \in B, \quad \nu \in N, \qquad (2.9)$$

where c is any positive constant and $p(\beta,\nu)$ is integrable in $\nu$ over N, lead to the marginalized likelihood

$$p(y|\beta) = c\Gamma(\frac{n}{2})\pi^{-\frac{n}{2}} |V|^{-\frac{1}{2}} \{[y-h(X,\beta)]'V^{-1}[y-h(X,\beta)]\}^{-\frac{n}{2}} \qquad (2.10)$$

for any g(.). Note that (2.10) is the same as the marginalized likelihood obtained from (2.8) under the prior structure

$$p_*(\beta,\varphi) = p_*(\varphi)p(\beta) = c\varphi^{-1}p(\beta), \quad \varphi \in R_+, \quad \beta \in B. \tag{2.11}$$

Since for the Bayesian models (2.7), (2.9) and (2.8), (2.11) marginal equivalence holds for any g(.) in P, Osiewalski and Steel (1992) arrive at **robustness** of posterior and predictive results with respect to departures from Normality within a broad class of multivariate elliptical data densities.

In this paper we are also looking for marginal equivalence of the Bayesian models with sampling densities (2.7) and (2.8), but under **proper** prior densities $p_g(\delta|\beta)$. As could be expected, the results will be much more modest than under the improper priors in (2.9) and (2.11).

## 3. Marginal equivalence under proper priors on $\delta$ given $\beta$

Consider the two alternative sampling families P and $P_*$ in (2.7) and (2.8). **Given** $\beta$, the precision parameter $\varphi$ of the Normal data density (2.8) is now assigned the very convenient natural conjugate gamma prior

$$p_*(\varphi|\beta) = f_G(\varphi|\frac{a}{2}, \frac{d_\beta}{2}), \tag{3.1}$$

where a is a positive constant, and $d_\beta$ is a known positive function of $\beta$ ($d_\beta$ may be a constant function d). Our interest is in a conditional prior $p_g(\delta|\beta)$ of $\delta$ in (2.7) such that the Bayesian models

$$p(y,\beta,\delta) = f_g^n(y|h(X,\beta),\tau^{-1}V)p_g(\delta|\beta)p(\beta) \tag{3.2}$$

and

$$p_*(y,\beta,\varphi) = f_N^n(y|h(X,\beta),\varphi^{-1}V)f_G(\varphi|\frac{a}{2}, \frac{d_\beta}{2})p(\beta) \tag{3.3}$$

are marginally equivalent. If such a prior $p_g(\delta|\beta)$ exists, it must solve the equation

$$\int_\Delta f_g^n(y|h(X,\beta),\tau^{-1}V)p_g(\delta|\beta)d\delta = f_S^n(y|a, h(X,\beta), \frac{a}{d_\beta}V^{-1}), \tag{3.4}$$

6

where the RHS, the marginalized likelihood from (3.3), is an n-variate Student t as defined in Appendix B. For non-Normal elliptical densities, this $p_g(\delta|\beta)$ will generally differ from $p_*(\wp|\beta)$ in (3.1), and will certainly not always exist in closed form from (3.4).

In the following subsections we give two very different examples of tail behaviour where closed form solutions of $p_g(\delta|\beta)$ can easily be derived.

## 3.1. The multivariate t data density

Consider the standard n-variate Student t data density with $\nu$ degrees of freedom:

$$f_g^n(y|h(X,\beta),\tau^{-1}V) = f_S^n(y|\nu, h(X,\beta), \tau V^{-1}),\qquad(3.5)$$

where $\nu \in R_+$. Tails are now thicker than in the Normal case (2.8) and moments exist up to (not including) $\nu$.

Proposition 1: the Student sampling model in (3.5) combined with the following conditional prior:

$$p_g(\delta|\beta) = p_g(\tau|\nu,\beta)p_g(\nu|\beta) = f_B(\tau|\tfrac{a}{2}, \tfrac{\nu-a}{2}, \tfrac{\nu}{d_\beta})p_g(\nu|\beta)\qquad(3.6)$$

where $p_g(\nu|\beta)$ is a proper density which is zero for $\nu \le a$, and with any (possibly improper) $p(\beta)$, is marginally equivalent with $p_*(y,\beta,\wp)$ in (3.3).

Proof: it suffices to check (3.4). In particular, we obtain

$$p(y|\beta) = \int_\Delta f_S^n(y|\nu, h(X,\beta), \tau V^{-1}) \, f_B(\tau|\tfrac{a}{2}, \tfrac{\nu-a}{2}, \tfrac{\nu}{d_\beta})d\tau \, p_g(\nu|\beta)d\nu.$$

$$(3.7)$$

Applying (A.2) in Appendix A to the inner integral with $q = \nu-a$ and $z = \nu/d_\beta$, leaves us with

$$p(y|\beta) = f_S^n(y|a, h(X,\beta), \tfrac{a}{d_\beta} V^{-1}) \int_N p_g(\nu|\beta)d\nu,\qquad(3.8)$$

so that $\nu$ can only be updated through its prior links with $\beta$ and (3.4) is seen to hold given the properness of $p_g(\nu|\beta)$. □

The beta prior in (3.6) restricts the parameter space of $\tau$ to $(0,\nu/d_\beta)$. For marginal inference on $(y,\beta)$ this exactly compensates the influence of the heavy tails in (3.5) and leads to the same results as in the Normal model with gamma prior $p_*(\varphi|\beta)$. By choosing a to be a small enough positive number, we can span almost the entire Student t class, even including Cauchy densities ($\nu = 1$). The parameter $\nu$, which does not appear in the Normal model, can never be updated "directly" in the marginally equivalent model. See Chib **et al.** (1992) for a discussion of inference on $\nu$. The conditional posterior density of $\tau$ given $\beta$ and $\nu$, however, does not retain the form of the prior, but can be written as

$$p_g(\tau|\beta,\nu,y) \propto f_B(\tau|\frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta}) f_{IB}(\tau|\frac{\nu}{2} - 1, \frac{n}{2} + 1, \frac{\nu}{s_\beta}), \qquad (3.9)$$

where

$$s_\beta = [y-h(X,\beta)]'V^{-1}[y-h(X,\beta)]. \qquad (3.10)$$

Prior independence between $\beta$ and $\varphi$ in (3.3) amounts to taking $d_\beta = d$, a positive constant, and renders the beta prior in (3.6) independent of $\beta$. As $a \to 0$ and $d \to 0$, the kernel of this beta prior becomes proportional to $\tau^{-1}$ (for $\tau \in \mathbb{R}_+$), and the Student t marginalized likelihood in (3.8) becomes proportional to (2.10). Finally, as $\nu \to \infty$ the sampling models in (2.8) and (3.5) become indistinguishable and the prior of $\tau$ in (3.6) indeed tends to the gamma prior of $\varphi$ found in (3.1).

## 3.2. The multivariate Pearson II data density

Instead of the Student tails, which are thicker than the reference Normal ones, let us now consider a case with truncated tails. In particular, we choose the multivariate Pearson Type II distribution [see e.g. Fang **et al.** (1990)], defined in Appendix B, where g(u) is nonzero for u ∈ [0,· and zero otherwise. Then (2.7) becomes

$$f_g^n(y|h(X,\beta), \tau^{-1}V) = f_{PII}^n(y|\nu, h(X,\beta), \tau^{-1}V). \qquad (3.11)$$

The data density (3.11) spreads the probability mass over the ellipsoid $\{y \in \mathbb{R}^n : \tau s_\beta < 1\}$, where $s_\beta$ was defined in (3.10), and thus possesses no tails. In this subclass of P the parameter $\nu \in \mathbb{R}_+$, as in Subsection 3.1.

Proposition 2: the Pearson II sampling model in (3.11) combined with the following conditional prior:

$$p_g(\delta|\beta) = f_{IB}(\tau|\frac{n+\nu}{2}, \frac{a}{2}, d_\beta^{-1})p_g(\nu|\beta), \qquad (3.12)$$

where $p_g(\nu|\beta)$ is a proper density, and with any $p(\beta)$, is marginally equivalent with $p_*(y,\beta,\varphi)$ in (3.3).

Proof: Straightforward use of (A.3) in Appendix A with $q = \nu$ and $z = d_\beta^{-1}$ directly results in the Student t marginalized likelihood $p_g(y|\beta,\nu)$ which no longer depends on $\nu$ and coincides with (3.4). Therefore, like in the previous case, updating $\nu$ can only be done through its prior dependence on $\beta$. □

As the data density itself is now restricted, there is no need to restrict the parameter space in order to obtain the reference result $p_*(y,\beta)$. Indeed, the tail of (3.12) is even much thicker than for its gamma counterpart in the Normal case. Note that the inverted beta prior in (3.12) will be truncated by the sampling model to give the following posterior:

$$p_g(\tau|\beta,\nu,y) \propto f_{IB}(\tau|\frac{n+\nu}{2}, \frac{a}{2}, d_\beta^{-1})f_B(\tau|\frac{n}{2} + 1, \frac{\nu}{2}, s_\beta^{-1}). \qquad (3.13)$$

Again, if we assume prior independence by taking $d_\beta = d$, the kernel of the inverted beta prior (3.12) approaches $\tau^{-1}$ (for $\tau \in \mathbb{R}_+$) as both a and d go to zero, and the marginalized likelihood becomes proportional to (2.10).

The posterior results in (3.9) and (3.13) clearly indicate that marginal equivalence holds for a single vector observation, but not for repeated sampling from either Student t or Pearson II densities.

## 4. Linear regression and semi-conjugate priors

Assume that the prior density in the reference Bayesian model (3.3) takes the form

$$p_*(\beta,\varphi) = f_G(\varphi|\frac{k+e}{2}, \frac{1}{2} [f + (\beta-\bar{\beta})'A(\beta-\bar{\beta})]) \ f_S^k(\beta|e,\bar{\beta},\frac{e}{f} A), \qquad (4.1)$$

where e and f are positive constants, $\bar{\beta}$ is a k×1 vector and A is a PDS k×k matrix. Note that $d_\beta$ defined implicitly in (4.1) is not constant in $\beta$, and thus precludes prior independence, and a = k+e is greater than the dimension of $\beta$. Of course, (4.1) is the well-known Student-gamma prior (or Normal-gamma in the alternative factorization), natural conjugate for both $\beta$ and $\varphi$ in the linear case, i.e. when h(X,$\beta$) = X$\beta$.

Definition 2: any prior density $p_g(\beta,\delta)$ which makes the Bayesian model

$$p(y,\beta,\delta) = f_g^n(y|X\beta, \tau^{-1}V)p_g(\beta,\delta) \qquad (4.2)$$

marginally equivalent to

$$p_*(y,\beta,\varphi) = f_N^n(y|X\beta, \varphi^{-1}V)p_*(\beta,\varphi) \qquad (4.3)$$

where $p_*(\beta,\varphi)$ is as in (4.1), will be called **semi-conjugate**. □

Semi-conjugate priors exactly preserve the simple Student t forms of the marginal prior, posterior and predictive densities of $\beta$ and y which are obtained from (4.3). Semi-conjugate priors for the types of data distributions considered in Subsections 3.1 and 3.2 can immediately be obtained by taking in (3.2), (3.6) and (3.12) the same a, $d_\beta$ and p($\beta$) as in (4.1). Only the Student t density case will be discussed in some detail. From (3.6) and (4.1) the prior

$$p_g(\beta,\delta) = f_S^k(\beta|e,\bar{\beta},\frac{e}{f} A) \ f_B(\tau|\frac{k+e}{2}, \frac{\nu-k-e}{2}, \frac{\nu}{f + (\beta-\bar{\beta})'A(\beta-\bar{\beta})})p_g(\nu|\beta) \qquad (4.4)$$

with $p_g(\nu|\beta)$ proper and zero for $\nu \leq$ k+e, is semi-conjugate for the Student t data density

$$f_g^n(y|X\beta,\tau^{-1}V) = f_S^n(y|\nu,X\beta,\tau V^{-1}) \qquad (4.5)$$

The form of the semi-conjugate prior (4.4) shows that severe restrictions must be put on the Bayesian model (4.2) with the data density (4.5) if we want to mimic (for marginal inference on $\beta$ and y) the convenient reference model (4.3).

Firstly, the implied marginal density of $\tau$ is nonzero over $(0,\nu/f)$ and the conditional densities of $\beta$, given $\nu$ and values of $\tau \in (0,\nu/f)$, are nonzero over ellipsoids $(\beta-\bar{\beta})'A(\beta-\bar{\beta}) < \frac{\nu}{\tau} - f$. Thus, an upper bound $\nu/f$ is put on the precision parameter $\tau$, and $\beta$ values far from the prior mean (in the metric induced by A) are allowed only for very small values of $\tau$, i.e. for noisy data processes, or for large values of $\nu$, i.e. densities close to Normality. Secondly, there is a lower bound on the degrees of freedom of the Student t sampling process in (4.5), namely $\nu > k+e$. Very thick (e.g. Cauchy) tails are ruled out. If the tails become too thick, i.e. if $\nu \leq k$, even restricting the parameter space of $(\beta,\delta)$ no longer suffices to obtain the same results for $p(y,\beta)$ as in (4.3). Even for very thick tails of (4.5), however, it is possible to obtain marginal equivalence (see Subsection 3.1) and thus mimic the Bayesian results for y and $\beta$, but then outside the natural conjugate framework (4.1). For example, by Proposition 1 the Normal data density under the Student-gamma prior,

$$p_*(y,\beta,\varphi) = f_N^n(y|X\beta,\varphi^{-1}V)f_S^k(\beta|e,\bar{\beta},A)f_G(\varphi|\frac{a}{2}, \frac{d_\beta}{2}), \tag{4.6}$$

is marginally equivalent to the Student t data density with Student-beta prior

$$p(y,\beta,\delta) = f_S^n(y|\nu,X\beta,\tau V^{-1})f_S^k(\beta|e,\bar{\beta},A)f_B(\tau|\frac{a}{2}, \frac{\nu-a}{2}, \frac{\nu}{d_\beta})p_g(\nu|\beta) \tag{4.7}$$

where $p_g(\nu|\beta)$ is proper and only nonzero for $\nu > a > 0$. Here we have at least one more free hyperparameter than in the semi-conjugate prior (4.4), and, therefore, $\nu$ need not be related to k. If we also take $d_\beta = d$ and $p_g(\nu,\beta) = p_g(\nu)$ we have prior independence between $(\tau,\nu)$ and $\beta$, and the only remaining restriction on the parameter space will be that $\tau \in (0,\nu/d)$, which will become less binding if $p_g(\nu)$ puts more mass on large values of $\nu$.

## 5. Conclusion

The use of Bayesian regression analysis in practice often relies on the Normal sampling model and its natural conjugate prior structure, since this leads to predictive and posterior densities with convenient properties. We ask whether the aspects which are typically of interest carry over to the general class of elliptical regression models. In particular, we examine the marginal equivalence for $(y,\beta)$ of non-Normal elliptical sampling models to the Normal

model with a convenient gamma prior on the precision factor $\varphi$, which is natural conjugate given $\beta$. For linear models, the specific prior structure that ensures marginal equivalence for $(y,\beta)$ under a fully natural-conjugate density of $(\beta,\varphi)$ in the Normal model is called semi-conjugate. The latter is of particular interest since it completely preserves the very convenient predictive and posterior results for $\beta$.

As already mentioned in Section 2, much stronger robustness results can be achieved if one allows for improper Jeffreys' type prior densities on the nuisance precision parameter $\tau$ of the elliptical model [see Osiewalski and Steel (1992)]. However, the analysis is similar in the sense that the difference between members of the elliptical class is entirely isolated in $\tau$. Under a Jeffreys' prior on $\tau$ only the conditional posterior on $\tau$ is affected by the choice of elliptical sampling model and the inference on $(y,\beta)$ is the same whatever the model chosen. The price to pay for restricting attention to proper prior families of the nuisance parameter $\delta = (\tau,\nu)$ is that the robustness results are more modest. Not just the posterior, but also the prior of $\delta$ given $\beta$ will now vary over elliptical models. A **specific** prior linked to a **particular** non-Normal elliptical model will exactly mimic the marginal results for $(y,\beta)$ that the natural conjugate prior structure produces with the Normal model. Differences in tail behaviour of the sampling model are entirely compensated by the properties of the conditional prior of $\delta$. The effect of heavy tails in the Student t case, for example, is neutralized by the beta form of $p_g(\tau|\beta,\nu)$, which restricts the parameter space. In the case of Pearson II densities with truncated tails, marginal equivalence requires an inverted beta density of $\tau$ given $(\beta,\nu)$ with a thicker tail than the natural conjugate gamma prior. So both prior and posterior distributions of $\tau$ given $(\beta,\nu)$ vary here with the choice of sampling model, allowing the rest of the analysis to remain unaffected.

## Appendix A. Representations of multivariate Student t densities

Any multivariate Student t density, as defined in Appendix B, can be represented as

(i)    a gamma scale mixture of Normal densities:

$$f_S^n(y|r,b,A) = \int_0^\infty f_N^n(y|b, \frac{r}{z}(\tau A)^{-1})f_G(\tau|\frac{r}{2}, \frac{z}{2})d\tau, \qquad (A.1)$$

for any z > 0,

(ii)  a beta scale mixture of Student t densities with larger degrees of freedom:

$$f_S^n(y|r,b,A) = \int_0^z f_S^n(y|r+q, \; b, \; \frac{r+q}{rz} \tau A) f_B(\tau|\frac{r}{2}, \; \frac{q}{2}, \; z) d\tau, \qquad (A.2)$$

for any q,z > 0,

(iii) a beta prime scale mixture of Pearson II densities:

$$f_S^n(y|r,b,A) = \int_0^\infty f_{PII}^n(y|q, \; b, \; rz(\tau A)^{-1}) f_{IB}(\tau|\frac{n+q}{2}, \; \frac{r}{2}, \; z) d\tau, \qquad (A.3)$$

for any q,z > 0, and where the mixand is only nonzero for
$\tau < rz/(y-b)'A(y-b)$.

The parameterizations of (A.1)-(A.3) were inspired by the specific use we make of these formulas, but they can clearly be simplified by convenient choices of z. Whereas representation (i) is widely known, (ii) and (iii) have, to the best knowledge of the authors, not yet appeared in the literature.
To prove (A.2), we use a variable transformation from $\tau$ to $\zeta = \frac{\tau}{r+q}/(1 - \frac{\tau}{z})$ and integrate out $\zeta$ using a beta prime density. A proof of (A.3) can be based on the transformation from $\tau$ to $\zeta = \frac{\tau}{z}/(1 + \frac{\tau}{z})$ which allows easy analytical integration of $\zeta$ through a beta density.

## Appendix B. Probability density functions

A k-variate Normal density on $x \in R^k$ with mean vector $b \in R^k$ and PDS k×k covariance matrix C:

$$f_N^k(x|b,C) = [(2\pi)^k|C|]^{-\frac{1}{2}} \exp \; \frac{1}{2} (x-b)'C^{-1}(x-b).$$

A k-variate Student t density on $x \in R^k$ with r > 0 degrees of freedom, location vector $b \in R^k$ and PDS k×k precision matrix A:

$$f^k_S(x|r,b,A) = \frac{\Gamma(\frac{r+k}{2})}{\Gamma(\frac{r}{2})(r\pi)^{k/2}} |A|^{\frac{1}{2}}[1 + \frac{1}{r}(x-b)'A(x-b)]^{-\frac{r+k}{2}} .$$

A k-variate Pearson II density defined on the support $E = \{x \in R^k : (x-b)'C^{-1}(x-b) < 1\}$ with $r > 0$, location vector $b \in R^k$ and PDS $k \times k$ scale matrix C:

$$f^k_{PII}(x|r,b,C) = \begin{cases} \dfrac{\Gamma(\frac{k+r}{2})}{\Gamma(\frac{r}{2})\pi^{k/2}} |C|^{-\frac{1}{2}}[1 - (x-b)'C^{-1}(x-b)]^{\frac{r}{2}-1} & \text{if } x \in E, \\[2em] 0 & \text{if } x \notin E. \end{cases}$$

A gamma density on $z > 0$ with $a,b > 0$:

$$f_G(z|a,b) = b^a[\Gamma(a)]^{-1}z^{a-1}\exp(-bz).$$

A beta density on $v \in (0,c)$ with $a,b > 0$:

$$f_B(v|a,b,c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} (\frac{v}{c})^{a-1}(1 - \frac{v}{c})^{b-1}.$$

A three-parameter inverted beta or beta prime density on $w > 0$ with $a,b,c > 0$ [see Zellner (1971, p. 376)]:

$$f_{IB}(w|a,b,c) = \frac{\Gamma(a+b)}{c\Gamma(a)\Gamma(b)} (\frac{w}{c})^{b-1}(1 + \frac{w}{c})^{-(a+b)} .$$

**References**

Berger, J.O., 1985, Statistical decision theory and Bayesian analysis (Springer-Verlag, New York).

Berger, J.O., 1990, Robust Bayesian analysis: Sensitivity to the prior, Journal of Statistical Planning and Inference 25, 303-328.

Box, G.E.P., and G.C. Tiao, 1973, Bayesian inference in statistical analysis, (Addison-Wesley, Reading).

Chib, S., J. Osiewalski, and M.F.J. Steel, 1992, Posterior inference on the degrees of freedom parameter in multivariate-t regression models, Economics Letters, forthcoming.

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

Dickey, J.M., 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.

Fang, K.-T., S. Kotz and K.W. Ng, 1990, Symmetric multivariate and related distributions (Chapman and Hall, London).

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

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.

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

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, 1992, Robust Bayesian inference in elliptical regression models, Journal of Econometrics, 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.

Zellner, A., 1971, An introduction to Bayesian inference in econometrics, (Wiley, New York).

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.