Edinburgh School of Economics

Discussion Paper Series

Number 25

Robust Bayesian inference on scale parameters

Carmen Fernandez (CentER and Tilburg University)
Jacek Osiewalski (Academy of Economics, Poland)
Mark F J Steel (CentER and Tilburg University)

Date
1996

Published by
School of Economics
University of Edinburgh
30 -31 Buccleuch Place
Edinburgh EH8 9JT
+44 (0)131 650 8361
http://www.ed.ac.uk/schools-departments/economics

THE UNIVERSITY *of* EDINBURGH

ROBUST BAYESIAN INFERENCE ON SCALE PARAMETERS

Carmen Fernández[a], Jacek Osiewalski[b] and Mark F.J. Steel[a,*]

[a] *CentER and Department of Econometrics, Tilburg University, 5000 LE Tilburg, The Netherlands*

[b] *Department of Econometrics, Academy of Economics, 31-510 Kraków, Poland*

ABSTRACT

We represent random variables Z that take values in $\Re^n - \{0\}$ as $Z = RY$, where R is a positive random variable and Y takes values in an $(n-1)$-dimensional space $\mathcal{Y}$. By fixing the distribution of either R or Y, while imposing independence between them, different classes of distributions on $\Re^n$ can be generated. As examples, the spherical, l_q-spherical, υ-spherical and anisotropic classes can be interpreted in this unifying framework.

We present a robust Bayesian analysis on a scale parameter in the pure scale model and in the regression model. In particular, we consider robustness of posterior inference on the scale parameter when the sampling distribution ranges over classes related to those mentioned above. Some links between Bayesian and sampling-theory results are also highlighted.

AMS Subject Classifications: 62F15, 60E05.

Keywords: Posterior distribution; Scale invariance; Scale model; Regression model.

* Corresponding author.

1. INTRODUCTION

This paper investigates perfectly robust Bayesian inference on scale parameters. We shall call an inference procedure robust if it is not affected by changes in the sampling distribution over a particular class. Thus, we are not considering robustness with respect to the data (extreme observations), or with respect to the prior specification, but we focus on exact robustness with respect to the specification of the sampling model. In particular, we analyze what Berger (1985, p.248) calls "model robustness".

In order to provide us with a natural way to define classes of sampling distributions over which we investigate robustness, Section 2 contains a representation of points in $\Re^n - \{0\}$ in terms of pairs (y, r), where $r > 0$ and y lies in some $(n-1)$-dimensional space $\mathcal{Y}$. This can be thought of as a generalization of the usual polar coordinates, where r is the Euclidean norm and $\mathcal{Y}$ is the unit sphere S^{n-1}. Applying this representation to n-variate random variables Z, we generate classes of distributions through fixing the distribution of either Y or R, as a number of examples illustrate.

The remaining sections of the paper examine inference robustness on scale parameters within classes of sampling distributions generated in such a way. Section 3 is devoted to the case of pure scale models, whereas the more limited robustness results for a scale parameter in the regression model are presented in Section 4. The final section groups some conclusions.

2. CLASSES OF MULTIVARIATE DISTRIBUTIONS

2.1. A representation of $\Re^n$

Fernández, Osiewalski and Steel (1996) introduce a general representation of points in $\Re^n$, which has proven to be a useful tool for generating and analyzing multivariate distributions and naturally induces certain useful classes of distributions in which robustness results can be derived. They consider a one-to-one correspondence between points $z \in \Re^n - \{0\}$ and pairs (y, r), where $r > 0$ and y is in some set $\mathcal{Y}$, such that $z = ry$. This implicitly implies that $\mathcal{Y}$ is an $(n-1)$-dimensional manifold and can be represented as $\mathcal{Y} = \{k(w) : w \in \mathcal{W}\}$, where $\mathcal{W} \subset \Re^{n-1}$ is the set of angular polar coordinates and $k(\cdot)$ is a one-to-one function. Thus, we can also uniquely identify $z \in \Re^n - \{0\}$ with a pair $(w, r) \in \mathcal{W} \times \Re_+$, such that $z = rk(w)$.

In this representation, different choices of $\mathcal{Y}$ lead to different interpretations of r and y. For instance, if $\mathcal{Y}$ is the unit sphere S^{n-1}, we obtain $y = z/\|z\|_2 \in S^{n-1}$ and $r = \|z\|_2$, the Euclidean or l_2-radius. This leads to the usual polar representation. Another possibility would be to take $\mathcal{Y} = \{x \in \Re^n : \|x\|_q \equiv (\sum_{i=1}^n |x_i|^q)^{1/q} = 1\}$, the unit l_q-sphere, for some $q \in [1, \infty)$, in which case $r = \|z\|_q$ describes the l_q-norm or l_q-radius, whereas $y = z/\|z\|_q$ is the corresponding point on the unit l_q-sphere.

The representation described here allows us to uniquely identify a random variable Z that takes values in $\Re^n - \{0\}$ with a pair (Y, R), where R is a positive random variable and Y takes values in $\mathcal{Y}$, through

$$Z = RY. \tag{2.1}$$

This naturally leads to the definition of classes of n-variate distributions characterized by fixing the distribution (marginal or conditional) of either R or Y. Many well-known families of distributions can be generated in this way while imposing, in addition, independence between R and Y. The next subsections discuss some of these classes.

2.2. Fixing the distribution of Y

Given a choice of $\mathcal{Y}$ and any probability distribution P_1 on $\mathcal{Y}$, we define the class $\mathcal{S}$ as the following set of random variables on $\Re^n - \{0\}$

$$\mathcal{S} = \{Z : Z = RY \text{ through (2.1), with } R \text{ and } Y \text{ independent and } Y \text{ distributed as } P_1\}. \tag{2.2}$$

We now consider the class of distributions corresponding to the random variables in $\mathcal{S}$. By varying the choice of $\mathcal{Y}$ and P_1, we generate many classes of multivariate distributions that have appeared in the literature. In particular, we can mention the class of **v-spherical distributions**, introduced in Fernández, Osiewalski and Steel (1995). They define a continuous v-spherical distribution through the following probability density function (p.d.f.) of $Z = (Z_1, \dots, Z_n)'$ in $\Re^n$:

$$p(z) = g\{v(z)\}, \tag{2.3}$$

where $v(\cdot)$ is a scalar function, strictly positive for $z \in \Re^n - \{0\}$ and such that $v(\alpha z) = \alpha v(z)$ for all $\alpha > 0$ and $z \in \Re^n$, and $g(\cdot)$ is a nonnegative labelling function. Furthermore, $v(\cdot)$ and $g(\cdot)$ are such that (2.3) is proper. For every admissible choice of $v(\cdot)$, the corresponding class of continuous v-spherical distributions is obtained by allowing $g(\cdot)$ to be free. Thus, the class of continuous v-spherical distributions exactly corresponds to the class of all the densities that share the isodensity sets $\{z \in \Re^n : v(z) = \eta\}$ (denoted as the "v-sphere" of "v-radius" η) for $\eta > 0$.

Through judicious choices of $v(\cdot)$ we can accommodate a wide range of possible isodensity contours. For example, continuous **spherical distributions** are a special case of continuous v-spherical distributions where $v(\cdot)$ is the Euclidean norm, whereas continuous **l_q-spherical distributions**, introduced in Osiewalski and Steel (1993), correspond to $v(\cdot) = v_q(\cdot)$ with

$$v_q(z) = \begin{cases} \left(\sum_{i=1}^n |z_i|^q\right)^{1/q} & \text{if } 0 < q < \infty \\ \max_{i=1,\dots,n} |z_i| & \text{if } q = \infty, \end{cases} \tag{2.4}$$

i.e. the l_q-norm when $q \geq 1$ and its generalization when $0 < q < 1$. Another interesting example are the **elliptical distributions**, where $v(z) = (z'V^{-1}z)^{1/2}$ for any positive definite symmetric matrix V.

Consider the class $\mathcal{S}$ in (2.2) with $\mathcal{Y}$ chosen as the unit v-sphere [for some given $v(\cdot)$] and P_1 given through the following p.d.f. for the polar angles

$$f_1(w) \propto s(w)[v\{h(w)\}]^{-n}, \tag{2.5}$$

where $\{h(w) : w \in \mathcal{W}\} = S^{n-1}$ is the Euclidean unit sphere and $s(w)$ is the factor in the Jacobian of the polar transformation that involves w. Fernández *et al.* (1996) show that the class of continuous v-spherical distributions is the subset of this $\mathcal{S}$ corresponding to any continuous distribution for the υ-radius $R = \upsilon(Z)$. By also allowing for any noncontinuous distribution for R, we generate the entire υ-spherical class, which therefore fits in the framework of (2.2).

Most of our robustness results will not require the independence between R and Y assumed in (2.2). We will then instead consider the class

$$\tilde{\mathcal{S}} = \{Z : Z = RY \text{ through (2.1), with } Y \text{ distributed as } P_1\}. \tag{2.6}$$

Although $\tilde{\mathcal{S}}$ is much broader than $\mathcal{S}$ in (2.2), the latter will typically constitute its subclass of greater interest. As an illustration of this fact, the shape of the isodensity sets is no longer preserved when we move from υ-sphericity, which is an example of $\mathcal{S}$, to the associated class $\tilde{\mathcal{S}}$.

2.3. Fixing the distribution of R

We now consider classes of multivariate distributions which are generated from the representation in (2.1) by choosing a particular distribution for R.

Given a particular choice of $\mathcal{Y}$ and a fixed probability distribution P_2 on $\Re_+$, we define the class of random variables

$$\mathcal{R} = \{Z : Z = RY \text{ through (2.1), with } R \text{ and } Y \text{ independent and } R \text{ distributed as } P_2\}. \tag{2.7}$$

Classes of distributions generated through choosing $\mathcal{Y}$ and P_2 in (2.7) have received much less attention in the literature than their counterparts based on $\mathcal{S}$ in (2.2). The only example that we are aware of is the class of **anisotropic distributions** introduced in Nachtsheim and Johnson (1988), which exactly fits into the framework of (2.7), by choosing $\mathcal{Y} = S^{n-1}$ and a $\sqrt{\chi^2_n}$ distribution for the Euclidean radius $R = \|Z\|_2$. Note that Normality, which corresponds to uniformity of $Y = Z/\|Z\|_2$ on S^{n-1}, defines the intersection of the anisotropic and spherical classes.

Similarly, we can start from a situation where the n components of Z are independently sampled from an exponential power distribution with p.d.f.

$$p_q(z_i) = \begin{cases} \{2^{1+1/q}\Gamma(1+1/q)\}^{-1}\exp(-|z_i|^q/2), & \text{for } 0 < q < \infty, \\ (1/2)I_{(-1,1)}(z_i), & \text{for } q = \infty \end{cases} \tag{2.8}$$

[see Box and Tiao (1973, ch. 3)]. If we then fix the distribution of $Y = Z/\upsilon_q(Z)$ and leave that of the l_q-radius $R = \upsilon_q(Z)$ free, but independent of Y, we generate an l_q-spherical class as discussed in Subsection 2.2. Conversely, keeping R independent of Y and distributed as $(\chi^2_{2n/q})^{1/q}$ for finite values of q and as Beta$(n, 1)$ for $q = \infty$ [see Osiewalski and Steel (1993)], and letting the distribution of Y change, we will define the class of **l_q-anisotropic distributions** through (2.7), for any given $q \in (0, \infty]$.

A more general version of the class $\mathcal{R}$ in (2.7) which will appear in connection with our robustness results is given by

$$\check{\mathcal{R}} = \{Z : Z = RY \text{ through (2.1), with } R \text{ given } Y \text{ distributed as } P_2^Y\}. \tag{2.9}$$

As was the case with (2.6), we have again relaxed the independence constraint. Note, however, that $\mathcal{R}$ in (2.7) is not a subset of $\check{\mathcal{R}}$, but it is instead obtained as a particular case where $P_2^Y = P_2$.

3. BAYESIAN INFERENCE IN THE SCALE MODEL

In this section we focus on independent and identically distributed (i.i.d.) sampling from the scale model

$$X_i = \sigma\varepsilon_i, \quad i = 1, \ldots, p, \tag{3.1}$$

where $\varepsilon_1, \ldots, \varepsilon_p$ are i.i.d. n-dimensional variables with $\varepsilon_i \stackrel{d}{=} \varepsilon$ (the distribution of which does not depend on σ), and $\sigma > 0$ is a scale parameter. Using (2.1), we represent

$$\varepsilon_i = R_i Y_i, \ i = 1, \ldots, p, \text{ and } \varepsilon = RY.$$

Thus, whenever we refer to Y_i, R_i, Y or R later in the section, it should be understood in this way. With a slight abuse of notation, we will not distinguish between the random variable σ and its realizations. The same convention will apply to β in Section 4.

We shall be concerned with robustness of inference on σ with respect to the choice of the sampling distribution. The next Theorem states our main result.

Theorem 1. *Consider the scale model in* (3.1)*, with the joint probability distribution of* (Y, R) *factorized into the marginal distribution of* Y*,* P_Y*, and the conditional distribution of* R *given* Y*,* $P_{R|Y}$*. Then, under any* σ*-finite prior measure for* σ*, the posterior distribution of* σ *given* $(X_1, \ldots, X_p)$*, provided it exists, does not depend on* P_Y*.*

Proof: see Appendix. •

Theorem 1 establishes perfect robustness of posterior inference on the scale parameter σ with respect to the choice of P_Y. We therefore obtain exactly the same inference on σ under any distribution for ε in a class $\check{\mathcal{R}}$ as in (2.9). Most practically relevant choices of this class are given through $\mathcal{R}$ in (2.7), where independence between R and Y is assumed (this corresponds to $P_{R|Y} = P_R$ in the notation of Theorem 1). Thus, inference on scale remains entirely unaffected by changes in the sampling distribution within *e.g.* the anisotropic or any l_q-anisotropic class. In simple terms, Theorem 1 states that whenever we know the distribution of the "radius" R given the "direction" Y, we can conduct Bayesian inference on the scale, irrespective of the particular distribution assigned to the "direction" Y.

From the proof of the Theorem follows that the key to this result is the product structure between the distributions of σ and $Y_1, \ldots, Y_p$, which is preserved under the transformation from $(\sigma, Y_1, R_1, \ldots, Y_p, R_p)$ to $(\sigma, Y_1, \Lambda_1, \ldots, Y_p, \Lambda_p)$, where $\Lambda_i = \sigma R_i$. Therefore, the conditional distribution of σ given $(Y_1, \Lambda_1, \ldots, Y_p, \Lambda_p)$, whenever it is defined, does not depend on P_Y. Note that, since (Y_i, Λ_i) are the coordinates of X_i in the chosen representation of $\Re^n$, we can immediately derive the posterior distribution of σ given $(X_1, \ldots, X_p)$ from the conditional distribution of σ given $(Y_1, \Lambda_1, \ldots, Y_p, \Lambda_p)$.

If the prior distribution of σ is a probability measure, the posterior distribution of σ given $(X_1, \ldots, X_p)$ will always exist. However, if the prior distribution of σ is improper, we require the predictive distribution [*i.e.* the marginal distribution of $(X_1, \ldots, X_p)$] to be σ-finite in order to obtain a proper posterior. From expression $(A.2)$ in the Appendix we can deduce that a necessary and sufficient condition for a σ-finite distribution of $(Y_1, \Lambda_1, \ldots, Y_p, \Lambda_p)$, and, thus, of $(X_1, \ldots, X_p)$, is that there exist p sequences of measurable sets in $\Re_+$, $\{C_i^{(l)}\}_{l \in \mathcal{N}}$ $(i = 1, \ldots, p)$, such that $\cup_{l \in \mathcal{N}} C_i^{(l)} = \Re_+$, for which

$$\int_0^\infty \prod_{i=1}^p P_{R|Y=y_i}\{r : \sigma r \in C_i^{(l)}\} \mathcal{D}_\sigma(d\sigma) < \infty, \ a.e. - (P_Y)^p, \tag{3.2}$$

where $\mathcal{D}_\sigma$ denotes the prior distribution of σ. We stress that in most practical situations the prior chosen for σ will either be a probability distribution, in which case the posterior distribution is obviously defined, or the standard non-informative (Jeffreys') prior, with density $p(\sigma) \propto \sigma^{-1}$, which can be shown from (3.2) to lead to a proper posterior even after just one observation from (3.1). For this particular prior, however, a more straightforward proof of the existence of the posterior (after one observation $X = \sigma\varepsilon = \sigma RY$) can be given as follows: Due to the invariance of this prior with respect to scale transformations, the joint distribution of $(\Lambda = \sigma R, Y, R)$ is the product measure

$$\mathcal{D}_{(\Lambda, Y, R)} = \mathcal{D}_\Lambda \times P_{(Y,R)} = \mathcal{D}_\Lambda \times P_Y \times P_{R|Y}, \tag{3.3}$$

where $\mathcal{D}_\Lambda$ again has density $p(\lambda) \propto \lambda^{-1}$. Thus, the conditional distribution of R given (Y, Λ) is just $P_{R|Y}$, which is proper by assumption. This directly implies that the distribution of $\sigma = \Lambda/R$ given (Y, Λ), or alternatively X, is also well-defined.

Often, both $P_{R|Y}$ and the prior distribution of σ are given through density functions (with respect to Lebesgue measure), say $f_2(r; Y)$ and $p(\sigma)$, respectively. The posterior distribution of σ given $(X_1, \ldots, X_p)$ is then characterized through the density

$$p(\sigma | X_1, \ldots, X_p) \propto p(\sigma) \prod_{i=1}^p \frac{1}{\sigma} f_2\left(\frac{\Lambda_i}{\sigma}; Y_i\right), \tag{3.4}$$

where (Y_i, Λ_i) are the coordinates of X_i, $i = 1, \ldots, p$, which is well-defined provided that

$$\int_0^\infty p(\sigma) \prod_{i=1}^p \frac{1}{\sigma} f_2\left(\frac{\lambda_i}{\sigma}; y_i\right) d\sigma < \infty,$$

for almost all $(y_1, \ldots, y_p)$ with respect to $(P_Y)^p$ and almost all $(\lambda_1, \ldots, \lambda_p)$ with respect to Lebesgue measure in $\Re_+^p$. Clearly, the density in (3.4) does not depend on the distribution of Y, P_Y, as we already knew from Theorem 1.

To summarize our discussion so far, in the pure scale model in (3.1), inference on σ, whenever it can be conducted, does not depend on P_Y. Therefore, posterior inference on σ is perfectly robust when the distribution of the error term $\varepsilon = RY$ ranges within any class $\check{\mathcal{R}}$ in (2.9). To illustrate this result, we now present an example concerning l_q-anisotropic classes.

Example 3.1. Bayesian inference under independent sampling from an l_q-anisotropic scale model

Assume that p n-variate observations are generated from the scale model (3.1), where the components of ε are i.i.d. with an exponential power distribution as in (2.8) for some value $q \in (0, \infty]$. For the scale parameter σ, we assume a proper prior, corresponding to

$$b\sigma^{-q} \sim \chi^2_{2a/q}, \text{ for some } a, b > 0, \text{ for } q \in (0, \infty), \quad \text{and}$$
$$b\sigma^{-1} \sim \text{Beta}(a, 1), \text{ for some } a, b > 0, \text{ for } q = \infty.$$

Since we have a proper prior, the posterior distribution of σ will clearly exist. In order to calculate the latter, we recall from Subsection 2.3 that the distribution of $R = \upsilon_q(\varepsilon)$, with $\upsilon_q(\cdot)$ defined in (2.4), given $Y = \varepsilon/\upsilon_q(\varepsilon)$ is characterized by the p.d.f.

$$f_2(r; Y) = n\left\{\Gamma\left(1 + \frac{n}{q}\right)\right\}^{-1} 2^{-n/q} r^{n-1} \exp(-r^q/2), \text{ for } 0 < q < \infty, \tag{3.5}$$

and by its limit

$$f_2(r; Y) = nr^{n-1} I_{(0,1)}(r), \text{ for } q = \infty. \tag{3.6}$$

Direct calculations lead to the posterior distribution of σ, described by

$$\left(b + \sum_{i=1}^{p}\{\upsilon_q(X_i)\}^q\right)\sigma^{-q}|(X_1, \ldots, X_p) \sim \chi^2_{2(a+pn)/q}, \text{ for } 0 < q < \infty, \tag{3.7}$$

and by

$$\max\{b, \max_{j=1,\ldots,n;i=1,\ldots,p} |X_{ji}|\}\sigma^{-1}|(X_1, \ldots, X_p) \sim \text{Beta}(a + np, 1), \text{ for } q = \infty, \tag{3.8}$$

where X_{ji}, $j = 1, \ldots, n$, denote the n components of the i^{th} observation X_i. As long as the distribution of R given Y remains as in (3.5) or (3.6), *i.e.* when ε ranges in the entire l_q-anisotropic class for given q, Theorem 1 tells us that the posterior distribution of σ does not change.

We note that in the limit as a and b both tend to zero, the kernels of our prior densities of σ tend to $p(\sigma) \propto \sigma^{-1}$, which corresponds to the standard non-informative prior, and the posterior distributions of σ [in (3.7) and (3.8)] tend to the sampling distributions for the pivots in the next example [see (3.10) and (3.12) below], considered as a function of σ.
•

Let us now look at inference robustness in the scale model (3.1) from a classical (sampling theory) perspective. In some applied statistics problems, pivotal quantities can be found that are only a function of $(R_1, \ldots, R_p)$. Clearly, such quantities will be distribution-free as long as the marginal distribution P_R of R is fixed. Thus, we will directly obtain perfect robustness under independent sampling from $\mathcal{R}$ in (2.7), and even in the wider class where we fix the marginal distribution of R without, however, imposing independence between R and Y. The latter class, say $\tilde{\mathcal{R}}$, is the counterpart of $\tilde{\mathcal{S}}$ in (2.6), where the marginal distribution of Y is fixed. Whereas $\tilde{\mathcal{R}}$ is obviously much wider than $\mathcal{R}$, the latter typically constitutes its most interesting subclass.

Example 3.2. Classical inference under independent sampling from an l_q-anisotropic scale model

As in the previous example, we consider p i.i.d. replications from the scale model in (3.1), where the n-vector ε is the result of independent sampling from an exponential power distribution, with fixed $q \in (0, \infty]$.

Writing $\varepsilon_i = R_i Y_i$ with $R_i = \upsilon_q(\varepsilon_i)$ and $Y_i = \varepsilon_i / \upsilon_q(\varepsilon_i)$, (3.5) directly leads to

$$\left(\frac{\upsilon_q(X_i)}{\sigma}\right)^q = R_i^q \sim \chi^2_{2n/q}, \; i = 1, \ldots, p, \text{ for } 0 < q < \infty, \tag{3.9}$$

and taking the sum over all the observations

$$t_q\left(\frac{X_1}{\sigma}, \ldots, \frac{X_p}{\sigma}\right) \equiv \sigma^{-q} \sum_{i=1}^{p} \{\upsilon_q(X_i)\}^q | \sigma \sim \chi^2_{2np/q}. \tag{3.10}$$

On the other hand, we know from (3.6)

$$\frac{\max_{j=1,\ldots,n} |X_{ji}|}{\sigma} = R_i \sim \text{ Beta}(n, 1), \; i = 1, \ldots, p, \text{ for } q = \infty, \tag{3.11}$$

where $X_i = (X_{1i}, \ldots, X_{ni})$ for $i = 1, \ldots, p$, and maximizing across observations

$$t_\infty\left(\frac{X_1}{\sigma}, \ldots, \frac{X_p}{\sigma}\right) \equiv \sigma^{-1}\{\max_{j=1,\ldots,n; i=1,\ldots,p} |X_{ji}|\} | \sigma \sim \text{ Beta}(np, 1). \tag{3.12}$$

The pivotal quantities $t_q(X_1/\sigma, \ldots, X_p/\sigma)$, which could be used for inference on σ, only depend on $(R_1, \ldots, R_p)$. As an immediate consequence, their distributions remain as in (3.10) and (3.12) as long as the marginal distribution of R stays fixed as in (3.9) or, respectively, (3.11). If, in addition, we impose independence between R and Y, we obtain perfect robustness whenever ε ranges in the l_q-anisotropic class, for given q. •

We remark that the pivotal quantities considered here are only functions of $(R_1 = \Lambda_1/\sigma, \ldots, R_p = \Lambda_p/\sigma)$, where $(\Lambda_1, \ldots, \Lambda_p)$ denote the radial coordinates of each of the observations in the representation chosen for $\Re^n$. As a consequence, the sampling distributions of these pivotal quantities only depend on the marginal distribution of R, P_R, derived from $P_{(Y,R)}$. This immediately leads to robustness of classical inference on σ with respect to the choice of the conditional distribution $P_{Y|R}$. On the other hand, Theorem 1 establishes robustness of Bayesian inference on σ with respect to the choice of the marginal distribution of Y, P_Y. Thus, our classical and Bayesian robustness results refer to opposite factorizations of the joint distribution of $P_{(Y,R)}$. From the proof of Theorem 1 we deduce, however, that drawing inferences only on the basis of $(\Lambda_1, \ldots, \Lambda_p)$, the radial coordinates of the observations, while discarding the data on $(Y_1, \ldots, Y_p)$, leads to a loss of relevant information about σ if R and Y are not independent [see also (3.4)]. If we only focus on distributions of the error term $\varepsilon = RY$ that impose independence between R and Y, both factorizations of $P_{(Y,R)}$ coincide and we obtain a parallelism between classical and Bayesian results. In this case, both classical and Bayesian inference on scale are completely robust when ε ranges in the class $\mathcal{R}$ in (2.7). Examples 3.1 and 3.2 illustrate this parallelism between Bayesian and classical robustness.

It is worthwhile to note that some Bayesian robustness results for certain location models follow directly from Theorem 1. We now assume that X_i in (3.1) only takes values in $\Re^n_+$, the positive orthant of $\Re^n$, which simply means that we restrict attention to those marginal distributions of Y, P_Y, that concentrate all the mass on $\mathcal{Y}_+ = \mathcal{Y} \cap \Re^n_+$. By taking natural logarithms, (3.1) can be transformed to

$$X_i^\dagger = \mu \iota_n + \varepsilon_i^\dagger, \quad i = 1, \ldots, p, \tag{3.13}$$

i.e. a pure location model, where $\mu = \ln(\sigma) \in \Re$, ι_n is an n-dimensional vector of ones and the superscript $\dagger$ denotes the coordinatewise natural logarithm transformation. Clearly, the errors $\varepsilon_i^\dagger$ are i.i.d. with $\varepsilon_i^\dagger \overset{d}{=} \varepsilon^\dagger = R^\dagger \iota_n + Y^\dagger$, where $R^\dagger$ and $Y^\dagger$ respectively take values in $\Re$ and $\mathcal{Y}_+^\dagger = \{y^\dagger \in \Re^n : y \in \mathcal{Y}_+\}$. Direct application of Theorem 1 now implies that posterior inference on μ in the model (3.13) only depends on the distribution of $(Y^\dagger, R^\dagger)$ through the conditional distribution of $R^\dagger$ given $Y^\dagger$, and is thus perfectly robust with respect to the choice of the marginal distribution of $Y^\dagger$ [on the $(n-1)$-dimensional manifold $\mathcal{Y}_+^\dagger$]. An example of this robustness is now provided.

Example 3.3. Bayesian inference on the location parameter of an extreme value distribution.

Assume that p observations are generated from the scale model (3.1), where the n components of ε are i.i.d. following a Gamma distribution with known mean and variance, both equal to a. Note that this is the distribution of $|Z|^{1/a}/2$, if Z is a random variable with the exponential power distribution in (2.8) for $q = 1/a$. Let $R = \iota_n' \varepsilon \in \Re_+$ and $Y = (\iota_n' \varepsilon)^{-1} \varepsilon \in \mathcal{Y}_+ = \{y \in \Re^n_+ : \iota_n' y = 1\}$ correspond to the representation chosen for $\Re^n$. Our distributional assumption on ε implies that R is Gamma distributed, independent of Y, with mean and variance both equal to na, whereas Y has a Dirichlet distribution with all n parameters equal to a. In terms of the equivalent location model (3.13), the n components of $\varepsilon^\dagger$ are i.i.d. with p.d.f. in $\Re$

$$p(t) = \{\Gamma(a)\}^{-1} \exp(at) \exp\{-\exp(t)\}, \tag{3.14}$$

which, for $a = 1$, is the density of an extreme value (Gumbel) distribution. Clearly, $R^\dagger$ and $Y^\dagger$ are independent, the distribution of $R^\dagger$ is given through the p.d.f.

$$f_2(r^\dagger) = \{\Gamma(na)\}^{-1} \exp(nar^\dagger) \exp\{-\exp(r^\dagger)\}, \tag{3.15}$$

in $\Re$, whereas the distribution of $Y^\dagger$ on $\mathcal{Y}_+^\dagger = \{(y_1^\dagger, \ldots, y_n^\dagger)' \in \Re^n : \sum_{j=1}^n \exp(y_j^\dagger) = 1\}$ corresponds to the Dirichlet distribution of Y on $\mathcal{Y}_+$. From Theorem 1 we can now deduce that posterior inference on the common location μ in (3.13) remains exactly the same if, instead of pn independent errors with the extreme-value type distribution in (3.14), we have p independent n-variate errors $\varepsilon_i^\dagger \overset{d}{=} \varepsilon^\dagger = R^\dagger \iota_n + Y^\dagger$, where $R^\dagger$ and $Y^\dagger$ are independent, $R^\dagger$ has the p.d.f. in (3.15) and $Y^\dagger$ has any distribution on $\mathcal{Y}_+^\dagger$. •

Let us now go back to the general scale model in (3.1) and, again, focus on inference on scale. In a Bayesian context we can also obtain perfect robustness with respect to the form of $P_{R|Y}$, although the results are much more moderate. In particular, our findings only relate to posterior moments of the scale parameter rather than to its entire posterior distribution. Moreover, they rely on a particular prior distribution, namely the scale invariant Jeffreys' prior with density $p(\sigma) \propto \sigma^{-1}$, and can only be obtained for the case of one observation $X = \sigma\varepsilon = \sigma RY$ from (3.1). These findings are based on the fact that the conditional distribution of R given (Y, Λ) is then just $P_{R|Y}$, as shown in (3.3). Thus, for any $\alpha \in \Re$,

$$E(\sigma^{\alpha}|X) = \Lambda^{\alpha} E(R^{-\alpha}|Y, \Lambda) = \Lambda^{\alpha} E(R^{-\alpha}|Y),$$

i.e. the posterior expectation of σ^{α}, provided it exists, only depends on $P_{R|Y}$ through its $(-\alpha)^{th}$ moment. Since, from Theorem 1, P_Y does not affect the posterior distribution of σ, we obtain perfect robustness of $E(\sigma^{\alpha}|X)$ when ε ranges in the very large class of n-variate distributions where P_Y is left free and only the $(-\alpha)^{th}$ moment of $P_{R|Y}$ is specified.

We stress again that this robustness of the α^{th} posterior moment with respect to the form of $P_{R|Y}$ only holds for a single n-variate observation, X, and does not extend to independent sampling from (3.1). The reason is that such robustness relies on the scale invariance of Jeffreys' prior for σ, which can then absorb the influence of R and preserve the product structure when we transform from (σ, Y, R) to $(\Lambda = \sigma R, Y, R)$. In the case of independent sampling we would need as many individual scale parameters (with Jeffreys' prior on each of them) as vector observations in order to achieve such robustness.

Finally, in practice, we may want the scale parameter σ to have some further meaning in terms of sampling properties of the observables. A natural such condition would be, for example, $V(X|\sigma) = \sigma^2 I_n$, *i.e.* where σ^2 describes the variance of the sampling distribution. This condition translates into $V(\varepsilon) = V(RY) = I_n$, which is clearly not fulfilled in general, unless we impose some restrictions on the distribution of (Y, R). Thus, in order for σ^2 to have the interpretation of the sampling variance, we need to narrow down our previous classes of distributions through incorporating this moment condition into the distribution of (Y, R).

4. INFERENCE ON THE SCALE IN THE REGRESSION MODEL

Let us now consider the model

$$X = b(\beta) + \sigma\varepsilon, \tag{4.1}$$

where X is an n-variate vector observation, ε is an n-dimensional random variable (the distribution of which does not depend on β and σ), $\sigma > 0$ is the scale parameter and $b(\beta)$ is the location, parameterized in terms of a vector $\beta \in \mathcal{B} \subset \Re^m (m \leq n)$ through a known function $b(\cdot)$ from $\Re^m$ to $\Re^n$. Again, using (2.1), we shall represent ε as

$$\varepsilon = RY.$$

Important examples of (4.1) are the standard location-scale model, where $b(\beta) = \beta$ with $\beta \in \Re^n$, the case of a common location, where $b(\beta) = \beta\iota_n$ with scalar β, and the regression context, where $b(\cdot)$ depends on a matrix of exogenous variables D.

Given β, (4.1) leads to the pure scale model $X - b(\beta) = \sigma\varepsilon$, and the results of the previous section can thus be applied to the conditional posterior distribution of σ [*i.e.* the distribution of σ given (X, β)]. Since, however, interest usually focuses on marginal rather than on conditional posterior inference, we need to integrate out β from the conditional posterior distribution of σ given (X, β), using the marginal posterior distribution of β given X. As the latter distribution typically depends on both $P_{R|Y}$ and P_Y, our robustness results in the previous section may not carry over to marginal posterior inference on σ. In particular, the robustness with respect to the choice of P_Y (Theorem 1), no longer holds after integrating out β. On the other hand, the robustness with respect to the choice of $P_{R|Y}$ presented at the end of Section 3 can be shown to also apply to marginal posterior inference on σ in the context of (4.1). This section is devoted to establishing and discussing this result.

We assume that X is generated by the model in (4.1), and that the prior distribution of (β, σ) has the product structure

$$\begin{aligned} &\mathcal{D}_{(\beta,\sigma)} = \mathcal{D}_\beta \times \mathcal{D}_\sigma, \\ &\text{where } \mathcal{D}_\beta \text{ is any } \sigma\text{-finite measure on } \mathcal{B} \\ &\text{and } \mathcal{D}_\sigma \text{ has density } p(\sigma) \propto \sigma^{-1}. \end{aligned} \tag{4.2}$$

Following the same reasoning as in Section 3 [see (3.3)], the scale invariance property of $\mathcal{D}_\sigma$ leads to the following joint distribution for $(\beta, \Lambda = \sigma R, Y, R)$:

$$\mathcal{D}_{(\beta,\Lambda,Y,R)} = \mathcal{D}_\beta \times \mathcal{D}_\Lambda \times P_Y \times P_{R|Y}, \tag{4.3}$$

where $\mathcal{D}_\Lambda = \mathcal{D}_\sigma$. From (4.3) it is immediate that for any $\alpha \in \Re$,

$$E(\sigma^\alpha|X, \beta) = E(\sigma^\alpha|Y, \Lambda, \beta) = \Lambda^\alpha E(R^{-\alpha}|Y, \Lambda, \beta) = \Lambda^\alpha E(R^{-\alpha}|Y), \tag{4.4}$$

where (Y, Λ) are the coordinates of $X - b(\beta)$ in the representation chosen for $\Re^n$. In order to obtain $E(\sigma^\alpha|X)$, we need to compute the expectation of $E(\sigma^\alpha|X, \beta)$ with respect to the distribution of β given X. Using Theorem 2 of Fernández *et al.* (1996) we know that, under the prior in (4.2), the joint distribution of (X, β) does not depend on $P_{R|Y}$ [*i.e.* it is exactly the same for any choice of ε in the class $\tilde{\mathcal{S}}$ in (2.6)]. Obviously, the same will hold for the marginal posterior distribution of β, provided it exists, which immediately implies that $E(\sigma^\alpha|X)$ only depends on $P_{R|Y}$ through $E(R^{-\alpha}|Y)$. The following Theorem highlights this finding.

Theorem 2. *Consider an n-variate observation X from (4.1), and the prior in (4.2). Then, for any $\alpha \in \Re$, $E(\sigma^\alpha|X)$, provided it exists, only depends on $P_{R|Y}$ through $E(R^{-\alpha}|Y)$.*

Theorem 2 thus establishes perfect robustness of the α^{th} order marginal posterior moment of σ, provided it exists, when ε ranges in a subclass of $\tilde{\mathcal{S}}$ in (2.6) characterized by a certain fixed value of $E(R^{-\alpha}|Y)$.

Often, the underlying distribution P_1 in the class $\tilde{\mathcal{S}}$ is given through a p.d.f., say $f_1(\cdot)$, for the polar angles w. For instance, when $\tilde{\mathcal{S}}$ corresponds to a class $\mathcal{S}$ in (2.2) of υ-spherical distributions, P_1 is given through $f_1(w)$ in (2.5). In the case of a continuous P_1, we can deduce from Fernández *et al.* (1996) the following expression for the joint distribution of (X, β):

$$\mathcal{D}_{(X,\beta)}(A \times B) \propto \int_{A\times B} \|x - b(\beta)\|_2^{-n} s(w)^{-1} f_1(w) dx \mathcal{D}_\beta(d\beta), \tag{4.5}$$

for each pair of measurable sets $A \subset \Re^n$ and $B \subset \mathcal{B}$, and where $\|\cdot\|_2$ stands for the Euclidean norm, w are the angular polar coordinates of $x - b(\beta)$ and $s(w)$ is the factor in the Jacobian of the polar transformation that depends on w. Properness of the posterior distribution requires that the marginal distribution of X, computed from $\mathcal{D}_{(X,\beta)}$ in (4.5), is σ-finite. As noted in Fernández *et al.* (1996), in most practically relevant situations [like *e.g.* any υ-spherical class where $\upsilon(\cdot)$ is such that the isodensity sets do not touch the origin], the latter requirement will not be met in the pure location-scale model, where $b(\beta) = \beta$, and $\beta \in \Re^n$ is not restricted to a lower dimensional subspace. Thus, in order to have a well-defined posterior in the Bayesian model (4.1) – (4.2), we typically require a regression context with $m < n$. Provided that the posterior distribution exists, we can now combine (4.4) and (4.5) to obtain an exact expression for $E(\sigma^\alpha|X)$:

$$E(\sigma^\alpha|X) = \frac{\int_{\mathcal{B}} \Lambda^\alpha E(R^{-\alpha}|Y) \|X - b(\beta)\|_2^{-n} s(W)^{-1} f_1(W) \mathcal{D}_\beta(d\beta)}{\int_{\mathcal{B}} \|X - b(\beta)\|_2^{-n} s(W)^{-1} f_1(W) \mathcal{D}_\beta(d\beta)}, \tag{4.6}$$

where (Y, Λ), $\Lambda > 0$, $Y \in \mathcal{Y}$, are the coordinates of $X - b(\beta)$ in the representation chosen for $\Re^n$ and $W = k^{-1}(Y)$ denotes the angular polar coordinates of $X - b(\beta)$ as explained in Subsection 2.1.

In an applied setup, we may require that σ has some further meaning in terms of sampling properties of the observables, such as

$$V(X|\beta, \sigma) = \sigma^2 I_n, \text{ or, equivalently, } V(\varepsilon) = V(RY) = I_n. \tag{4.7}$$

This implies a moment restriction on $P_{(Y,R)}$. As a consequence, we can not just consider any class $\tilde{\mathcal{S}}$ in (2.6), since certain choices of P_Y [P_1 in the notation of (2.6)] will preclude (4.7). Once we have a valid choice for P_Y, we still need to restrict the corresponding class $\tilde{\mathcal{S}}$ by only considering those distributions $P_{R|Y}$ for which (4.7) is fulfilled. There are, nevertheless, many rich classes of n-variate distributions that are compatible with these assumptions, as will be illustrated in the ensuing examples.

Example 4.1. l_q-spherical distributions with unitary variance

Assume the model in (4.1), $X = b(\beta) + \sigma\varepsilon$, where ε is an n-dimensional random vector following an l_q-spherical distribution with unitary variance, for some fixed value of $q \in (0, \infty]$. We further consider the prior distribution in (4.2).

Following Osiewalski and Steel (1993), if ε is l_q-spherical,

$$E(\varepsilon) = 0 \text{ and } V(\varepsilon) = c_q^{-1} E[\upsilon_q(\varepsilon)^2] I_n, \tag{4.8}$$

where $\upsilon_q(\cdot)$, defined in (2.4), coincides with the l_q-norm for $q \geq 1$,

$$c_\infty = \frac{3n}{n+2}, \text{ and } c_q = c_\infty \frac{\Gamma\left(1 + \frac{1}{q}\right)\Gamma\left(1 + \frac{n+2}{q}\right)}{\Gamma\left(1 + \frac{3}{q}\right)\Gamma\left(1 + \frac{n}{q}\right)} \text{ for finite } q.$$

The unitary variance assumption [as in (4.7)] implies $E[\upsilon_q(\varepsilon)^2] = c_q$, defining a subset of the class of l_q-spherical distributions through fixing the second order moment of $R = \upsilon_q(\varepsilon)$.

Applying (4.6) to $\alpha = -2$, with $f_1(w)$ given by (2.5) with $\upsilon(\cdot) = \upsilon_q(\cdot)$, the posterior mean of the precision takes the following form, common to all distributions in this restricted l_q-spherical class:

$$E(\sigma^{-2}|X) = c_q \frac{\int_{\mathcal{B}} [\upsilon_q\{X - b(\beta)\}]^{-(n+2)} \mathcal{D}_\beta(d\beta)}{\int_{\mathcal{B}} [\upsilon_q\{X - b(\beta)\}]^{-n} \mathcal{D}_\beta(d\beta)}. \tag{4.9}$$

From (4.8) follows that $E(Y) = 0$ and $V(Y) = c_q^{-1} I_n$, where $Y = \varepsilon/\upsilon_q(\varepsilon)$. It is then easy to see that $V(\varepsilon) = I_n$ for any choice of $P_{R|Y}$ such that $E(R|Y)$ does not depend on Y and $E(R^2|Y) = c_q$. From (4.6), the expression in (4.9) for $E(\sigma^{-2}|X)$ extends to this wider class. However, the case of independence between R and Y, *i.e.* the subset of the l_q-spherical class, seems the most interesting from a practical perspective. For instance, we know that, in the continuous case, all distributions for ε will only then share the same isodensity sets. $\bullet$

Example 4.2. Linear regression with elliptical errors

Let us consider the model $X^* = D\beta + \sigma\varepsilon^*$, where ε^* is an n-variate elliptical random vector with mean zero and a known positive definite symmetric covariance matrix V, and D is a known $n \times m$ matrix of full colum rank with $n > m$. In terms of $X \equiv V^{-1/2}X^*$ and $\varepsilon \equiv V^{-1/2}\varepsilon^*$, we have a linear regression model with a spherical error vector,

$$X = V^{-1/2} D\beta + \sigma\varepsilon,$$

where $E(\varepsilon) = 0$ and $V(\varepsilon) = I_n$, which directly fits into the framework of (4.7). Since ε follows a spherical distribution, $Y = \varepsilon/\|\varepsilon\|_2$ is uniformly distributed over the unit sphere S^{n-1} and independent of $R = \|\varepsilon\|_2$. As the covariance matrix of the uniform distribution over S^{n-1} is $\frac{1}{n} I_n$ [see Fang *et al.* (1990, p.34)], we obtain $E(R^2) = n$ as the only restriction on the distribution of $R = \|\varepsilon\|_2$, which can also be seen from the previous example with $q = 2$.

Under the prior density $p(\beta, \sigma) = p(\beta)p(\sigma) \propto \sigma^{-1}$, the marginal posterior density for β, derived from (4.5), takes the form

$$p(\beta|X) = p(\beta|X^*) = f_S^m(\beta|n - m, \hat{\beta}, \hat{\sigma}^{-2} D'V^{-1}D),$$

which corresponds to the m-variate Student-t distribution with $n - m$ degrees of freedom, location vector $\hat{\beta} = (D'V^{-1}D)^{-1}D'V^{-1}X^*$, and precision matrix $\hat{\sigma}^{-2}D'V^{-1}D$, where $\hat{\sigma}^{-2} = (\hat{\sigma}^2)^{-1} = (n - m)\{(X^* - D\hat{\beta})'V^{-1}(X^* - D\hat{\beta})\}^{-1}$.

Integrating $E(\sigma^{-2}|X^*, \beta)$, obtained as in (4.4), with $p(\beta|X^*)$, leads to the posterior mean of the inverse variance

$$E(\sigma^{-2}|X^*) = n \int_{\Re^m} \frac{1}{(X^* - D\beta)'V^{-1}(X^* - D\beta)} f_S^m(\beta|n - m, \hat{\beta}, \hat{\sigma}^{-2}D'V^{-1}D) d\beta = \hat{\sigma}^{-2},$$

common to all spherical distributions for ε with unitary variance. From (4.6) follows that the latter expression for $E(\sigma^{-2}|X^*)$ also applies in the wider class where Y and R are no longer independent but $P_{R|Y}$ is such that $E(R^2|Y) = n$ (while P_Y is uniform over S^{n-1}). If, in addition, $E(R|Y)$ does not depend on Y, σ^2 keeps its interpretation as the sampling variance.

Since $\hat{\sigma}^2$ is an unbiased estimator of σ^2, we have, for any elliptical distribution of ε^* with a fixed covariance matrix V, an interesting classical-Bayesian parallel:

$$E(\sigma^{-2}\hat{\sigma}^2|\beta,\sigma) = 1 \quad \text{and} \quad E(\hat{\sigma}^2\sigma^{-2}|X^*) = 1,$$

first noted by Osiewalski and Steel (1996). •

Finally, we remark that the robustness results in this section only hold for the case of one single n-variate replication of (4.1), since they crucially hinge upon the scale invariance of the prior distribution for σ, which is lost after one observation (see also the penultimate paragraph in Section 3). From the findings in Fernández *et al.* (1996) follows that, in order to obtain such robustness under independent sampling from (4.1), it is required that each vector observation has its own scale parameter. Since our interest here focuses on a particular scale parameter, we have only considered one n-variate observation X.

5. CONCLUSIONS

Bayesian robust inference on the scale parameter under i.i.d. sampling from the pure scale model in (3.1) is seen to hold for any prior. In particular, we have found perfect robustness with respect to the marginal distribution of Y. The Bayesian perspective clearly shows that inference on the scale based on only R (as is the case for classical inference based on some natural pivotal quantities) involves a loss of information, unless R and Y are independent. However, under independence between R and Y, we have a perfect parallel between robustness of sampling theory inference based on any function of R and of Bayesian inference.

If we are interested in the posterior moments of the scale parameter in the context of the regression model (4.1) under the prior (4.2), we find that a Bayesian analysis will lead to exactly the same values for such moments when ε ranges over a certain subclass of $\tilde{\mathcal{S}}$ in (2.6). As an interesting example, the posterior mean of the precision (inverse variance) of the observables will only depend on the distribution of R given Y through the second order moment $E(R^2|Y)$, so fixing the latter will naturally lead to robustness. This result, however, crucially depends on the particular choice of the prior for σ made in (4.2), and does not extend to independent sampling from (4.1). In the special case of a linear regression model with elliptical errors, this finding also has a classical counterpart (which, however, does not rely on ellipticity).

APPENDIX

Proof of Theorem 1

We present the proof for the case of one single observation $X = \sigma\varepsilon = \sigma RY$ from (3.1); the extension to repeated sampling is straightforward.

The joint distribution of (σ, Y, R), denoted by $\mathcal{D}_{(\sigma,Y,R)}$, can be factorized as

$$\mathcal{D}_{(\sigma,Y,R)} = \mathcal{D}_\sigma \times P_{(Y,R)} = \mathcal{D}_\sigma \times P_Y \times P_{R|Y},$$

where $\mathcal{D}_\sigma$ denotes the σ-finite prior distribution of σ. Transforming from (σ, Y, R) to (σ, Y, Λ), where $\Lambda = \sigma R$, we obtain for any measurable sets A and B

$$\begin{aligned} &\mathcal{D}_{(\sigma,Y,\Lambda)}\{(\sigma, y, \lambda) : \sigma \in A, (y, \lambda) \in B\} = \mathcal{D}_{(\sigma,Y,R)}\{(\sigma, y, r) : \sigma \in A, (y, \sigma r) \in B\} \\ &= \int_{B_M} \int_A P_{R|Y=y}\{r : \sigma r \in B^y\} \mathcal{D}_\sigma(d\sigma) P_Y(dy), \end{aligned} \tag{A.1}$$

where $B_M = \{y \in \mathcal{Y} : (y, \lambda) \in B \text{ for some } \lambda > 0\}$ and $B^y = \{\lambda > 0 : (y, \lambda) \in B\}$. From $(A.1)$ it is immediate that the conditional distribution of σ given (Y, Λ), provided it exists, does not depend on P_Y. As (Y, Λ) are the coordinates of X in the representation chosen for $\Re^n$, Theorem 1 follows.

Marginal distribution of (Y, Λ)

In order for the conditional probability distribution of σ given (Y, Λ) to be defined, we require that the marginal distribution of (Y, Λ) is σ-finite. This marginal distribution, derived from $(A.1)$, takes the form

$$\mathcal{D}_{(Y,\Lambda)}(B) = \mathcal{D}_{(\sigma,Y,\Lambda)}\{(0, \infty) \times B\} = \int_{B_M} \int_0^\infty P_{R|Y=y}\{r : \sigma r \in B^y\} \mathcal{D}_\sigma(d\sigma) P_Y(dy). \tag{A.2}$$

REFERENCES

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

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

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

Fernández, C., J. Osiewalski and M.F.J. Steel (1995). Modeling and inference with υ-spherical distributions. *J. Amer. Statist. Assoc.* **90**, 1331-1340.

Fernández, C., J. Osiewalski and M.F.J. Steel (1996). Classical and Bayesian inference robustness in multivariate regression models. Mimeo (Tilburg University, Tilburg).

Nachtsheim, C.J. and M.E. Johnson (1988). A new family of multivariate distributions with applications to Monte Carlo studies. *J. Amer. Statist. Assoc.* **83**, 984-989.

Osiewalski, J. and M.F.J. Steel (1993). Robust Bayesian inference in l_q-spherical models. *Biometrika* **80**, 456-460.

Osiewalski, J. and M.F.J. Steel (1996). Posterior moments of scale parameters in elliptical sampling models. In: D. Berry, K. Chaloner and J. Geweke, Eds., *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, Ch. 27. Wiley, New York.