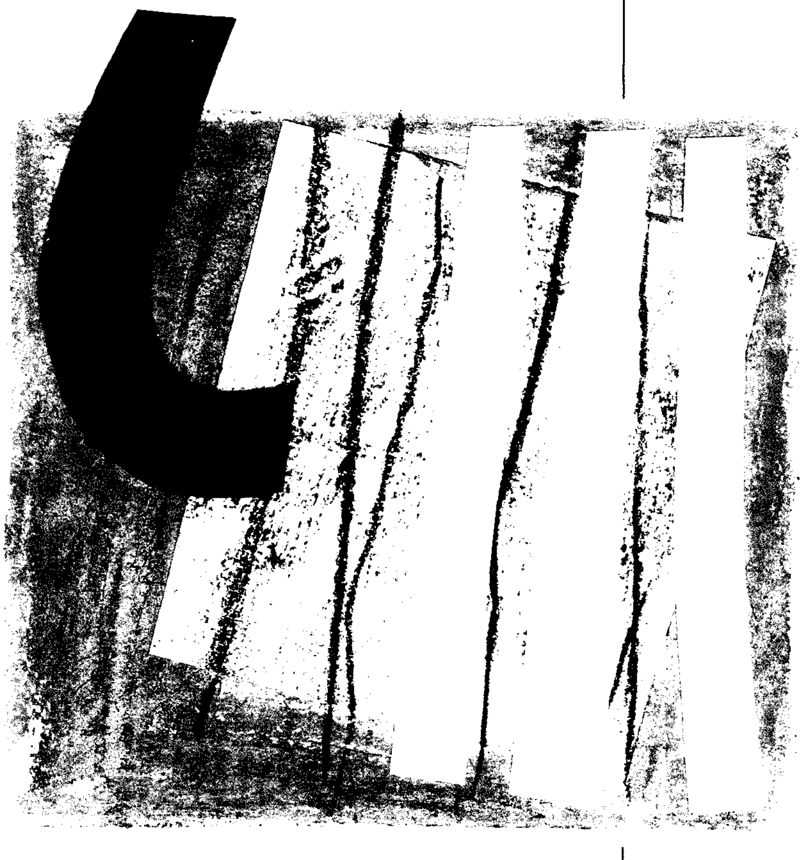
POSTERIOR MOMENTS OF SCALE PARAMETERS IN ELLIPTICAL REGRESSION MODELS

Jacek Osiewalski and Mark F.J. Steel

92 - 05



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 92-05
February 1992

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

POSTERIOR MOMENTS OF SCALE PARAMETERS IN ELLIPTICAL REGRESSION MODELS*

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

Abstract

In the general multivariate elliptical class of data densities we define a scalar precision parameter τ^2 through a normalization of the scale matrix V. Using the improper prior on τ^2 which preserves the results under Normality for all other parameters and prediction, we consider the posterior moments of τ^2. For the subclass of scale mixtures of Normals we derive the Bayesian counterpart to a sampling theory result concerning uniformly minimum variance unbiased estimation of τ^{-2}. If the sampling variance exists, we single out the common variance factor Ψ as the scalar multiplying V in this sampling variance. Moments of Ψ are examined for various elliptical subclasses and a sampling theory result regarding its unbiased estimation is mirrored.

Key words: multivariate elliptical data densities, Bayesian analysis, unbiased estimation.

* We benefited from useful discussions with Julián de la Horra, but, of course, the usual disclaimer applies. The first author acknowledges the hospitality of Tilburg University and the Universidad Carlos III de Madrid. The second author held a research fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW).

** 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

Much of the recent Bayesian work[1] on regression models with non-Normal data densities was prompted by Zellner (1976), who considered linear models with multivariate Student t error terms, a particular member of the elliptical family.

The general multivariate elliptical (or ellipsoidal) class of data densities, introduced in Section 2 of this paper, is characterized by ellipsoidal isodensity curves and the way the density function changes over the ellipsoids is indicated by a nonnegative labelling function $g_{n,\nu}(\cdot)$ indexed by a parameter ν of finite dimension. Location and shape of the ellipsoids are parameterized by θ, and we distinguish a scalar precision parameter τ^2 which is implicitly defined through a normalization of the scale matrix. Given θ, the volume of the ellipsoids is determined by τ^2. In Osiewalski and Steel (1992) it was shown that an improper reference prior on τ^2 combined with prior independence between τ^2 and (θ,ν) will preserve the "usual" posterior results on (θ,ν) that we find under Normality as well as the predictive results in the entire multivariate elliptical class. Such robustness with respect to the sampling model is only possible because the entire influence of the choice of $g_{n,\nu}(\cdot)$ is taken up by the conditional posterior distribution of τ^2, given the other parameters.

In this paper, we focus on the posterior properties of τ^2, in particular its moments and the moments of its inverse, σ^2. We find in Section 3 that these moments can, provided they exist, be expressed as a product of two factors: one data-dependent but not influenced by the particular choice of labelling function $g_{n,\nu}(\cdot)$, and the other depending on the form of $g_{n,\nu}(\cdot)$ but not on the data. In other words, the posterior moments of τ^2 or σ^2 under the multivariate elliptical sampling model and a commonly used improper prior are equal to the corresponding moments under Normality times some correcting factor, which completely captures the effect of the particular tail behaviour assumed and does not depend on the observed sample. This result holds given θ and ν, and is also shown to hold after marginalizing with respect to (θ,ν), under the additional assumption of prior independence between θ and ν (Proposition 1).

For some important subclasses of elliptical densities, the posterior moment of σ^2 are compared with those under Normality (Section 4), and it is found that a multivariate Student t density leads to

[1] We mention Jammalamadaka *et al.* (1987) and Chib *et al.* (1988) who consider scale mixtures of Normals, an elliptical subclass that contains the Student t case, whereas Osiewalski (1991) and Chib *et al.* (1992) generalize to nonlinear models. Osiewalski and Steel (1992) examine the entire family of multivariate elliptical densities.

the same posterior mean of σ^2 as obtained in the Normal case. The latter was noted in Zellner (1976) for linear spherical models.

Under the extra assumptions of linearity and an improper uniform prior on θ, Section 5 derives the Bayesian counterpart to a sampling theoretic result concerning uniformly minimum variance unbiased estimation of σ^2 found by Girón *et al.* (1989), in the context of scale mixtures of Normals. The latter is an important subclass of the elliptical family, with the Student t as a prominent member.

Section 6 is devoted to elliptical data densities that allow a finite sampling variance. The scalar factor that multiplies the normalized scale matrix in the sampling variance is then defined as the common variance factor Ψ. This Ψ is related to σ^2, but generally not equal to it. Its posterior moments are equal to the corresponding moments of σ^2 under Normality times a correcting factor, which is again not dependent on the observed data (Proposition 2). Interestingly, the inverse Ψ^{-1} always has the same mean as under Normality for the whole elliptical class with finite sampling variance (Corollary 2). This ties in directly with a sampling theory result on unbiased estimation of Ψ in linear models.

Section 7 concludes and an appendix groups some probability density functions used in the course of the paper.

2. THE ELLIPTICAL SAMPLING MODEL

In this paper we assume that the observation vector $y \in \mathbf{R}^n$ possesses a multivariate elliptical distribution around some location vector $h(X,\beta)$ with scale matrix $\tilde{V}(X,\tilde{\eta})$ and the density of y exists. The latter is then necessarily equal to

$$p(y|X,\omega) = |\tilde{V}(X,\tilde{\eta})|^{-1/2} g_{n,\nu}[(y-h(X,\beta))'\tilde{V}(X,\tilde{\eta})^{-1}(y-h(X,\beta))], \tag{2.1}$$

where the labelling function or density generator $g_{n,\nu}(\cdot)$ is indexed by n and ν and satisfies

$$\int_{\mathbf{R}_+} u^{\frac{n}{2}-1} g_{n,\nu}(u)\,du = \Gamma(\frac{n}{2})\pi^{-\frac{n}{2}}. \tag{2.2}$$

For general properties of such symmetric multivariate distributions we refer the reader to Kelker (1970), Cambanis *et al.* (1981), Dickey and Chen (1985) and Fang *et al.* (1990). The location vector in this regression context is a known function of a set of exogenous variables in X and a coefficient vector β of finite dimension. We reparameterize $\tilde{\eta}$ into (η, τ^2) such that

$$\tilde{V}(X,\tilde{\eta}) = \frac{1}{\tau^2} V(X,\eta), \tag{2.3}$$

where $\tau^2 \in \mathbf{R}_+$ is a scalar precision parameter and $V(X,\eta)$ is a normalized PDS scale matrix function of X and a finite parameter vector η. For notational convenience, we now define $\theta = (\beta,\eta) \in \Theta$, which summarizes all the information about location and shape of the ellipsoids.

Under (2.1)-(2.3) the squared radius r^2, which is equal to $r^2 = \tau^2 d(y,X,\theta)$ with

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

has the following density function:

$$p(r^2|X,\omega) = p(r^2|\nu) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}(r^2)^{\frac{n}{2}-1} g_{n,\nu}(r^2), \tag{2.4}$$

which was essentially proven in Kelker (1970) and can be found in Dickey and Chen (1985, p.161) and Fang *et al.* (1990, p.36).

3. THE PRIOR STRUCTURE AND POSTERIOR RESULTS

In Osiewalski and Steel (1992) it was shown that the improper prior

$$p(\omega) = p(\tau^2)p(\theta,\nu) = \frac{c}{\tau^2}\, p(\theta,\nu), \tag{3.1}$$

where $c>0$ and $p(\theta,\nu)$ is functionally independent of $\tau^2 \in \mathbf{R}_+$, renders predictive inference and posterior inference on θ completely robust with respect to the choice of the labelling function $g_{n,\nu}(\cdot)$. In other words, as the multivariate reference[2] Normal sampling model is a special case of (2.1) by choosing $g_{n,\nu}(\cdot) = g_n(\cdot) = (2\pi)^{-n/2} \exp(-\cdot/2)$, we obtain exactly the same predictive results and posterior results on θ as under Normality for <u>any</u> member of the elliptical class in (2.1) - (2.3). Indeed, after integrating out τ^2 using (2.2), the marginal density of (y,θ,ν) given X is [see Osiewalski and Steel (1992, Section 3)]

$$p(y,\theta,\nu|X) = c\,\Gamma(\tfrac{n}{2})\pi^{-\frac{n}{2}}p(\theta,\nu)\,|V(X,\eta)|^{-\frac{1}{2}}\, d(y,X,\theta)^{-\frac{n}{2}}, \tag{3.2}$$

which no longer depends on $g_{n,\nu}(\cdot)$. The entire influence of the choice of tail-behaviour through $g_{n,\nu}(\cdot)$ is captured by the conditional posterior of τ^2. Let us therefore examine the posterior of τ^2, and in particular its moments, under the prior structure in (3.1).

We find that the prior in (3.1) preserves the form of the sampling density (2.4) for the squared radius $r^2 = \tau^2 d(y,X,\theta)$ in its conditional posterior

$$p(r^2|y,X,\theta,\nu) = p(r^2|\nu) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}(r^2)^{\frac{n}{2}-1}g_{n,\nu}(r^2)\,. \tag{3.3}$$

Equivalently, the posterior of the precision parameter τ^2 is

$$p(\tau^2|y,X,\theta,\nu) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\, d(y,X,\theta)^{\frac{n}{2}}(\tau^2)^{\frac{n}{2}-1}g_{n,\nu}[\tau^2 d(y,X,\theta)]\,. \tag{3.4}$$

[2] Generally, the labelling function $g_{n,q}(\cdot)=(q/\pi)^{n/2}\exp(-q\cdot)$ leads to the Normal data density with variance $(\sigma^2/2q)V(X,\eta)$. As our reference Normal case we choose the value $q=1/2$.

Since (3.3) only depends on ν and not on the other parameters nor the data, posterior moments of τ^2 or of its inverse $\sigma^2 = \tau^2$ have exactly the <u>same</u> dependence on (y,X,θ) for all elliptical data densities:

$$E[(\tau^2)^m|y,X,\theta,\nu] = [d(y,X,\theta)]^{-m}\, E[(r^2)^m|\nu], \tag{3.5}$$

where $m \in M_\nu$ and $M_\nu = \{\alpha \in \mathbf{R}: \int_{\mathbf{R}_+} (r^2)^\alpha\, p(r^2|\nu)\, dr^2 < \infty\}$. In particular, if $m = -1$ is an element of M_ν, then (3.5) gives us the conditional posterior mean of the scale parameter σ^2. In the reference Normal case, where ν is absent and $p(r^2) = f_G(r^2|n/2, 1/2)$, conditional posterior moments are given by

$$E_N[(\tau^2)^m|y,X,\theta] = \frac{\Gamma(\frac{n}{2}+m)}{\Gamma(\frac{n}{2})}\left[\frac{1}{2}d(y,X,\theta)\right]^{-m} \tag{3.6}$$

and exist if and only if $m > -n/2$.

From (3.5) and (3.6) we now derive that for $m \in M_\nu \cap (-n/2, \infty)$

$$E[(\tau^2)^m|y,X,\theta,\nu] = f(m,n,\nu)\; E_N[(\tau^2)^m|y,X,\theta], \tag{3.7}$$

the product of the moment under Normality in (3.6), which depends on $d(y,X,\theta)$, and the factor

$$f(m,n,\nu) = \frac{\Gamma(\frac{n}{2})}{2^m\,\Gamma(\frac{n}{2}+m)}\; E[(r^2)^m|\nu], \tag{3.8}$$

which entirely captures the influence of the choice of a particular elliptical data density but does not depend on the observed sample. Under prior (functional) independence between θ and ν this interesting product structure is preserved after marginalizing (3.6) and (3.7) with respect to θ, leading to our main result.

Proposition 1. For any elliptical data density (2.1),(2.2) with the reparameterization (2.3) and using the improper prior (3.1) with $p(\theta,\nu) = p(\theta)\,p(\nu)$ and $p(\nu)$ proper, we obtain for $m \in M_\nu \cap M_N$:

$$E[(\tau^2)^m|y,X,\nu] = f(m,n,\nu)\; E_N[(\tau^2)^m|y,X], \tag{3.9}$$

where the last factor denotes the corresponding posterior moment in the reference case of a Normal data density with mean $h(X,\beta)$ and covariance matrix $\tau^{-2}V(X,\eta)$, and M_N is the set of all $\alpha \in \mathbf{R}$ for which $E_N[(\tau^2)^\alpha|y,X]$ exists.

Proof. From Osiewalski and Steel (1992) we know that under the prior $p(\omega) = (c/\tau^2)\, p(\theta)\, p(\nu)$ the marginal posterior of (θ,ν) factorizes as $p(\theta,\nu|y,X) = p(\nu)\, p(\theta|y,X)$, where the posterior of θ is the same as under Normality, i.e.,

$$p(\theta|y,X) = p_N(\theta|y,X) \propto p(\theta)\, |V(X,\eta)|^{-\frac{1}{2}} d(y,X,\theta)^{-\frac{n}{2}}.$$

Thus, from (3.7), we obtain

$$\begin{aligned} E[(\tau^2)^m|y,X,\nu] &= f(m,n,\nu) \int_{\Theta} E_N[(\tau^2)^m|y,X,\theta]\, p_N(\theta|y,X)\, d\theta \\ &= f(m,n,\nu)\, E_N[(\tau^2)^m|y,X]\,. \end{aligned}$$

■

An obvious extension of Proposition 1 is to marginalize (3.9) with respect to ν, which, given the proof above, leads to

$$E[(\tau^2)^m|y,X] = \left(\int f(m,n,\nu)\, p(\nu)\, d\nu\right) E_N[(\tau^2)^m|y,X]\,, \tag{3.10}$$

where we integrate $f(m,n,\nu)$ with the proper prior $p(\nu)$ over the parameter space of ν. Clearly, if $f(-1,n,\nu) = 1$ the posterior mean of σ^2 is the same as in the reference Normal case, whatever the form of $p(\nu)$.

In the next section we shall examine $f(m,n,\nu)$ in some important special cases of the elliptical family. As a byproduct, we shall arrive at the result that $f(-1,n,\nu) = 1$ for multivariate Student t data densities with precision matrix $\tau^2 V(X,\eta)^{-1}$, and where $\nu \in \mathbf{R}_+$ is the degrees of freedom parameter. Zellner's (1976) seminal paper contains the latter result for linear functions $h(X,\beta) = X\beta$ and spherical models, i.e. $V(X,\eta) = I_n$.

4. SOME SUBCLASSES OF ELLIPTICAL DATA DENSITIES

4.1 Scale Mixtures of Normals.

The form of the data density is then

$$p(y|X,\omega) = \int f_N^n\left(y|h(X,\beta),\ \frac{\lambda}{\tau^2}V(X,\eta)\right) dF_\nu(\lambda), \tag{4.1}$$

where $f_N^n(.|.)$ denotes the Normal density function (see the Appendix) and $F_\nu(.)$ is a distribution function over $\mathbf{R}_+$, parameterized by ν. The squared radius density is now a mixture of gamma densities $f_G(r^2|(n/2),(1/2\lambda)$ and

$$E[(r^2)^m|\nu] = 2^m \frac{\Gamma\left(\frac{n}{2}+m\right)}{\Gamma\left(\frac{n}{2}\right)} \int_{\mathbf{R}_+} \lambda^m\, dF_\nu(\lambda), \tag{4.2}$$

so that

$$f(m,n,\nu) = \int_{\mathbf{R}_+} \lambda^m\, dF_\nu(\lambda), \tag{4.3}$$

provided that $m > -n/2$ and the integral in (4.3) is finite. Generally, for scale mixtures of Normals the m^{th} moment of r^2 is the same as under Normality if $E(\lambda^m|\nu) = 1$. In particular, $f(-1,n,\nu) = 1$ if and only if $n>2$ and $E(\lambda^{-1}|\nu) = 1$; this holds e.g. for the case where λ^{-1} is gamma distributed with parameters $(\nu/2,\ \nu/2)$ for $\nu>0$, which induces a Student t data density with ν degrees of freedom, location $h(X,\beta)$ and precision $\tau^2V(X,\eta)^{-1}$.

4.2. Beta-prime Distributed Squared Radius.

Dickey and Chen (1985, Appendix A.3) introduce the subclass with an F-distributed squared radius. Here, equivalently, we assume that r^2 has a beta-prime or inverted beta density (see Appendix):

$$p(r^2|\nu) = f_{IB}(r^2|a,b,c) \tag{4.4}$$

with $\nu = (a,b,c) \in \mathbf{R}_+^3$. Moments for $-b<m<a$ are given by

$$E\,[(r^2)^m|\nu] = \frac{\Gamma(a-m)\Gamma(b+m)}{\Gamma(a)\Gamma(b)}\, c^m, \tag{4.5}$$

leading to the expression

$$f(m,n,\nu) = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma(a-m)\Gamma\ (b+m)}{\Gamma\left(\frac{n}{2}+m\right)\Gamma(a)\Gamma(b)}\left(\frac{c}{2}\right)^m \tag{4.6}$$

for m $\in$ (-b,a) $\cap$ (-n/2, ∞). By taking b= n/2 we obtain the Pearson Type VII family (see Fang *et al.*, p.81) for which f(-1,n,ν) becomes 2a/c. Restricting this family even further by taking c=2a, thus equating the posterior mean of σ^2 with that under reference Normality, we end up in the multivariate Student case with 2a degrees of freedom and location and precision as in Subsection 4.1. The latter density is thus a common element of both subclasses considered sofar.

From (4.6) a general necessary and sufficient condition to obtain f(-1,n,ν)=1 is b>1, n>2 and (n-2)a=(b-1)c.

4.3. Beta Distributed Squared Radius.

This subclass induces data densities that are nonzero only inside the ellipsoid $E_\ell = \{ y \in \mathbf{R}^n: \tau^2\, d(y,X,\theta) \leq \ell\}$ by assuming [see Dickey and Chen (1985, Appendix A.2)]

$$p(r^2|\nu) = f_B(r^2|\nu_1, \nu_2, \ell), \tag{4.7}$$

with $\nu = (\nu_1,\nu_2,\ell) \in \mathbf{R}_+^3$, and leading to

$$E[(r^2)^m|\nu] = \frac{\Gamma(\nu_1+\nu_2)\Gamma(\nu_1+m)}{\Gamma(\nu_1+\nu_2+m)\Gamma(\nu_1)}\,\ell^m \tag{4.8}$$

for m> $-\nu_1$, and thus

$$f(m,n,\nu) = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma(\nu_1+\nu_2)\Gamma(\nu_1+m)}{\Gamma\left(\frac{n}{2}+m\right)\Gamma(\nu_1+\nu_2+m)\Gamma(\nu_1)}\left(\frac{\ell}{2}\right)^m, \tag{4.9}$$

provided m>max {-n/2, $-\nu_1$}.

Generally, (4.9) becomes unity for m= -1 if and only if n>2, ν_1 >1 and $(n-2)(\nu_1+\nu_2-1)=\ell(\nu_1-1)$.

The special case $\nu_1 = n/2$ corresponds to the multivariate Pearson Type II data density, treated in e.g. Johnson (1987, Section 6.2) and Fang *et al.* (1990, Section 3.4). In that case, the posterior mean of σ^2 is the same as under Normality if and only if $\ell = n+2\nu_2-2$ as well as $n>2$. Reducing the subclass even further by also taking $\nu_2=1$ we obtain a data density that is uniform over the ellipsoid E_ℓ and $f(-1,n,\nu)=1$ if and only if $\ell=n$.

4.4. Symmetric Kotz type Distributions.

The density generator $g_{n,\nu}(\cdot)$ is now of the form [see Fang *et al.* (1990, p.76)]

$$g_{n,\nu}(u) = C_n\, u^{N-1} \exp(-qu^s), \tag{4.10}$$

where $\nu = (q,s,N)$ with $q,s>0$, $2N+n>2$ and, from (2.2), we obtain

$$C_n = \frac{s\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}}\Gamma\left(\frac{2N+n-2}{2s}\right)}\, q^{\frac{(2N+n-2)}{2s}}. \tag{4.11}$$

Using (4.10) in (2.4) we can derive

$$p(r^2|\nu) = \frac{\pi^{\frac{n}{2}}C_n}{\Gamma\left(\frac{n}{2}\right)}(r^2)^{N+\frac{n}{2}-2}\exp(-qr^{2s}), \tag{4.12}$$

which becomes a gamma density $f_G(r^2 \mid N+(n/2)-1,\, q)$ for $s=1$. Clearly, for $N=1$ and $s=1$ the data density (2.1) is Normal, but with $(\sigma^2/2q)V(X,\eta)$ as its covariance matrix. If also $q=1/2$, we are back in the reference Normal case. Moments of the squared radius are easily calculated as [see Fang *et al.* (1990, p.77)]:

$$E[(r^2)^m|\nu] = q^{-\frac{m}{s}}\,\frac{\Gamma\left(\frac{2N+n+2m-2}{2s}\right)}{\Gamma\left(\frac{2N+n-2}{2s}\right)}, \tag{4.13}$$

from which

$$f(m,n,\nu) = \left(2q^{\frac{1}{s}}\right)^{-m} \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{2N+n+2m-2}{2s}\right)}{\Gamma\left(\frac{n}{2}+m\right)\Gamma\left(\frac{2N+n-2}{2s}\right)}, \tag{4.14}$$

for $m > \max\left\{-\frac{n}{2}, 1-N-\frac{n}{2}\right\}$.

The original Kotz distribution introduced in Kotz (1975) assumes $s=1$, in which case $f(-1,n,\nu) = 1$ if and only if $n>2$, $2N+n>4$ and $(n-2)q= N+(n/2)-2$. For general $s>0$, these conditions become $n>2$ and $(n-2)\,q^{\frac{1}{s}}\,\Gamma\left(\frac{2N+n-4}{2s}\right) = \Gamma\left(\frac{2N+n-2}{2s}\right)$.

4.5. Summary

Table 1 summarizes some results concerning $f(m,n,\nu)$ and the posterior mean of σ^2 in the subclasses of elliptical distributions considered here.

Table 1: Expressions for $f(m,n,\nu)$ and conditions for $E(\sigma^2 | y,X,\nu)=E_N(\sigma^2 | y,X)$ provided $-1 \in M_\nu \cap M_N$.

Subclass of elliptical distribution	$f(m,n,\nu)$	Necessary and sufficient conditions for $f(-1,n,\nu)=1$
Scale mixtures of Normals (4.1)	(4.3)	$n>2$ and $E(\lambda^{-1} \| \nu)=1$
* Student t with ν degrees of freedom, $f_G\left(\lambda^{-1} \| \frac{\nu}{2}, \frac{\nu}{2}\right)$, $\nu>0$	$\frac{\Gamma\left(\frac{\nu}{2}-m\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2}\right)^m$, $m<\frac{\nu}{2}$	$n>2$
** Cauchy, $\nu = 1$	$\frac{\Gamma\left(\frac{1}{2}-m\right)}{2^m\sqrt{\pi}}$, $m<\frac{1}{2}$	$n>2$

Table 1: continued		
Beta-prime distr. r^2 (4.4)	(4.6)	$n>2$, $b>1$ and $(n-2)a=(b-1)c$
* Pearson VII $b=n/2$	$\frac{\Gamma(a-m)}{\Gamma(a)}\left(\frac{c}{2}\right)^m$, $m<a$	$n>2$ and $c=2a$
** Student t with 2a degrees of freedom $b=n/2$, $c=2a$	$\frac{\Gamma(a-m)}{\Gamma(a)}a^m$, $m<a$	$n>2$
Beta distr. r^2 (4.7)	(4.9)	$n>2$, $\nu_1>1$ and $(n-2)(\nu_1+\nu_2-1)=\ell(\nu_1-1)$
* Pearson II $\nu_1=n/2$	$\frac{\Gamma\left(\nu_2+\frac{n}{2}\right)}{\Gamma\left(\nu_2+\frac{n}{2}+m\right)}\left(\frac{\ell}{2}\right)^m$, $m>-\nu_2-\frac{n}{2}$	$n>2$ and $\ell=n+2\nu_2-2$
** Uniformity over ellipsoid $\nu_1=n/2$, $\nu_2=1$	$\frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}+m+1\right)}\left(\frac{\ell}{2}\right)^m$, $m>-\frac{n}{2}-1$	$n>2$ and $\ell=n$
Kotz type distribution (4.10)	(4.14)	$n>2$ and $(n-2)\,q^{\frac{1}{s}}\Gamma\left(\frac{2N+n-4}{2s}\right)=$ $=\Gamma\left(\frac{2N+n-2}{2s}\right)$
* Gamma distr. r^2 $f_G\left(r^2\mid N+\frac{n}{2}-1,q\right)$ $s=1$	$(2q)^{-m}\frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(N+\frac{n}{2}+m-1\right)}{\Gamma\left(\frac{n}{2}+m\right)\Gamma\left(N+\frac{n}{2}-1\right)}$, $m>\max\left\{-\frac{n}{2},1-N-\frac{n}{2}\right\}$	$n>2$, $2N+n>4$ and $(n-2)q=N+(n/2)-2$
** Normal $N=1$, $s=1$	$(2q)^{-m}$, $m>-n/2$	$n>2$ and $q=1/2$

Note: * indicates a special case of the subclass, whereas ** identifies a further specialization of *.

5. LINEAR REGRESSION WITH KNOWN COVARIANCE STRUCTURE.

In the elliptical sampling model (2.1) - (2.3) we shall now make the assumptions that the location is a linear function $h(X,\beta) = X\beta$ with X of full column rank k and that the normalized scale matrix is known $V(X,\eta) = V$. Furthermore, we add the prior assumption that $p(\theta)$ is improper uniform on $\mathbf{R}^k$, the entire parameter space of $\theta=\beta$. Our main result in Proposition 1 then reduces to:

Corollary 1. If the sampling density of $y \in \mathbf{R}^n$ is multivariate elliptical with location vector $X\beta$, scale matrix $\tau^{-2}V$ and labelling function $g_{n,\nu}(\cdot)$ satisfying (2.2), and if we assume the improper prior $p(\beta,\tau^2,\nu) = p(\beta)p(\tau^2)p(\nu) \propto \tau^{-2}p(\nu)$ with proper $p(\nu)$, we obtain for $m \in M_\nu \cap (-(n-k)/2, \infty)$

$$E[(\tau^2)^m | y, X, \nu] = f(m, n, \nu) \frac{\Gamma\left(\frac{n-k}{2}+m\right)}{\Gamma\left(\frac{n-k}{2}\right)} \left(\frac{y'Py}{2}\right)^{-m}, \tag{5.1}$$

where $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

Proof. In this linear case with uniform prior, the posterior density of β is simply the Student t density $p(\beta | y,X) = f_t^k(\beta | n-k, (X'X)^{-1}X'y, [(n-k)/y'Py]\, X'V^{-1}X)$, which is used to marginalize (3.6) with respect to β. Then (3.9) trivially gives (5.1). ■

Applying Corollary 1 to the special case of scale mixtures of Normals as in Subsection 4.1, we obtain the Bayesian counterpart to the sampling-theory result in Theorem 1.2 of Girón *et al.* (1989). The latter find that if we consider mixing distributions with finite first and second order moments:

$$\mu_F = \int_{\mathbf{R}_+} \lambda \, dF_\nu(\lambda) \tag{5.2}$$

and

$$\sigma_F^2 = \int_{\mathbf{R}_+} (\lambda - \mu_F)^2 \, dF_\nu(\lambda), \tag{5.3}$$

then

$$\hat{\sigma}^2 = (y'Py) / [(n-k)\mu_F] \tag{5.4}$$

is the uniformly minimum variance unbiased estimator for $\sigma^2 = \tau^{-2}$, i.e. that

$$E(\tau^2\hat{\sigma}^2|X,\omega)=1, \tag{5.5}$$

and also that

$$Var(\tau^2\hat{\sigma}^2|X,\omega)=\frac{2}{n-k}+\frac{n-k+2}{n-k}\frac{\sigma_F^2}{\mu_F^2}. \tag{5.6}$$

Using (4.3) and (5.1) for m= 1,2 we find that for scale mixtures of Normals with μ_F and σ_F^2 both finite the following posterior results hold

$$E(\hat{\sigma}^2\tau^2|y,X,\nu)=1, \tag{5.7}$$

and

$$Var(\hat{\sigma}^2\tau^2|y,X,\nu)=\frac{2}{n-k}+\frac{n-k+2}{n-k}\frac{\sigma_F^2}{\mu_F^2}, \tag{5.8}$$

which are, indeed, the exact Bayesian counterparts to (5.5) and (5.6). The only difference is that Girón *et al.* (1989) condition on all parameters and examine the sampling properties of $\hat{\sigma}^2$, whereas we condition on the sample and ν [which means that we take $\hat{\sigma}^2$ in (5.4) to be given] and consider the posterior moments of τ^2.

Of course, Corollary 1 can also be used to find the posterior moments of σ^2 in the case of scale mixtures of Normals. In particular, we obtain that for n>k+2.

$$E(\sigma^2|y,X,\nu)=E(\lambda^{-1}|\nu)\,\frac{y'Py}{n-k-2} \tag{5.9}$$

and for n>k+4

$$Var(\sigma^2|y,X,\nu)=\frac{(y'Py)^2\{(n-k-2)E(\lambda^{-2}|\nu)-(n-k-4)\,[E(\lambda^{-1}|\nu)]^2\}}{(n-k-2)^2(n-k-4)} \tag{5.10}$$

For the special case of the multivariate Student t, (5.9) simplifies to

$$E(\sigma^2|y,X,\nu) = E_N(\sigma^2|y,X) = \frac{y'Py}{n-k-2}, \tag{5.11}$$

as obtained in Zellner (1976) and which is exactly the same as in the reference Normal case (see also Subsections 4.1 and 4.2), whereas (5.10) becomes

$$Var(\sigma^2|y,X,\nu) = \frac{(y'Py)^2[2+\frac{2}{\nu}(n-k-2)]}{(n-k-2)^2(n-k-4)}, \tag{5.12}$$

which reduces to the expression under Normality only if $\nu \to \infty$, i.e. if the Student t tends to the reference Normal data density.

6. POSTERIOR MOMENTS OF THE COMMON VARIANCE FACTOR

Let us now restrict our attention to those elliptical distributions which possess second order moments. This rules out cases for which $E(r^2|\nu)$ is not finite, like the Cauchy and the Student t density with 2 degrees of freedom, but allows us to consider the variance of the data density (2.1) under (2.2) and (2.3):

$$Var(y|X,\omega) = \frac{\sigma^2}{n} E(r^2|\nu)\, V(X,\eta) \tag{6.1}$$

[see Dickey and Chen (1985, p. 161) or Fang *et al.* (1990, p.34)].

Note that $E_N(r^2|\nu) = n$ so that in the case of reference Normality the common variance factor, which multiplies the properly normalized matrix $V(X,\eta)$, is just σ^2 itself. In general, however, this common variance factor, say Ψ, is

$$\Psi = \frac{\sigma^2}{n} E(r^2|\nu) = f(1,n,\nu)\,\sigma^2, \tag{6.2}$$

using (3.8). Combining (6.2) with Proposition 1 we obtain

Proposition 2. For the elliptical data densities (2.1)-(2.3) where $1 \in M$, and under the improper prior (3.1) with $p(\theta, \nu) = p(\theta)\, p(\nu)$ and $p(\nu)$ proper, we obtain for $m \in M_1 \cap M_N$:

$$E(\Psi^{-m}|y,X,\nu)=\frac{f(m,n,\nu)}{f(1,n,\nu)^{m}}E_N[(\tau^2)^m|y,X]. \tag{6.3}$$

■

This proposition allows the following interesting corollary for m=1:

Corollary 2. Under the conditions of Proposition 2 we obtain

$$E(\Psi^{-1}|y,X,\nu)=E(\Psi^{-1}|y,X)=E_N(\tau^2|y,X), \tag{6.4}$$

so that the inverse common variance factor always has the same mean as under Normality, regardless of the choice of the particular elliptical density with finite sampling variance. ■

Generally, $E_N(\tau^2|y,X) = n \int_\Theta d(y,X,\theta)^{-1} p_N(\theta|y,X)d\theta$, but in the linear case of Corollary 1 it becomes the simple expression [see (5.1)]:

$$E_N(\tau^2|y,X)=\frac{n-k}{y'Py}\equiv\hat{\Psi}^{-1}, \tag{6.5}$$

with leads to

$$E[\hat{\Psi}\Psi^{-1}|y,X,\nu]=1. \tag{6.6}$$

Expression (6.6) is the Bayesian counterpart, under elliptical data densities with $1 \in M$, and the prior in Propositions 1 and 2, of the well-known classical result that $\hat{\Psi}$ is an unbiased estimator of the common variance factor Ψ under any sampling distribution with mean $X\beta$ and allowing for a covariance matrix ΨV.

Focusing attention on the mean of the common variance factor itself, rather than of its inverse, we use Proposition 2 for m=-1 and obtain that

$$E(\Psi|y,X,\nu)=f(-1,n,\nu)\,f(1,n,\nu)\,E_N(\sigma^2|y,X). \tag{6.7}$$

Clearly, the nice robustness results in Corollary 2 do not hold for the posterior mean of Ψ. With the help of the expressions in Section 4 we can, however, evaluate (6.7) in certain elliptical subclasses, provided M_ν contains the elements (-1, 1) and -1 is in M_N. The multivariate Student t data density with $\nu > 2$ degrees of freedom, for example, does not lead to the same posterior mean of the common variance factor as under Normality, but to

$$E(\Psi|y,X,\nu) = \frac{\nu}{\nu-2} E_N[\sigma^2|y,X], \tag{6.8}$$

where the proportionality factor tends to unity as $\nu \to \infty$, i.e. as the Student t tends to Normality. Table 2 summarizes the results for the elliptical families considered in Section 4.

Table 2: Expressions for $f(-1,n,\nu)\, f(1,n,\nu)$ and conditions for $E(\Psi|y,X,\nu) = E_N(\sigma^2|y,X)$ provided $1 \in M_\nu$ and $-1 \in M_\nu \cap M_N$.

Subclass of elliptical distribution	$f(-1,n,\nu)\, f(1,n,\nu)$	Necessary and sufficient conditions for $f(-1,n,\nu)\, f(1,n,\nu) = 1$
Scale mixtures of Normals (4.1), $n > 2$	$E(\lambda^{-1}\|\nu)\, E(\lambda\|\nu)$	$n > 2$ and $E(\lambda^{-1}\|\nu)\, E(\lambda\|\nu) = 1$
* Student t with ν degrees of freedom $f_G\left(\lambda^{-1}\|\frac{\nu}{2}, \frac{\nu}{2}\right)$, $\nu > 2$	$\frac{\nu}{\nu-2}$	$n > 2$ and $\nu \to \infty$ (Normality)
Beta-prime distr. r^2 (4.4), $n>2$, $a>1$, $b>1$	$\frac{(n-2)ab}{n(a-1)(b-1)}$	$n>2, a>1, b>1$ and $\frac{ab}{a+b-1} = \frac{n}{2}$
* Pearson VII $b = n/2$	$\frac{a}{a-1}$	$n>2$ and $a \to \infty$ (Normality)
** Student t with 2a degrees of freedom $b=n/2$, $c=2a$	$\frac{a}{a-1}$	$n>2$ and $a \to \infty$ (Normality)

Table 2: continued

Beta distr. r^2 (4.7), $n>2$, $\nu_1>1$	$\frac{(n-2)(\nu_1+\nu_2-1)\nu_1}{n(\nu_1+\nu_2)(\nu_1-1)}$	$n>2$, $\nu_1>1$ *and* $2\nu_1(\nu_1+\nu_2-1) = n\nu_2$
* Pearson II $\nu_1 = n/2$	$\frac{\nu_2+\frac{n}{2}-1}{\nu_2+\frac{n}{2}}$	$n>2$ and $\nu_2 \to \infty$ (Normality)
** Uniformity over ellipsoid $\nu_1 = n/2$, $\nu_2 = 1$	$\frac{n}{n+2}$	
Kotz type distribution (4.10), $n>2$	$\frac{(n-2)\Gamma\left(\frac{2N+n-4}{2s}\right)\Gamma\left(\frac{2N+n}{2s}\right)}{n\left[\Gamma\left(\frac{2N+n-2}{2s}\right)\right]^2}$	$n>2$ and $(n-2)\Gamma\left(\frac{2N+n-4}{2s}\right)\Gamma\left(\frac{2N+n}{2s}\right) = n\left[\Gamma\left(\frac{2N+n-2}{2s}\right)\right]^2$
* Gamma distr. r^2 $f_G\left(r^2 \mid N+\frac{n}{2}-1, q\right)$, $s=1$, $2N+n>4$	$\frac{(n-2)\left(N+\frac{n}{2}-1\right)}{n\left(N+\frac{n}{2}-2\right)}$	$n>2$ and $N=1$
** Normal $N=1$, $s=1$	1	$n>2$

The elliptical distributions with gamma distributed squared radius $f_G(r^2 \mid n/2, q)$ ($q \in \mathbf{R}_+$) is the Normal family with variance $(\sigma^2/2q)V(X,\eta)$, where $\sigma^2/2q$ is equal to the common variance factor Ψ from (6.1). Therefore, within this Normal class, moments of Ψ are not affected by the choice of q (see Table 2) whereas moments of σ^2 are (see Table 1). This illustrates the difference between σ^2, an arbitrarily chosen parameter, and Ψ, which is linked to the sampling variance of the observables. Given a normalization rule for $V(X,\eta)$ in (2.3) Ψ is then uniquely identified in the entire class $g_{n,\nu,a}(\cdot) = a^{-n/2}g_{n,\nu}(\cdot/a)$ ($a \in \mathbf{R}_+$) which satisfies (2.2) if $g_{n,\nu}(\cdot)$ does. If, in addition, we wish to identify σ^2, a particular subclass needs to be chosen by fixing a.

7. CONCLUDING REMARKS

By using a standard Jeffreys' type improper prior on a scalar precision parameter τ^2 we isolate the consequences of choosing a non-Normal data density within the much richer multivariate elliptical class in the conditional posterior density of τ^2. If, in addition, we impose prior (functional) independence between the parameters θ, describing the ellipsoid, and ν, which indexes the labelling function, the posterior moments of τ^2 given ν can be expressed as their marginal counterparts under reference Normality multiplied by a factor which does not depend on the observed sample.

Examining the posterior moments of τ^2 and, in particular, the mean of its inverse, σ^2, we derive conditions under which $E(\sigma^2 \mid y,X,\nu)$ is the same as that in the reference Normal case for several well-known subclasses of elliptical distributions. Within the family of scale mixtures of Normals we find Bayesian counterparts to sampling-theoretical results in Girón *et al.* (1989) for linear regression models, and we derive the result, pre-empted in Zellner (1976) that the posterior mean of σ^2 is the same for multivariate Student t as for reference Normal sampling models.

If we express the sampling variance of the data as a product of a properly normalized matrix and a scalar common variance factor Ψ, we can consider the posterior moments of Ψ which is related to, but generally not equal to, σ^2. For any member of the multivariate elliptical class with finite sampling variance the mean of the inverse of Ψ turns out to be the same, given our previous prior assumptions. In the linear case, this leads to a Bayesian counterpart for elliptical data densities of the classical result that $\hat{\Psi} = y'Py/(n-k)$ is an unbiased estimator of Ψ under any sampling distribution with mean $X\beta$ and covariance matrix ΨV. The posterior mean of Ψ itself, however, is generally found to differ from that under Normality.

Appendix: Probability density functions.

A k-variate Normal density on $x \in \mathbf{R}^k$ with mean vector $b \in \mathbf{R}^k$ and PDS $k \times k$ 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 $x \in \mathbf{R}^k$ with $r>0$ degrees of freedom, location vector $b \in \mathbf{R}^k$ and PDS $k \times k$ precision matrix A:

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

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 $a,b,c>0$:

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

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

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

REFERENCES

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, 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, 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.

Fang, K.- T., S. Kotz and K.-W. Ng, 1990, *Symmetric Multivariate and Related Distributions* (Chapman and Hall, London).

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

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.

Kotz, S., 1975, Multivariate distributions at a cross-road, in G. P. Patil, S. Kotz and J. K. Ord, eds., *Statistical Distributions in Scientific Work, 1* (Reidel, Dordrecht).

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.

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.