NISS

# Default Priors for Gaussian Processes

Rui Paulo

Technical Report Number 139
December, 2003

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

## Acknowledgment and Disclaimer

This material is based upon work supported by the National Science Foundation under Grant No. DMS-00739520

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

# DEFAULT PRIORS FOR GAUSSIAN PROCESSES

By Rui Paulo*

*National Institute of Statistical Sciences*

*and*

*Statistical and Applied Mathematical Sciences Institute*

## Abstract

Motivated by the statistical evaluation of complex computer models, we deal with the issue of objective prior specification for the parameters of Gaussian processes. In particular, we derive the Jeffreys-rule, independence Jeffreys and reference priors for this situation, and prove that the resulting posterior distributions are proper under a quite general set of conditions. Another prior specification strategy, based on maximum likelihood estimates, is also considered, and all priors are then compared on the grounds of the frequentist properties of the ensuing Bayesian procedures. Computational issues are also addressed in the paper, and we illustrate the proposed solutions by means of an example taken from the field of complex computer model validation.

*Research was supported by the U.S. National Science Foundation, Grants DMS-0073952 at the National Institute of Statistical Sciences, DMS-0103265, and the Statistical and Applied Mathematical Sciences Institute. This research formed part of the author's Ph.D. thesis at Duke University, where he was partly supported by a Ph.D. Fellowship awarded by Fundação para a Ciência e a Tecnologia, Portugal, with reference PRAXIS XXI/BD/15703/98.

*Key Words and phrases.* Gaussian process, Jeffreys prior, reference prior, integrated likelihood, frequentist coverage, posterior propriety, computer model.

# 1 Introduction

In this paper, we address the problem of specifying objective priors for the parameters of general Gaussian processes. We derive formulas for the Jeffreys-rule prior; for an independence Jeffreys prior; and for a reference prior on the parameters involved in the parametric specification of the mean and covariance functions of a Gaussian process. The mean is assumed to be a $q$-dimensional linear model on location-dependent covariates, and the correlation function involves an $r$-dimensional vector of unknown parameters. The resulting posteriors are shown to be proper under a more restrictive, yet with practical relevance, scenario. We also address computational issues, and in particular we devise a sampling scheme to draw from the resulting posteriors that requires very little input from the user. An empirical Bayes procedure, based on maximum likelihood estimates, is also described, and we present the results of a simulation study designed in order to compare the frequentist properties of all the Bayesian methods described in the paper. An example, based on real data, is used to illustrate some of the proposed solutions.

The motivation for considering the problems addressed in this paper is both theoretical and practical. From the applied point of view, these results are of interest in general for the field of spatial statistics, but especially for the analysis and validation of complex computer models. Indeed, one prominent approach to this problem involves fitting a Gaussian process to the computer model output, and a separable correlation function involving several parameters, typically a multidimensional power exponential, is frequently assumed — cf., e.g., Sacks et al. (1989), Kennedy and O'Hagan (2000, 2001) and Bayarri et al. (2002). The computer models are often computationally extremely demanding and this is a way of providing a cheaper surrogate that can be used in the design of computer experiments; optimization problems; prediction and uncertainty analysis; calibration; and validation of the model. In the Bayesian approach, one must specify prior distributions for the parameters of the Gaussian processes. Typically, little or no prior information about the parameters is available, and their interpretation is not always straightforward, so that automatic or default procedures are sought. The need for default specification of priors, and also for the development of computational schemes for the resulting posteriors, is thus considerable in this area.

From a more theoretical perspective, this article is also relevant. Berger, De Oliveira, and Sansó (2001) consider objective Bayesian analysis of Gaussian spatial processes with a quite general correlation structure, but the study is restricted to the situation where only the (one-dimensional) range parameter is considered to be unknown. The original motivation for their paper was the observation that commonly prescribed default priors could fail to yield proper posteriors. A more in-depth study of the problem revealed very interesting and unusual facts. In the presence of an unknown mean level for the Gaussian process, the integrated likelihood for the parameters governing the correlation structure is typically bounded away from zero, which explains the difficulty with posterior propriety. Also, the independence Jeffreys prior (assuming the parameters in the mean level are *a priori* independent of the ones involved in the covariance), which is often prescribed, fails to yield a proper posterior when the mean function includes an unknown constant level. The usual algorithm for the reference prior of Bernardo (1979) and Berger and Bernardo (1992), in which asymptotic marginalization is used, also fails to produce a proper posterior — this is the first known example in which exact marginalization is required to achieve posterior propriety. The authors end up recommending this "exact" reference prior.

The next logical step is to investigate whether these results still hold in higher dimensions, which is precisely one of the aims of the present article. The answer is no, in the sense that there are no surprises in terms of posterior propriety, and we describe in detail the reasons for that.

The paper is organized as follows: Section 2 sets up some notation and establishes formulas for the objective priors that are valid in a very general setting. The next section begins by describing the scenario where analytical results have been achieved, and in the sequel we analyze the behavior of the integrated likelihood and of the priors. Finally, conditions ensuring posterior propriety are determined.

Section 4 addresses computational aspects of the problem, and in particular we describe a Markov chain algorithm to sample from the posterior that requires very little input from the user. This algorithm involves computing maximum likelihood estimates and the Fisher information matrix. Taking advantage of the availability of these quantities, we propose also an Empirical Bayes approach to the problem. This section ends with an example, using real data, that illustrates some of the proposed solutions.

Section 5 presents the results of a simulation study designed to compare the frequentist properties of the Bayesian procedures proposed in the paper, and ends with some final recommendations.

In the Appendix we present the proofs of the various results that are described in the body of the paper.

## 2 Notation and the Objective Priors

Let us consider the following rather general situation: $Y(\cdot)$ is a Gaussian process on $\mathscr{S} \subset \mathbb{R}^p$ with mean and covariance functions given respectively by

$$\mathbb{E}\ Y(\boldsymbol{x}) \equiv \boldsymbol{\Psi}(\boldsymbol{x}) \cdot \boldsymbol{\theta}$$

and

$$\mathbb{C}\mathsf{ov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}^\star)) \equiv \sigma^2\ c(\boldsymbol{x}, \boldsymbol{x}^\star \mid \boldsymbol{\xi})\ ,$$

where $c(\cdot,\cdot)$ is the correlation function, $\sigma^2$ is the variance, and $\boldsymbol{\xi}$ is an $r$-dimensional vector of unknown parameters. The vector $\boldsymbol{\Psi}(\boldsymbol{x}) = (\psi_1(\boldsymbol{x}), \ldots, \psi_q(\boldsymbol{x}))'$ is a $q$-vector of location-dependent covariates. Define $\boldsymbol{\eta} = (\sigma^2, \boldsymbol{\theta}', \boldsymbol{\xi}')'$.

The stochastic process $Y(\cdot)$ is observed at locations $S = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and hence the resulting random vector, $\boldsymbol{Y} = (Y(\boldsymbol{x}_1), \ldots, Y(\boldsymbol{x}_n))'$, satisfies

$$\boldsymbol{Y} \mid \boldsymbol{\eta} \sim \mathsf{N}(\boldsymbol{X}\boldsymbol{\theta}, \sigma^2\boldsymbol{\Sigma}) \tag{2.1}$$

where $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\boldsymbol{\xi}) = [c(\boldsymbol{x}_i, \boldsymbol{x}_j \mid \boldsymbol{\xi})]_{ij}$ and

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{\Psi}(\boldsymbol{x}_1)' \\ \vdots \\ \boldsymbol{\Psi}(\boldsymbol{x}_n)' \end{pmatrix} . \tag{2.2}$$

As a result, the associated likelihood function based on the observed data $\boldsymbol{y}$ is given by

$$L(\boldsymbol{\eta} \mid \boldsymbol{y}) \propto (\sigma^2)^{-n/2}\ |\boldsymbol{\Sigma}|^{-1/2}\ \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})\right\} . \tag{2.3}$$

The priors on $\boldsymbol{\eta}$ that we will consider are of the form

$$\pi(\boldsymbol{\eta}) \propto \frac{\pi(\boldsymbol{\xi})}{(\sigma^2)^a} \tag{2.4}$$

for different instances of $\pi(\boldsymbol{\xi})$ and $a$. Indeed, we will in the next two propositions show that this is the case for the Jeffreys-rule prior, for an independence Jeffreys prior and for a reference prior. This general prior follows a familiar form: flat on the parameters that specify the mean and usual forms for the variance $\sigma^2$.

Following Berger et al. (2001), in order to derive the reference prior we specify the parameter of interest to be $(\sigma^2, \boldsymbol{\xi})$ and consider $\boldsymbol{\theta}$ to be the nuisance parameter. This corresponds in the reference prior algorithm to factoring the prior as

$$\pi^R(\boldsymbol{\eta}) = \pi^R(\boldsymbol{\theta} \mid \sigma^2, \boldsymbol{\xi})\ \pi^R(\sigma^2, \boldsymbol{\xi})$$

and selecting $\pi^R(\boldsymbol{\theta} \mid \sigma^2, \boldsymbol{\xi}) \propto 1$, because that is the Jeffreys-rule prior for the model at hand when $\sigma^2$ and $\boldsymbol{\xi}$ are considered known. Next, $\pi^R(\sigma^2, \boldsymbol{\xi})$ is calculated as the reference prior but for the marginal model defined by the integrated likelihood with respect to $\pi^R(\boldsymbol{\theta})$. It is this marginalization step that usually is carried in an asymptotic fashion; Berger et al. (2001) recommend that, for statistical models with a complicated covariance structure, the exact marginalization should become standard practice.

Before stating the formula for this reference prior, let us define $\dot{\boldsymbol{\Sigma}}^k$ as the matrix that results from $\boldsymbol{\Sigma}$ by differentiating each of its components with respect to $\xi_k$, the $k^{\text{th}}$ component of $\boldsymbol{\xi}$. Also, note that the marginal model with respect to $\pi^R(\boldsymbol{\theta}) \propto 1$ is in this case

$$\begin{aligned} L^I(\sigma^2, \boldsymbol{\xi} \mid \boldsymbol{y}) &= \int_{\mathbb{R}^q} L(\boldsymbol{\eta} \mid \boldsymbol{y})\ \pi^R(\boldsymbol{\theta})\ d\boldsymbol{\theta} \\ &\propto (\sigma^2)^{-(n-q)/2} |\boldsymbol{\Sigma}|^{-1/2}\ |\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{-1/2}\ \exp\left\{-\frac{S^2_{\boldsymbol{\xi}}}{2\,\sigma^2}\right\} \qquad (2.5) \end{aligned}$$

where $S^2_{\boldsymbol{\xi}} = \boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}$, $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}\boldsymbol{P}$ and $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}$.

PROPOSITION 2.1 The reference prior $\pi^R(\boldsymbol{\eta})$ is of the form (2.4) with

$$a = 1 \qquad \text{and} \qquad \pi^R(\boldsymbol{\xi}) \propto |I_R(\boldsymbol{\xi})|^{1/2} \qquad (2.6)$$

where, with $\boldsymbol{W}_k = \dot{\boldsymbol{\Sigma}}^k \boldsymbol{Q}$, $k = 1, \ldots, r$,

$$I_R(\boldsymbol{\xi}) = \begin{pmatrix} n-q & \operatorname{tr}\boldsymbol{W}_1 & \operatorname{tr}\boldsymbol{W}_2 & \cdots & \operatorname{tr}\boldsymbol{W}_r \\ & \operatorname{tr}\boldsymbol{W}_1^2 & \operatorname{tr}\boldsymbol{W}_1\boldsymbol{W}_2 & \cdots & \operatorname{tr}\boldsymbol{W}_1\boldsymbol{W}_r \\ & & \ddots & \cdots & \vdots \\ & & & & \operatorname{tr}\boldsymbol{W}_r^2 \end{pmatrix}. \qquad (2.7)$$

PROOF: See Appendix B. ■

In the next proposition, we present the formulas for the two Jeffreys-type priors we will consider.

PROPOSITION 2.2 The independence Jeffreys prior, $\pi^{J1}$, obtained by assuming $\boldsymbol{\theta}$ and $(\sigma^2, \boldsymbol{\xi})$ *a priori* independent, and the Jeffreys-rule prior, $\pi^{J2}$, are of the form (2.4) with, respectively,

$$a = 1 \qquad \text{and} \qquad \pi^{J1}(\boldsymbol{\xi}) \propto |I_J(\boldsymbol{\xi})|^{1/2} \ , \tag{2.8}$$

and

$$a = 1 + \frac{q}{2} \qquad \text{and} \qquad \pi^{J2}(\boldsymbol{\xi}) \propto |\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{1/2} \ \pi^{J1}(\boldsymbol{\xi}) \ , \tag{2.9}$$

where $\boldsymbol{U}_k = \dot{\boldsymbol{\Sigma}}^k \boldsymbol{\Sigma}^{-1}$ and

$$I_J(\boldsymbol{\xi}) = \begin{pmatrix} n & \text{tr } \boldsymbol{U}_1 & \text{tr } \boldsymbol{U}_2 & \cdots & \text{tr } \boldsymbol{U}_r \\ & \text{tr } \boldsymbol{U}_1^2 & \text{tr } \boldsymbol{U}_1\boldsymbol{U}_2 & \cdots & \text{tr } \boldsymbol{U}_1\boldsymbol{U}_r \\ & & \ddots & \cdots & \vdots \\ & & & & \text{tr } \boldsymbol{U}_r^2 \end{pmatrix} . \tag{2.10}$$

PROOF: See Appendix B. ■

For priors of the form (2.4), it is possible to integrate explicitly the product of the likelihood and the prior over $(\sigma^2, \boldsymbol{\theta})$. Indeed, standard calculations yield (as long as $a > 1 - (n-q)/2$)

$$\int_{\mathbb{R}^q \times \mathbb{R}_+} L(\boldsymbol{\eta} \mid \boldsymbol{y}) \pi(\boldsymbol{\eta}) \ d\boldsymbol{\theta} \ d\sigma^2 = L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \pi(\boldsymbol{\xi})$$

where

$$L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto |\boldsymbol{\Sigma}|^{-1/2} |\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{-1/2} (S^2_{\boldsymbol{\xi}})^{-((n-q)/2+a-1)} \ . \tag{2.11}$$

It is clear that the posterior associated with the prior (2.4) is proper if and only if $0 < \int_\Omega L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \ \pi(\boldsymbol{\xi}) d \ \boldsymbol{\xi} < \infty$, where $\Omega \subset \mathbb{R}^r$ is the parametric space of $\boldsymbol{\xi}$.

It appears to be difficult to study analytically the properties of both the integrated likelihood $L^I(\boldsymbol{\xi} \mid \boldsymbol{y})$ and the function $\pi(\boldsymbol{\xi})$ — which we will often abuse terminology and refer to as the marginal prior for $\boldsymbol{\xi}$ —, in such a general setting. In the next section we will make the problem more amenable to analytical treatment by introducing a number of assumptions defining a more restrictive, yet with practical relevance, scenario.

# 3 Posterior Propriety

In this section, we show that the objective priors yield proper posteriors for an important special case. The proof is constructed as follows. In §3.1 we introduce the notion of a separable correlation function and the restrictions we assume on the linear model specifying the mean function. Next, under added assumptions, we describe the behavior of the integrated likelihood, whereas in §3.3 we achieve the same goal but for each of the objective priors determined in Section 2. Putting these results together, we are able to prove the theorem stated in §3.4. Throughout the present section, we will number the assumptions in order to easily be able to refer to them.

## 3.1 Separable Correlation Functions

Let $p \geq 2$. It is a simple consequence of Bochner's Theorem, cf. Cressie (1993), that if $c_i(x, x^\star) \equiv c_i(|x - x^\star|)$, $i = 1, \ldots, p$, are isotropic correlation functions in $\mathbb{R}$, then $c(\boldsymbol{x}, \boldsymbol{x}^\star) = \prod_{i=1}^{p} c_i(|x_i - x_i^\star|)$ is a valid correlation function in $\mathbb{R}^p$. Such correlation functions are called *separable*. If a separable correlation function is used and if furthermore the set of locations at which the process is observed forms a Cartesian product, it is easy to see that the correlation matrix of the data is the Kronecker product of the individual correlation matrices associated with each dimension. (Recall that the Kronecker product of two matrices, $\boldsymbol{A} = [a_{ij}]$ and $\boldsymbol{B}$, is defined by $\boldsymbol{A} \otimes \boldsymbol{B} = [a_{ij}\ \boldsymbol{B}]$.) It is essentially in this setting that we will study the analytical properties of the integrated likelihood and priors, along with establishing sufficient conditions for posterior propriety. We next make these statements precise.

We will henceforth assume that $r \equiv p \geq 2$ and that

$$c(\boldsymbol{x}, \boldsymbol{x}^\star \mid \boldsymbol{\xi}) = \prod_{k=1}^{p} \rho(|x_k - x_k^\star|; \xi_k) \ , \tag{3.1}$$

where $\rho$ is a valid correlation function in $\mathbb{R}$. Also, the set of locations at which the Gaussian process is observed will be assumed to follow a Cartesian product, i.e.

$$S = S_1 \times S_2 \times \cdots \times S_p \tag{3.2}$$

where $S_k = \{x_{1k}, \cdots, x_{n_k,k}\} \subset \mathbb{R}$, $\#S_k = n_k$, so that $\#S \equiv n = \prod_{k=1}^{p} n_k$. One should think about the parameter $\xi_k$ as representing the range parameter of the associated correlation structure; if other parameters are present, e.g. roughness parameters, those will be assumed known.

In terms of the mean structure, we will restrict attention exclusively to the situation where there is only one unknown parameter $\theta$, i.e. $q = 1$, and the design matrix (2.2) can be written in the form

$$\boldsymbol{X} = \boldsymbol{X}_1 \otimes \boldsymbol{X}_2 \otimes \cdots \otimes \boldsymbol{X}_p \tag{3.3}$$

with each $\boldsymbol{X}_k$ of dimension $n_k \times 1$. Within this framework we can find, for instance, the unknown constant mean case, i.e.,

$$q = 1 \quad \text{and} \quad \boldsymbol{\Psi}(\boldsymbol{x}) \equiv 1 \quad \Rightarrow \quad \mathbb{E}\, Y(\boldsymbol{x}) = \theta \quad \forall\, \boldsymbol{x} \ , \tag{3.4}$$

and $\boldsymbol{X} \equiv \mathbf{1}_n = \mathbf{1}_{n_1} \otimes \cdots \otimes \mathbf{1}_{n_p}$. This is in turn an instance of the more general situation described by

$$q = 1 \quad \text{and} \quad \boldsymbol{\Psi}(\boldsymbol{x}) \equiv \prod_{k \in A} \psi_k(x_k) \quad \Rightarrow \quad \mathbb{E}\, Y(\boldsymbol{x}) = \theta \prod_{k \in A} \psi_k(x_k) \quad \forall\, \boldsymbol{x} \ , \tag{3.5}$$

where $A \subset \{1, \ldots, n\}$. In this circumstance, $\boldsymbol{X}_k = (\psi_k(x_{ik}),\ i = 1, \ldots, n_k)'$, if $k \in A$; otherwise, $\boldsymbol{X}_k = \mathbf{1}_{n_k}$.

Above, and throughout the paper, when we write $\boldsymbol{X}_k = \mathbf{1}_{n_k}$ what we really mean is that $\boldsymbol{X}_k$ is a multiple of the $n_k$-dimensional vector of all ones, or, in a more mathematical sense, that $\mathbf{1}_{n_k}$ is in the space spanned by $\boldsymbol{X}_k$. The same comments apply to the statement $\boldsymbol{X}_k \neq \mathbf{1}_{n_k}$.

As we alluded to before, the Cartesian product structure of the design set, and the separable correlation function that we assumed, jointly allow for a very convenient Kronecker product expression for the correlation matrix of the data. To make that clear and to set up some notation as well, we have

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \cdots \otimes \boldsymbol{\Sigma}_p \equiv \bigotimes_{k=1}^{p} \Sigma_k \tag{3.6}$$

where $\boldsymbol{\Sigma}_k = [\rho(|x_{ik} - x_{jk}|; \xi_k)]_{i,j=1,\ldots,n_k}$, $k = 1, \ldots, p$, are the correlation matrices associated to each dimension. Note that each of these matrices is of dimension $n_k \times n_k$. This fact is further explored in Section 4. For simplicity, we will from here on use the shorthand exemplified above: we represent by $\otimes_{k=1}^{p} \boldsymbol{A}_k$ the matrix $\boldsymbol{A}_1 \otimes \cdots \otimes \boldsymbol{A}_p$.

The $\rho$ functions we will consider are quite general, but we will focus particular attention on the families of correlation functions that we list in Table 1 for future reference. For more details and additional references, consult Cressie (1993).

**Spherical—**

$$\rho(d;\xi) = \left[1 - \tfrac{3}{2}\ d\xi + \tfrac{1}{2}(d\xi)^3\right] I\{d\xi \le 1\};\ \xi > 0\ .$$

**Power Exponential—**

$$\rho(d;\xi) = \exp\left[-(d\xi)^\alpha\right];\ \xi > 0,\ \alpha \in (0,2]\ .$$

**Rational Quadratic—**

$$\rho(d;\xi) = \left[1 + (d\xi)^2\right]^{-\alpha};\ \xi > 0,\ \alpha > 0\ .$$

**Matérn—** $\mathscr{K}_\alpha(\cdot)$ is the modified Bessel function of the second kind and order $\alpha$.

$$\rho(d;\xi) = \tfrac{1}{2^{\alpha-1}\Gamma(\alpha)}(d\xi)^\alpha \mathscr{K}_\alpha(d\xi);\ \xi > 0,\ \alpha > 0\ .$$

Table 1: Families of correlation functions: $\xi$ denotes a range parameter; $\alpha$ refers to a roughness parameter; $d$ represents the Euclidean distance between two locations.

We next summarize the assumptions introduced in this paragraph.

**Assumptions**:

$\mathscr{A}1$ Separability of the correlation function: $r \equiv p \ge 2$ and

$$c(\boldsymbol{x}, \boldsymbol{x}^\star \mid \boldsymbol{\xi}) = \prod_{k=1}^{p} \rho(|x_k - x_k^\star|; \xi_k)\ ,$$

where $\rho$ is a valid correlation function in $\mathbb{R}$ .

$\mathscr{A}2$ Cartesian product of the design set:

$$S = S_1 \times S_2 \times \cdots \times S_p$$

where $S_k = \{x_{1k}, \cdots, x_{n_k,k}\} \subset \mathbb{R}$, $\#S_k = n_k$, so that $\#S \equiv n = \prod_{k=1}^{p} n_k$ .

$\mathscr{A}3$ Mean structure: $q = 1$ and

$$\boldsymbol{X} = \boldsymbol{X}_1 \otimes \boldsymbol{X}_2 \otimes \cdots \otimes \boldsymbol{X}_p$$

with each $\boldsymbol{X}_k$ of dimension $n_k \times 1$. ■

## 3.2 Behavior of the Integrated Likelihood

In this section, we will focus attention on the properties of the integrated likelihood. Nonetheless, the next two lemmas are key also to the series of results that will follow concerning the analytical behavior of the priors.

Lemma 3.1 Under assumptions $\mathscr{A}1$–$\mathscr{A}3$, we have

$$\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1} \ \boldsymbol{P} = \bigotimes_{k=1}^{p} \boldsymbol{\Sigma}_k^{-1} - \bigotimes_{k=1}^{p} \boldsymbol{\Phi}_k \ , \tag{3.7}$$

where $\boldsymbol{\Phi}_k = \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k (\boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k)^{-1} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1}$, and

$$|\boldsymbol{X}' \boldsymbol{\Sigma}^{-1} \boldsymbol{X}| = \prod_{k=1}^{p} \boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k \ . \tag{3.8}$$

Proof: See Appendix C.1. ■

Lemma 3.2 Under assumptions $\mathscr{A}1$–$\mathscr{A}3$, suppose $\rho(d;\xi)$ is a continuous function of $\xi$ for every $d \geq 0$ such that

- $\rho(d;\xi) = \rho^0(d\ \xi)$, where $\rho^0(\cdot)$ is a correlation function satisfying $\lim_{u\to\infty} \rho^0(u) = 0$;
- as $\xi_k \to 0$, each of the correlation matrices $\boldsymbol{\Sigma}_k$ satisfies

$$\boldsymbol{\Sigma}_k = \mathbf{1}_{n_k} \mathbf{1}_{n_k}' + \nu(\xi_k)(\boldsymbol{D}_k + o(1)) \ , \quad k = 1, \ldots, p \tag{3.9}$$

  for some continuous non-negative function $\nu(\cdot)$ and fixed *nonsingular* matrix $\boldsymbol{D}_k$.

- Above, and for $k = 1, \ldots, p$, $\boldsymbol{D}_k$ should satisfy $\mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \neq 0$, $\boldsymbol{X}_k' \boldsymbol{D}_k^{-1} \boldsymbol{X}_k \neq 0$, and, if $\boldsymbol{X}_k \neq \mathbf{1}$,

$$\mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \neq \mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1} \boldsymbol{X}_k (\boldsymbol{X}_k' \boldsymbol{D}_k^{-1} \boldsymbol{X}_k)^{-1} \boldsymbol{X}_k' \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \ . \tag{3.10}$$

Define the quantities

$$\boldsymbol{F}_k = \frac{\boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1}}{(\mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k})^2} \tag{3.11}$$

$$\boldsymbol{G}_k = \boldsymbol{D}_k^{-1} - \frac{\boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1}}{\mathbf{1}_{n_k}' \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k}} \tag{3.12}$$

$$\boldsymbol{H}_k = \frac{\boldsymbol{G}_k \boldsymbol{X}_k \boldsymbol{X}_k' \boldsymbol{G}_k}{\boldsymbol{X}_k' \boldsymbol{G}_k \boldsymbol{X}_k} \ . \tag{3.13}$$

In this setting, we have that, as $\xi_k \to 0$, $k = 1, \dots, p$

$$\boldsymbol{\Sigma}_k^{-1} = \boldsymbol{G}_k(1 + o(1))/\nu(\xi_k) \tag{3.14}$$

$$|\boldsymbol{\Sigma}_k| = [\nu(\xi_k)]^{n_k - 1}|\boldsymbol{D}_k|(\boldsymbol{1}'_{n_k}\boldsymbol{D}_k\boldsymbol{1}_{n_k})\ (1 + o(1)) \tag{3.15}$$

$$\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k = \begin{cases} 1 + o(1) & \text{if } \boldsymbol{X}_k = \boldsymbol{1} \\ \left[\boldsymbol{X}'_k\boldsymbol{D}_k^{-1}\boldsymbol{X}_k - \dfrac{\boldsymbol{1}'\boldsymbol{D}_k^{-1}\boldsymbol{X}_k\boldsymbol{X}'_k\boldsymbol{D}_k^{-1}\boldsymbol{1}}{\boldsymbol{1}'\boldsymbol{D}_k^{-1}\boldsymbol{1}}\right](1 + o(1))/\nu(\xi_k) & \text{otherwise} \end{cases} \tag{3.16}$$

$$\boldsymbol{\Phi}_k = \begin{cases} \boldsymbol{F}_k(1 + o(1)) & \text{if } \boldsymbol{X}_k = \boldsymbol{1} \\ \boldsymbol{H}_k(1 + o(1))/\nu(\xi_k) & \text{otherwise .} \end{cases} \tag{3.17}$$

Proof: See Appendix C.1. ■

The following result describes the behavior of the integrated likelihood (2.11) in the present setting and under the assumptions of the previous lemma.

Proposition 3.3 Under the conditions of Lemma 3.2, and assuming a prior of the form (2.4), $L^I(\boldsymbol{\xi} \mid \boldsymbol{y})$ is a continuous function of $\boldsymbol{\xi}$ given by the expression

$$L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto \prod_{k=1}^{p}\left\{|\boldsymbol{\Sigma}_k|^{-n_{(k)}/2}(\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k)^{-1/2}\right\} \times [\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}]^{-(n-3+2a)/2} \tag{3.18}$$

where $\boldsymbol{Q}$ is given by (3.7) and $n_{(k)} \equiv \prod_{i \neq k} n_i$.

Additionally, let $A = \{k \in \{1, \dots, n\} : \boldsymbol{X}_k \neq \boldsymbol{1}\}$ and $B$ a nonempty, but otherwise arbitrary, subset of $\{1, \dots, n\}$. Denote by $\bar{E}$ the complement of a set $E$. Then,

(a) as $\xi_k \to 0$, $k \in B$ and $\xi_k \to \infty$, $k \in \bar{B}$,

$$L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto \prod_{k \in B}\nu(\xi_k)^{(n_k-3+2a)/2} \prod_{k \in A \cap B}\nu(\xi_k)^{1/2}\ (1 + g(\boldsymbol{\xi})) \tag{3.19}$$

with $g(\boldsymbol{\xi}) \to 0$;

(b) as $\xi_k \to 0$, $k \in B$ and $0 < \delta_k < \xi_k < M_k < \infty$, $k \in \bar{B}$,

$$\begin{aligned} L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto \prod_{k \in B}\nu(\xi_k)^{(n_k-3+2a)/2} \prod_{k \in A \cap B}\nu(\xi_k)^{1/2} \times \\ \times (1 + g(\{\xi_k, k \in B\})) \times h(\{\xi_k, k \in \bar{B}\})\ , \end{aligned} \tag{3.20}$$

with $g(\cdot) \to 0$ and $h(\cdot)$ continuous on $\{\delta_k < \xi_k < M_k, k \in \bar{B}\}$.

Proof: See Appendix C.1. ■

## 3.3 Behavior of the Priors

In the next two results, properties of both Jeffreys' priors and of the reference prior are described. They are based on a series of assumptions that we list below, and form (except for $\mathscr{A}8$) a subset of those introduced in Berger et al. (2001) in their study of the asymptotic properties of the priors. According to this paper, the families of correlation functions listed on Table 1 (virtually always) satisfy these properties. (For example, beware of the fact that for the power exponential with $\alpha = 2$ and more than 3 equally spaced points, it is not the case that $\boldsymbol{D}$ is invertible. It is if $\alpha \in (0, 2)$.) Note also how these assumptions imply those of Lemma 3.2. Below, we denote by $\dot{\boldsymbol{\Sigma}}_k$ the matrix of derivatives of $\boldsymbol{\Sigma}_k$ with respect to $\xi_k$, $k = 1, \ldots, p$.

**Assumptions**: Suppose that $\rho(d; \xi) = \rho^0(d\ \xi)$, where $\rho^0(\cdot)$ is a correlation function satisfying $\lim\limits_{u \to \infty} \rho^0(u) = 0$, is a continuous function of $\xi$ for any $d > 0$ and that, for $k = 1, \ldots, p$,

$\mathscr{A}4$ As $\xi_k \to 0$, there are fixed matrices, $\boldsymbol{D}_k$, *nonsingular*, and $\boldsymbol{D}_k^\star$; differentiable functions $\nu(\cdot)$ $(> 0)$ and $\omega(\cdot)$; and a matrix function $\boldsymbol{R}_k(\cdot)$ that is differentiable as well, such that

$$\boldsymbol{\Sigma}_k = \boldsymbol{1}_{n_k}\boldsymbol{1}_{n_k}' + \nu(\xi_k)\boldsymbol{D}_k + \omega(\xi_k)\boldsymbol{D}_k^\star + \boldsymbol{R}_k(\xi_k) \tag{3.21}$$

$$\dot{\boldsymbol{\Sigma}}_k = \nu'(\xi_k)\boldsymbol{D}_k + \omega'(\xi_k)\boldsymbol{D}_k^\star + \tfrac{\partial}{\partial \xi_k}\boldsymbol{R}_k(\xi_k)\ . \tag{3.22}$$

$\mathscr{A}5$ The functions $\nu$, $\omega$ and $\boldsymbol{R}_k$ further satisfy, as $\xi_k \to 0$ and with $||[a_{ij}]||_\infty = \max_{ij}\{|a_{ij}|\}$

$$\begin{array}{ll} \dfrac{\omega(\xi_k)}{\nu(\xi_k)} \to 0 & \qquad \dfrac{\omega'(\xi_k)}{\nu'(\xi_k)} \to 0 \\ \dfrac{||\boldsymbol{R}_k(\xi_k)||_\infty}{\nu(\xi_k)} \to 0 & \qquad \dfrac{||\frac{\partial}{\partial \xi_k}\boldsymbol{R}_k(\xi_k)||_\infty}{\nu'(\xi_k)} \to 0\ . \end{array}$$

$\mathscr{A}6$ Above, $\boldsymbol{D}_k$ satisfies $\boldsymbol{1}_{n_k}'\boldsymbol{D}_k^{-1}\boldsymbol{1}_{n_k} \neq 0$, $\boldsymbol{X}_k'\boldsymbol{D}_k^{-1}\boldsymbol{X}_k \neq 0$, and, if $\boldsymbol{X}_k \neq \boldsymbol{1}$,

$$\boldsymbol{1}_{n_k}'\boldsymbol{D}_k^{-1}\boldsymbol{1}_{n_k} \neq \boldsymbol{1}_{n_k}'\boldsymbol{D}_k^{-1}\boldsymbol{X}_k(\boldsymbol{X}_k'\boldsymbol{D}_k^{-1}\boldsymbol{X}_k)^{-1}\boldsymbol{X}_k'\boldsymbol{D}_k^{-1}\boldsymbol{1}_{n_k}\ .$$

$\mathscr{A}7$ $[\mathrm{tr}(\dot{\boldsymbol{\Sigma}}_k)^2]^{1/2}$ is integrable at infinity.

$\mathscr{A}8$ $n_k \geq 2$. ■

We start with the reference prior and then proceed to the Jeffreys-type priors.

Proposition 3.4 For the reference prior (2.6), and under assumptions $\mathscr{A}1$–$\mathscr{A}3$, there are functions $\pi_k^R(\xi_k)$, $k = 1, \ldots, p$ such that

$$\pi^R(\boldsymbol{\xi}) \leq \prod_{k=1}^{p} \pi_k^R(\xi_k) \ , \tag{3.23}$$

where, as $\xi_k \to 0$, $k = 1, \ldots, p$, and under added assumptions $\mathscr{A}4$–$\mathscr{A}8$,

$$\pi_k^R(\xi_k) \propto \left| \frac{\nu'(\xi_k)}{\nu(\xi_k)} \right| (1 + o(1)) \ , \tag{3.24}$$

and $\pi_k^R$ is integrable at infinity.

Proof: See Appendix C.2. ■

Proposition 3.5 Under assumptions $\mathscr{A}1$–$\mathscr{A}3$, for the independence Jeffreys prior, $\pi^{J1}$, given by (2.8), and the Jeffreys-rule prior, $\pi^{J2}$, given by (2.9), there are functions $\pi_k^{Ji}(\xi_k)$, $i = 1, 2$, $k = 1, \ldots, p$, such that

$$\pi^{Ji}(\boldsymbol{\xi}) \leq \prod_{k=1}^{p} \pi_k^{Ji}(\xi_k) \tag{3.25}$$

where, under added assumptions $\mathscr{A}4$–$\mathscr{A}8$ and for $k = 1, \ldots, p$, we have, as $\xi_k \to 0$

$$\pi_k^{Ji}(\boldsymbol{\xi_k}) \propto \frac{|\nu'(\xi_k)|}{|\nu(\xi_k)|^{\alpha_k}} \ (1 + o(1)) \tag{3.26}$$

with

$$\alpha_k = \begin{cases} 1 \ , & \text{if } i = 1 \text{ or } (i = 2 \text{ and } \boldsymbol{X}_k = \mathbf{1}) \\ 3/2 \ , & \text{if } i = 2 \text{ and } \boldsymbol{X}_k \neq \mathbf{1} \ . \end{cases} \tag{3.27}$$

Also, $\pi_k^{Ji}$ is integrable at infinity.

Proof: See Appendix C.2. ■

## 3.4 Results on Posterior Propriety

In this section we investigate the propriety of the formal posterior associated with a general prior of the form (2.4), in the scenario described throughout the previous sections. We end with a result that specifically addresses the case of both Jeffreys' priors and the reference prior.

Theorem 3.6 Under the set of assumptions $\mathscr{A}1$–$\mathscr{A}8$, the posterior associated with the general prior (2.4) — where $\pi(\boldsymbol{\xi})$ is either $\pi^R$, $\pi^{J1}$ or $\pi^{J2}$ —, is proper as long as

$$a > 1/2 \ . \tag{3.28}$$

Proof: See Appendix C.3. ■

If we restrict attention to the instances of the general prior that are of particular interest, then we have the following:

Corollary 3.7 Under assumptions $\mathscr{A}1$–$\mathscr{A}8$, the reference, the independence Jeffreys and the Jeffreys-rule priors yield proper posteriors.
Proof: Just recall that $a = 1$ in the case of the reference and independence Jeffreys prior, and that for the Jeffreys-rule $a = 1 + q/2$, where by assumption $q = 1$. ■

It is interesting to understand what makes the multidimensional problem different from the unidimensional one, the case covered in Berger et al. (2001). As we mentioned in the Introduction, the reason why posterior propriety is difficult to achieve in the unidimensional case has to do with the behavior of the integrated likelihood near the origin. For example, if $p = 1$ and $\boldsymbol{X} = \mathbf{1}$, we have, as $\xi \to 0$

$$L^I(\xi \mid \boldsymbol{y}) = O([\nu(\xi)]^{1-a})$$

which is independent of the sample size. Roughly speaking, the first and last factors of (2.11) produce powers of $\nu(\xi)$ depending on the sample size that essentially cancel.

The key formula in the multidimensional case is

$$\left| \bigotimes_{k=1}^{p} \boldsymbol{\Sigma}_k \right| = \prod_{k=1}^{p} |\boldsymbol{\Sigma}_k|^{n_{(k)}}$$

where $n_{(k)} = \prod_{i \neq k} n_i$. Considering (3.15), it is clear that, as $\xi_k \to 0$,

$$|\boldsymbol{\Sigma}|^{-1/2} \propto \left[ \prod_{i \neq k}^{p} |\boldsymbol{\Sigma}_i|^{-n_{(i)}/2} \right] \ \nu(\xi_k)^{-n/2} \ \nu(\xi_k)^{n_{(k)}/2} \ (1 + o(1)) \ .$$

The first factor involving $\nu(\xi_k)$ will again essentially cancel, but we still keep the second, which explains the different behaviors. For small $\xi_k$ the prior is not that relevant in determining posterior propriety in the multidimensional case, whereas in the unidimensional one it is of capital importance.

# 4 Computation

As we already mentioned, there is considerable need in the area of computer model validation for default prior specifications and also for efficient computational schemes. In this section, we address several computing issues, and in particular that of Bayesian learning in the presence of the objective priors derived in this paper. We present an example, taken from the field of computer model validation, to illustrate some of the proposed solutions.

We now return to the general setting of Section 2, and in particular, unless otherwise noted, $\boldsymbol{\xi}$ represents a general $r$-dimensional vector involved in the parametric formulation of the correlation function.

## 4.1 General Remarks

Evaluating any of the objective priors at one particular value of $\boldsymbol{\xi}$ is a computationally intensive task. The reference prior is the most computationally intensive of all three, followed by the independence Jeffreys prior and by the Jeffreys rule prior. In order to convince oneself of that, it suffices to consider the calculations involved in computing each of the entries of the matrices (2.7) and (2.10). For similar reasons, it is clear that the likelihood function (2.3) is computationally less expensive than either of the integrated likelihoods given by (2.5) and (2.11).

One should note that some of the assumptions of Section 3 have considerable potential impact on the computational side of the problem. When satisfied, the separability assumption ($\mathscr{A}1$) paired with the Cartesian product structure of the design set ($\mathscr{A}2$), which implies formula (3.6), can be exploited in order to tremendously simplify and speed up computation. Indeed, the most expensive (and possibly numerically unstable) calculations of this problem are certainly computing the inverse of the correlation matrix and its determinant. Formulas (II) and (III) of Appendix A essentially state that, in this case, we only have to compute the inverse and the determinant of each of the $p$ matrices $\boldsymbol{\Sigma}_k$ of dimension $n_k \times n_k$, and not for the $n \times n$ correlation matrix $\boldsymbol{\Sigma}$. Even the Cholesky decomposition of $\boldsymbol{\Sigma}$ can be obtained from that of the $\boldsymbol{\Sigma}_k$, as it is easy to see. This allows for much more freedom on the number of points at which one observes the stochastic process, as explored in Bayarri et al. (2002) when dealing with functional output of a computer model.

## 4.2 Maximum Likelihood Estimates

The calculation of maximum likelihood (ML) estimates turns out to be useful at least in two manners. On the one hand, from an applied perspective, it is often the case, and especially in the field of computer model validation, that there is not enough information in the data to learn about all the parameters that specify the correlation structure. In particular, it is quite plausible that there is very little information in the data about parameters that characterize geometric properties of the process, the so-called roughness parameters. If one is not willing to impose restrictions on those characteristics based on expert information, an alternative solution is to fix those parameters at some data-determined value, in particular at some kind of maximum likelihood estimate. The other circumstance where these estimates can be useful is in devising sampling schemes that require little input from the user, as we will see in §4.4.

The question of which kind of estimate should one compute is relevant, as here we have at our disposal explicit formulas for three distinct (integrated) likelihood functions.

There is considerable empirical evidence supporting the fact that, in general, maximum likelihood estimates derived from integrated likelihoods tend to be more stable and also more meaningful, so that one would in principle discard the idea of maximizing the joint likelihood (2.3). Also, these estimates are not as useful in the context of devising sampling schemes.

Maximizing the integrated likelihood (2.5) has some advantages over maximizing (2.11). First of all, it gives rise to considerable simpler formulas for the gradient and the associated expect information matrix. Indeed, it is easy to see from inspection of the proof of Proposition 2.1 that

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\ln L^I(\sigma^2,\boldsymbol{\xi}) &= \frac{1}{2}\,\left(\frac{1}{\sigma^4}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y} - \frac{1}{\sigma^2}(n-q)\right)\\ \frac{\partial}{\partial\xi_k}\ln L^I(\sigma^2,\boldsymbol{\xi}) &= \frac{1}{2}\,\left(\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y} - \operatorname{tr}\boldsymbol{W}_k\right)\ , \quad k=1,\ldots,r\ .\end{aligned}$$

and that the associated Fisher information matrix is

$$I(\sigma^2,\boldsymbol{\xi}) = \frac{1}{2}\begin{pmatrix} \frac{n-q}{\sigma^4} & \frac{1}{\sigma^2}\ \operatorname{tr}\boldsymbol{W}_1 & \frac{1}{\sigma^2}\ \operatorname{tr}\boldsymbol{W}_2 & \cdots & \frac{1}{\sigma^2}\ \operatorname{tr}\boldsymbol{W}_r \\ & \operatorname{tr}\boldsymbol{W}_1^2 & \operatorname{tr}\boldsymbol{W}_1\boldsymbol{W}_2 & \cdots & \operatorname{tr}\boldsymbol{W}_1\boldsymbol{W}_r \\ & & \ddots & \cdots & \vdots \\ & & & & \operatorname{tr}\boldsymbol{W}_r^2 \end{pmatrix}\ . \qquad (4.1)$$

This allows for a very simple optimization algorithm to be used in computing the associated estimate, namely Fisher's Scoring method, a variant of the Newton-Raphson method that results from approximating the Hessian of the logarithm of integrated likelihood by its expected value. To be more precise, the $(s+1)$-th iterate of the numerical method is given by

$$\begin{aligned}(\sigma^2)^{(s+1)} &= \frac{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}}{n-q} \\ \boldsymbol{\xi}^{(s+1)} &= \boldsymbol{\xi}^{(s)} + \lambda\ [I(\boldsymbol{\xi}^{(s)})]^{-1}\ \left.\frac{\partial\ \ln L^I(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}\right|_{\boldsymbol{\xi}=\boldsymbol{\xi}^{(s)}}\ ,\end{aligned}$$

where $I(\boldsymbol{\xi})$ results from (4.1) by dropping first row and column. The quantity $\lambda$ is the step size of the algorithm.

There are some drawbacks to this simple numerical method. First, it is possible that, initially, some iterates happen to lie outside the parameter space. We surround this problem by simply saying that $\xi_k^{(s+1)} = \xi_k^{(s)}$ whenever the $k$-th component of $\boldsymbol{\xi}^{(s+1)}$ happens to not belong to the corresponding parameter space. Also, it is well known that Newton-Raphson-type methods are quite sensitive to the starting points, sometimes becoming trapped in local maxima and sometimes simply not converging. This unfortunately is not easy to solve. Some experimentation and tuning of $\lambda$ are required in order to assure convergence.

Getting an estimate of $\boldsymbol{\xi}$ by maximizing (2.11) is not as attractive because of the fact that the formulas are computationally more involved. For example, there does not seem to be a closed form expression for the associated expected information matrix. Also, in the examples we have dealt with involving the power exponential family of correlation functions, we did not notice any improvements over the simpler method. The availability of an estimate of the variance will also be useful in the next section, where we will describe an Empirical Bayes alternative to the objective priors.

## 4.3 An Empirical Bayes Alternative

The availability of ML estimates allows for an alternative prior specification to be considered, one that is data-dependent. The basic idea is to place independent Exponential priors centered at a multiple of the ML estimates on the precision (reciprocal of variance) and on the components of vector $\boldsymbol{\xi}$, in effect specifying an Empirical Bayes procedure. The constant that multiplies the ML estimate in order to get the mean of the prior can be determined by experimentation, making sure that the effect of the prior on the posterior is relatively small. In other words, making sure that the prior is relatively flat in the region of the parametric space where the posterior mass accumulates. For an example, consider §4.5.

The reasons for entertaining this alternative method are as follows. First, having said that the evaluation of the objective priors is computationally intensive, it makes sense to consider this much simpler method and compare the results. Second, the objective priors we derived are all based on formal methods, and hence it is nice to be able to compare them with seemingly more intuitive approaches to the problem. Lastly, this comparison is even more important as people often in practice resort to similar stategies as an alternative to objective priors derived by formal methods.

## 4.4 Sampling from the Posterior

No matter which prior specification one uses — any of the objective priors or the Empirical Bayes method —, one can sample exactly from the full conditional of $\boldsymbol{\theta}$ and of $\sigma^2$ — these are respectively Gaussian and Inverse-Gamma:

$$\begin{aligned}\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{\xi}, \sigma^2 &\sim \mathsf{N}((\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y},\ \sigma^2(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1})\\ \sigma^2 \mid \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\xi} &\sim \Gamma^{-1}(n/2 + a - 1,\ \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}/2 + r_{\sigma^2})\end{aligned}$$

where, by convention, $a = 2$ in the case of the Empirical Bayes. Recall that in the case of the reference and independence Jeffreys priors $a = 1$, whereas in the Jeffreys-rule, $a = 1 + q/2$. Also, $r_{\sigma^2}$ is the rate of the Exponential prior in the Empirical Bayes case, and should be set to zero otherwise.

Then, one has the choice between sampling from the marginal posterior $[\boldsymbol{\xi} \mid \boldsymbol{y}]$ or from the full conditional $[\boldsymbol{\xi} \mid \boldsymbol{y}, \sigma^2, \boldsymbol{\theta}]$. We adopt this latter strategy since the integrated likelihood is computationally more expensive than the full likelihood and, based on the in the experiments we conducted, there are no apparent gains in terms of mixing in adopting the other strategy.

At this point we have also the option of drawing $\boldsymbol{\xi}$ as a block or not. To this end, notice that as long as one updates one of the components of $\boldsymbol{\xi}$ one has basically to recompute the likelihood, as it does not factor in any useful way. From that perspective, it is more efficient therefore to sample $\boldsymbol{\xi}$ as a block. On the other hand, it is reasonable to expect some correlation between the $\xi_k$, and sampling $\boldsymbol{\xi}$ as a block would at least reduce it.

The question is, can we get a proposal, in a more or less automatic fashion, that allows for this? The ML estimate and the expected information matrix we described in §4.2 are useful in accomplishing this goal. Although the resulting sampling scheme is certainly not new and has been widely used, it seems to work well in the examples we examined so far.

The idea is to consider a reparameterization

$$\boldsymbol{\zeta} = \boldsymbol{g}(\sigma^2,\ \boldsymbol{\xi})$$

so that the new parameter space is $\mathbb{R}^{r+1}$. For this alternative parameterization, it is easy to calculate both the marginal ML estimate, $\hat{\boldsymbol{\zeta}}$, and the associated expected information evaluated at the ML estimate, $I_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}})$, given that we can compute those quantities for the original parameterization. Consider the partition of $I_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}})$ given by ($u$ is a scalar, $\boldsymbol{v}$ is an $r \times 1$ vector)

$$I_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\zeta}}) = \begin{bmatrix} u & \boldsymbol{v}' \\ \boldsymbol{v} & \boldsymbol{A} \end{bmatrix} \tag{4.2}$$

and define

$$\hat{\boldsymbol{V}} = \boldsymbol{A} - \boldsymbol{v}\ \boldsymbol{v}'/u\ . \tag{4.3}$$

We perform a Metropolis step in the alternative parameterization with a $t$-density proposal centered at the previous iteration, with $d$ degrees of freedom and scale matrix $c^2\ \hat{\boldsymbol{V}}$.

The only quantities that need to be specified in order to implement this Markov chain Monte Carlo (MCMC) method is the number of degrees of freedom and the multiple $c$. The number of degrees of freedom is not an issue, one can even use a Normal instead of a $t$. As for $c$, one can always experiment with a few values, and a reasonable starting guess is $c \approx 2.4/\sqrt{r}$.

## 4.5 An Example

For illustrative purposes, we are going to consider a simplified version of a problem analyzed in detail in Bayarri et al. (2002). In this paper, they consider the analysis of a computer model that simulates the crash of prototype vehicles against a barrier, recording the velocity curve from the point of impact until the vehicle stops. For our purposes, we will imagine that if we input an impact velocity and an instant in time, the computer model will return the velocity of the vehicle at that point in time after impact.

The data consists of the output of the computer model at 19 values of $t$ and 9 initial velocities $v$, which corresponds to an 171-point design set that follows a Cartesian product. If we subtract from each of the individual curves the initial velocity, what happens is that all the curves will start at zero and decay at a rate roughly proportional to the initial velocity. For this transformed data, which is plotted on Figure 1, it seems reasonable to consider the mean function $\mathbb{E}\, Y(v,t) = v\ t\ \theta$, which in the notation of Section 2 translates into $q = 1$ and $\psi(v,t) = v\ t$. In particular, this problem satisfies assumptions $\mathscr{A}2$ and $\mathscr{A}3$.

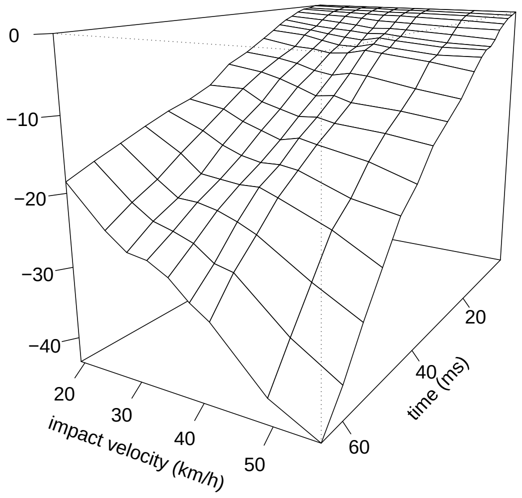

Figure 1: Relative velocity as a function of time after impact and of impact velocity.

In terms of correlation function, we will assume a two-dimensional separable power exponential function, with the roughness parameters fixed at 2, and the following parameterization

$$c((t_1,v_1),(t_2,v_2)) = \exp(-\beta_t\ |t_1 - t_2|^2)\ \exp(-\beta_v\ |v_1 - v_2|^2)\ . \tag{4.4}$$

Figure 2 shows estimates of the posterior distribution associated with each of the priors that we have introduced in the paper. The Empirical Bayes corresponds to centering the Exponential priors at 10 times the computed ML estimates, and the dotted lines on that figure correspond to those densities. Figure 3 shows the same samples now in the form of box-plots. The sampling mechanism was implemented exactly as detailed above, and we used a $t$ density with 3 degrees of freedom and $c = 1.7$. The transformation $\boldsymbol{g}$ that we used was the logarithmic one. The acceptance rate was roughly 0.27 in all cases, and the results correspond to 50,000 iterations, minus 500 that were discarded as burn-in. The chains seem to reach stationarity very quickly, and 50,000 iterations is certainly more than we actually need for reliable inference.

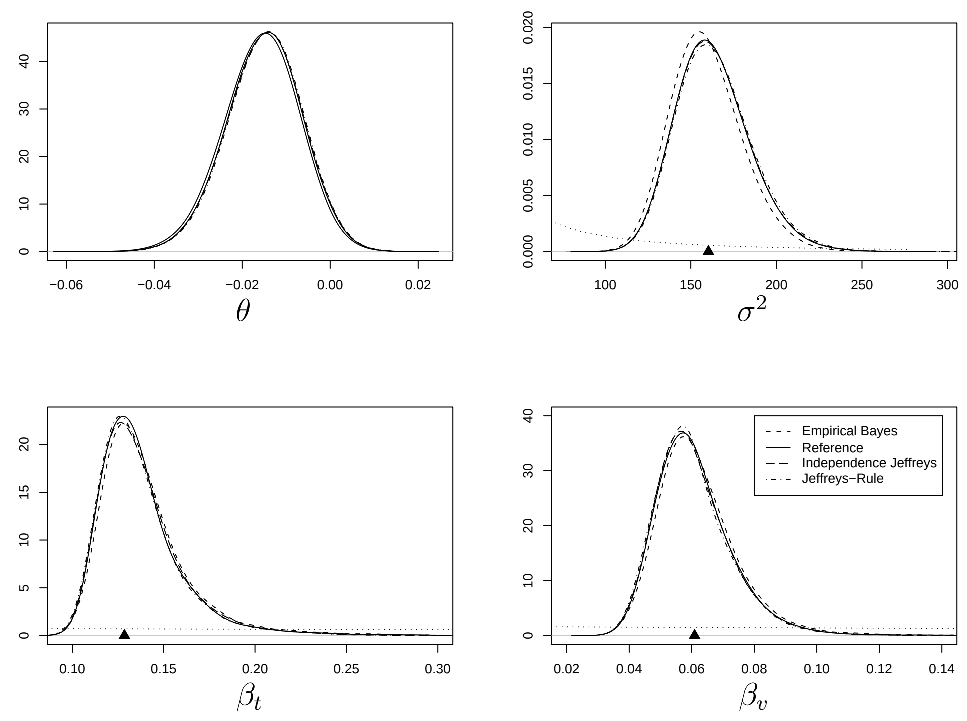

Figure 2: Smoothed histograms of samples drawn from the posterior distribution associated with each of the priors. Dotted lines correspond to Exponential priors of Empirical Bayes, and the triangles indicate the marginal likelihood estimates.

For these data, the answers are quite robust with respect to the type of default prior chosen. In the next section we study finer details of the methods proposed in the paper, in particular by studying the frequentist properties of the resulting Bayesian procedures.

# 5 Comparison of the Priors

When faced with more than one valid default prior specification strategy, it is often argued that one way to distinguish among these is by studying the frequentist properties of the

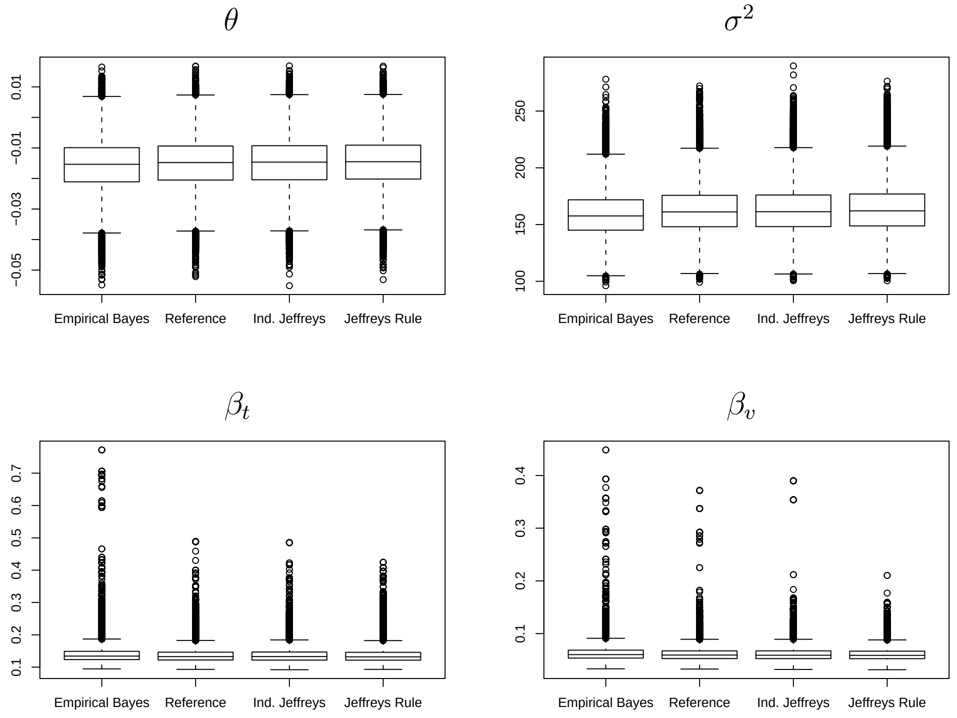

Figure 3: Box-plots of samples drawn from the posterior distribution associated with each of the priors.

resulting Bayesian inferential procedures. This usually takes the form of computing the frequentist coverage of the $(1-\gamma)\times 100\%$ equal-tailed Bayesian credible interval for one of the parameters of interest. The closer to the nominal level is this frequentist coverage, the 'better' is the prior.

This frequentist coverage is a computationally very intensive type of calculation since, if no other problem-specific method is available, one has to proceed as follows to get an estimate of such probability. Having fixed a value for the vector of unknown parameters, one has to generate data from the ensuing model and calculate the equal-tailed credible interval. Often, as in the present case, this is accomplished by generating an MCMC sample from the posterior and obtaining the associated $\gamma/2$- and $(1-\gamma/2)$-quantiles for the parameter of interest. Finally, one records the realization of the Bernoulli experiment in which the success corresponds to the parameter of interest lying within the limits of the computed credible interval. These steps have to be repeated a large number of times in order to get a reliable estimate of this success probability, which is typically large in magnitude. Obviously, this is only feasible if a nearly automated MCMC algorithm is available, which is presently the case.

Our study was conducted in the context of the power exponential correlation function with $p = 2$, parameterized as in (4.4), with the roughness parameters ($\alpha_1$ and $\alpha_2$) treated as known. For a $5 \times 5$ equally spaced grid in $[0, 1] \times [0, 1]$, we considered two options for the mean structure: an unknown constant term that we fixed at $\theta = 1$ and $\mathbb{E}[Y(\boldsymbol{x}) \mid \boldsymbol{\eta}] = \theta x_1$, where again $\theta$ was fixed at 1. Several choices for the other parameters were considered, and we simulated 3000 draws from the ensuing model. Each Markov chain consisted of 15000 iterations, the first 100 being discarded as burn-in. We calculated 95% credible intervals for $\sigma^2$, $\theta$ and $\beta_1$, and partial results are summarized in Table 2. Along with the estimate of the coverage probability, we present also an estimate of the expected length of the resulting credible interval and the standard deviation associated with the estimate.

A clear conclusion to extract from the results of these experiments is the comparatively poor behavior of the Empirical Bayes method, which resulted from centering the priors at 10 times the computed ML estimate. This conclusion is particularly important since, whenever a formal objective prior is not available, practitioners tend to resort to similar strategies to produce so-called “vague” or “diffuse” priors. On the basis of this study, one might argue against that type of approach to objective Bayesian analysis.

It is also possible to argue against the use of the Jeffreys-rule prior, and the reason for its inferior behavior has to do with the power $a = 1 + q/2$ in its formulation. This was already reported in Berger et al. (2001): this type of prior is known to add spurious degrees of freedom to uncertainty statements.

The present study is not quite decisive about how the reference prior and the independence Jeffreys prior compare, and one may conclude that they present virtually equivalent performance. On the basis that the reference prior is computationally more demanding, we recommend the independence Jeffreys prior as a default prior for the problem at hand.

# 6 Acknowledgments

The author would like to thank Professor James O. Berger for his guidance and for numerous insightful discussions.

| | Coverage Prob. | Expected Length | Std. Dev. |
|---|---|---|---|
| Empirical Bayes | 0.857 | 2.904 | 0.005 |
| Reference Prior | 0.956 | 11.354 | 0.003 |
| Independence Jeffreys Prior | 0.954 | 11.698 | 0.003 |
| Jeffreys-Rule Prior | 0.946 | 6.334 | 0.003 |

Interval for $\sigma^2$; parameter vector fixed at (1.5,3.2,3.6,1.5,1.7); $\mathbb{E}Y = 1$.

| | Coverage Prob. | Expected Length | Std. Dev. |
|---|---|---|---|
| Empirical Bayes | 0.805 | 3.156 | 0.007 |
| Reference Prior | 0.952 | 7.434 | 0.005 |
| Independence Jeffreys Prior | 0.956 | 7.410 | 0.005 |
| Jeffreys-Rule Prior | 0.902 | 5.265 | 0.012 |

Interval for $\theta$; parameter vector fixed at (1.5,3.2,3.6,1.5,1.7); $\mathbb{E}Y = \theta \equiv 1$.

| | Coverage Prob. | Expected Length | Std. Dev. |
|---|---|---|---|
| Empirical Bayes | 0.840 | 1.077 | 0.007 |
| Reference Prior | 0.948 | 0.950 | 0.003 |
| Independence Jeffreys Prior | 0.955 | 0.931 | 0.004 |
| Jeffreys-Rule Prior | 0.928 | 1.027 | 0.005 |

Interval for $\theta$; parameter vector fixed at (1.5,0.2,0.6,1.5,1.7); $\mathbb{E}Y = \theta x_1 \equiv x_1$.

| | Coverage Prob. | Expected Length | Std. Dev. |
|---|---|---|---|
| Empirical Bayes | 0.826 | 1.145 | 0.007 |
| Reference Prior | 0.946 | 0.901 | 0.004 |
| Independence Jeffreys Prior | 0.954 | 0.890 | 0.004 |
| Jeffreys-Rule Prior | 0.919 | 1.059 | 0.005 |

Interval for $\beta_1$; parameter vector fixed at (1.5,0.2,0.6,1.5,1.7); $\mathbb{E}Y = 1$.

| | Coverage Prob. | Expected Length | Std. Dev. |
|---|---|---|---|
| Empirical Bayes | 0.797 | 1.815 | 0.007 |
| Reference Prior | 0.948 | 1.042 | 0.004 |
| Independence Jeffreys Prior | 0.948 | 0.996 | 0.004 |
| Jeffreys-Rule Prior | 0.916 | 1.255 | 0.005 |

Interval for $\beta_1$; parameter vector fixed at (1.5,0.2,0.6,1.0,1.2); $\mathbb{E}Y = 1$.

Table 2: Coverage probability of 95% credible intervals. The order of the fixed parameter vector is $(\sigma^2, \beta_1, \beta_2, \alpha_1, \alpha_2)$.

## APPENDIX

# A Auxiliary Facts

Throughout the appendix, we will repeatedly use the following results, often without referring explicitly to them. They are either standard propositions, whose proof can easily be found (cf. e.g. Harville (1997) for matrix algebra and Tong (1990) for multivariate (normal) theory), or easy extensions of such results. Therefore, proofs will be omitted at this point.

I — If for $i = 1, \dots, p$ the product $\boldsymbol{A}_i\boldsymbol{B}_i$ is possible, then we have

$$\left(\bigotimes_{i=1}^{p} \boldsymbol{A}_i\right)\left(\bigotimes_{i=1}^{p} \boldsymbol{B}_i\right) = \bigotimes_{i=1}^{p} (\boldsymbol{A}_i\boldsymbol{B}_i)$$

II — If the matrices $\boldsymbol{A}_i$, $i = 1, \dots, p$ are invertible, then $\otimes_{i=1}^{p}\boldsymbol{A}_i$ is invertible and one has

$$\left(\bigotimes_{i=1}^{p} \boldsymbol{A}_i\right)^{-1} = \bigotimes_{i=1}^{p} \boldsymbol{A}_i^{-1}$$

III — If the matrices $\boldsymbol{A}_i$, $i = 1, \dots, p$ are are of dimension $n_i \times n_i$, then

$$\left|\bigotimes_{i=1}^{p} \boldsymbol{A}_i\right| = \prod_{i=1}^{p} |\boldsymbol{A}_i|^{n_{(i)}}$$

where $n_{(i)} = \prod_{k \neq i} n_k$.

IV — If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $m \times n$ matrices, and $\boldsymbol{C}$ is $p \times q$, one has

$$\boldsymbol{C} \otimes (\boldsymbol{A} + \boldsymbol{B}) = \boldsymbol{C} \otimes \boldsymbol{A} + \boldsymbol{C} \otimes \boldsymbol{B}$$
$$(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$$

V — $\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = (\operatorname{tr} \boldsymbol{A})(\operatorname{tr} \boldsymbol{B})$

VI — $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$

VII — If $\boldsymbol{A}$ is a function of $\theta$ but $\boldsymbol{B}$ is not, then

$$\frac{\partial}{\partial\theta}[\boldsymbol{A} \otimes \boldsymbol{B}] = [\frac{\partial}{\partial\theta}\boldsymbol{A}] \otimes \boldsymbol{B}$$
$$\frac{\partial}{\partial\theta}[\boldsymbol{B} \otimes \boldsymbol{A}] = \boldsymbol{B} \otimes [\frac{\partial}{\partial\theta}\boldsymbol{A}]$$

VIII — If $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\theta)$ is a positive definite matrix whose entries are differentiable with respect to $\theta$, then with $\dot{\boldsymbol{\Sigma}} = \frac{\partial}{\partial\theta}\boldsymbol{\Sigma}$,

$$\frac{\partial}{\partial\theta}\log|\boldsymbol{\Sigma}| = \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\right]$$
$$\frac{\partial}{\partial\theta}\boldsymbol{\Sigma}^{-1} = -\,\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1} \ ,$$

IX — Let $\boldsymbol{X} \sim \mathsf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{A}$ and $\boldsymbol{B}$ symmetric matrices. Then

$$\mathbb{E}\,\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X} = \operatorname{tr}\boldsymbol{A}\boldsymbol{\Sigma} + \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}$$
$$\mathbb{C}\mathsf{ov}(\boldsymbol{X}'\boldsymbol{A}\boldsymbol{X}, \boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}) = 2\ \operatorname{tr}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma} + 4\boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\mu} \ .$$

X — Let $\boldsymbol{A}$ be a nonsingular matrix. Then

$$|\boldsymbol{A} + \mathbf{1}\mathbf{1}'| = |\boldsymbol{A}|(1 + \mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1}) \ .$$

If, furthermore, $\mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1} \neq -1$, then $\boldsymbol{A} + \mathbf{1}\mathbf{1}'$ is nonsingular and

$$(\boldsymbol{A} + \mathbf{1}\mathbf{1}')^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{A}^{-1}}{1 + \mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1}} \ .$$

# B Proofs of Section 2

Proof of Proposition 2.1: Recall that

$$L^I(\sigma^2, \boldsymbol{\xi} \mid \boldsymbol{y}) \propto (\sigma^2)^{(n-q)/2}|\boldsymbol{\Sigma}|^{-1/2}\ |\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}|^{-1/2}\ \exp\left\{-\frac{S^2_{\boldsymbol{\xi}}}{2\ \sigma^2}\right\} \ ,$$

and define the shorthand $\ell^I = \log L^I(\sigma^2, \boldsymbol{\xi} \mid \boldsymbol{y})$. Berger, De Oliveira, and Sansó (2001) prove that

$$\frac{\partial}{\partial\sigma^2}\ell^I = \frac{1}{2\sigma^2}\left(\frac{S^2_{\boldsymbol{\xi}}}{\sigma^2} - \mathbb{E}\frac{S^2_{\boldsymbol{\xi}}}{\sigma^2}\right)$$
$$\frac{\partial}{\partial\xi_i}\ell^I = \frac{1}{2\sigma^2}\left(R^i_{\boldsymbol{\xi}} - \mathbb{E}R^i_{\boldsymbol{\xi}}\right)$$

where $S^2_{\boldsymbol{\xi}}/\sigma^2 \sim \chi^2_{n-q}$ and $R^i_{\boldsymbol{\xi}}$ is a quadratic form on $\boldsymbol{P}\boldsymbol{Y} \sim \mathsf{N}(\mathbf{0}, \sigma^2\boldsymbol{P}\boldsymbol{\Sigma})$, associated with the matrix $\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}^i\boldsymbol{\Sigma}^{-1}$. As a consequence,

$$\mathbb{E}\left(\frac{\partial}{\partial\sigma^2}\ell^I\right)^2 = \frac{n-q}{2\sigma^2}$$
$$\mathbb{E}\left(\frac{\partial}{\partial\sigma^2}\ell^I\ \frac{\partial}{\partial\xi_i}\ell^I\right) = \frac{1}{2\sigma^2}\operatorname{tr}\boldsymbol{W}_i$$
$$\mathbb{E}\left(\frac{\partial}{\partial\xi_i}\ell^I\ \frac{\partial}{\partial\xi_j}\ell^I\right) = \frac{1}{2}\operatorname{tr}\boldsymbol{W}_i\boldsymbol{W}_j \ ,$$

where we have used the well-known fact that, whenever the products are possible, tr $\boldsymbol{ABC} =$ tr $\boldsymbol{CAB}$. The result follows from elementary properties of the determinant function. ■

Proof of Proposition 2.2: Suppose $\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{\xi} \sim \mathsf{N}(\boldsymbol{X\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \equiv \boldsymbol{\Sigma}(\boldsymbol{\xi})$, so that

$$L(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\theta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X\theta})\right\} .$$

Standard results of matrix differentiation plus facts VIII and IX of Appendix A, allows one to write, with $\ell^I \equiv \log L(\boldsymbol{\theta}, \boldsymbol{\xi} \mid \boldsymbol{y})$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}}\ell^I =& \frac{1}{2}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} - \mathbb{E}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right) \\ \frac{\partial}{\partial \xi_i}\ell^I =& \frac{1}{2}\left(R^i_{\boldsymbol{\xi}} - \mathbb{E}R^i_{\boldsymbol{\xi}}\right) \ , \end{aligned}$$

where $R^i_{\boldsymbol{\xi}}$ is a quadratic form on $(\boldsymbol{Y} - \boldsymbol{X\theta}) \sim \mathsf{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ associated with the matrix $\boldsymbol{\Sigma}^{-1}\dot{\boldsymbol{\Sigma}}^i\boldsymbol{\Sigma}^{-1}$. Then, it is easy to check that

$$\begin{aligned} \mathbb{E}\left(\frac{\partial}{\partial \xi_i}\ell^I \ \frac{\partial}{\partial \xi_j}\ell^I\right) &= \frac{1}{2}\text{tr } \boldsymbol{U}_i\boldsymbol{U}_j \\ \mathbb{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\ell^I \ \frac{\partial}{\partial \boldsymbol{\theta}'}\ell^I\right) &= \frac{1}{4}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \\ \mathbb{E}\frac{\partial}{\partial \xi_i}\frac{\partial \ell^I}{\partial \boldsymbol{\theta}} &= \boldsymbol{0} \end{aligned}$$

and now all the results follow easily from elementary properties of the determinant function. ■

# C Proofs of Section 3

## C.1 Proofs of Section 3.2

Proof of Lemma 3.1: We start out by proving (3.8). Since $\boldsymbol{\Sigma} = \otimes_{k=1}^p \boldsymbol{\Sigma}_k$, from Fact II of Appendix A it follows that $\boldsymbol{\Sigma}^{-1} = \otimes_{k=1}^p \boldsymbol{\Sigma}_k^{-1}$. Then, by Fact VI of Appendix A, repeated application of Fact I, and by Assumption $\mathscr{A}2$, one has

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} &= (\otimes_{k=1}^p \boldsymbol{X}_k'\boldsymbol{\Sigma}_k^{-1})(\otimes_{k=1}^p \boldsymbol{X}_k) \\ &= \otimes_{k=1}^p \boldsymbol{X}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k \ , \end{aligned}$$

which in conjunction with Fact III (or just by noting that $\boldsymbol{X}_k'\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k$ is a scalar) establishes the result. This last expression, Fact II and repeated application of Fact I shows (3.7). ■

Proof of Lemma 3.2: Formulas (3.14), (3.15) and (3.16) are shown in Berger et al. (2001) to follow essentially from assumption (3.9) and from the fact that $\lim_{\xi \to 0} \nu(\xi) = 0$, a simple consequence of (3.9) and $\rho(0 \mid \xi) = 0$. The assumption $\mathbf{1}'_{n_k} \boldsymbol{D}_k^{-1} \mathbf{1}_{n_k} \neq 0$ and (3.10) assure that the matrices are well-defined and that the expansions are meaningful.

We next prove (3.17). For simplicity, we drop the subscript $k$ in the sequel.

Suppose then that $\boldsymbol{X} = \mathbf{1}$, and write $\boldsymbol{\Sigma} = \boldsymbol{A} + \mathbf{1}\mathbf{1}'$. Using Fact X and simple manipulations, Berger et al. (2001) prove that (in a more general context)

$$\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}' = \mathbf{1}(\mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1})^{-1}\mathbf{1}' + \mathbf{1}\mathbf{1}'$$

Premultiplying this last equation by $\boldsymbol{\Sigma}^{-1} = \boldsymbol{A}^{-1} - (\boldsymbol{A}^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{A}^{-1})/(1 + \mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1})$ and simplifying, yields

$$\boldsymbol{\Sigma}^{-1}\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}' = \boldsymbol{A}^{-1}\mathbf{1}(\mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1})^{-1}\mathbf{1}' \ ;$$

post-multiplying this one by the same quantity finally produces, after some algebra,

$$\boldsymbol{\Sigma}^{-1}\mathbf{1}(\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{\Sigma}^{-1} = \frac{\boldsymbol{A}^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{A}^{-1}}{(1 + \mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1})\mathbf{1}'\boldsymbol{A}^{-1}\mathbf{1}} \ .$$

Substituting $\boldsymbol{A} = \nu(\xi)(\boldsymbol{D} + o(1))$ in the right-hand side yields the first part of (3.17). The second part follows directly from formula $\boldsymbol{\Phi}_k = \boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k(\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}\boldsymbol{X}_k)^{-1}\boldsymbol{X}'_k\boldsymbol{\Sigma}_k^{-1}$ and expansion (3.14). Assumption (3.10) assures that $\boldsymbol{H}_k$ is well defined, and it is also easy to show that $\boldsymbol{H}_k \neq \mathbf{0}$. ■

Proof of Proposition 3.3: Formula (3.18) is a simple consequence of Lemma 3.1 and Fact III of Appendix A. The continuity of $L^I(\boldsymbol{\xi} \mid \boldsymbol{y})$ as a function of $\boldsymbol{\xi}$ follows from the continuity of $\rho$ as a function of the parameter and the product form of the correlation function. We will use throughout this proof the fact that, as $\xi \to 0$, $\nu(\xi) \to 0$, and that, as $\xi_k \to \infty$, $\boldsymbol{\Sigma}_k \to \mathbf{I}_{n_k}$. These are simple consequences of the assumptions of Lemma 3.2.

Using formula (3.18) and the expansions of Lemma 3.2, it is possible to check that in the circumstances of part (a) we have

$$L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto \prod_{k \in B} \nu(\xi_k)^{(n_k - 3 + 2a)/2} \prod_{k \in A \cap B} \nu(\xi_k)^{1/2} \ (1 + g(\boldsymbol{\xi})) \ \times$$

$$\times \ \left\{ \boldsymbol{y}' \left[ \boldsymbol{C} - \prod_{k \in \bar{A} \cap B} \nu(\xi_k) \ \boldsymbol{C}^\star \right] \boldsymbol{y} \right\}^{-(n-3+2a)/2}$$

where $\boldsymbol{C} \equiv \otimes_{k=1}^{p} \boldsymbol{C}_k$, $\boldsymbol{C}^\star \equiv \otimes_{k=1}^{p} \boldsymbol{C}_k^\star$, with

$$\boldsymbol{C}_k = \begin{cases} \boldsymbol{G}_k, & \text{if } k \in B \\ \boldsymbol{I}_{n_k}, & \text{if } k \in \bar{B} \end{cases}$$

and

$$\boldsymbol{C}_k^\star = \begin{cases} \boldsymbol{F}_k, & \text{if } k \in \bar{A} \cap B \\ \boldsymbol{H}_k, & \text{if } k \in A \cap B \\ \boldsymbol{X}_k(\boldsymbol{X}_k'\boldsymbol{X}_k)^{-1}\boldsymbol{X}_k', & \text{if } k \in \bar{B} \ . \end{cases}$$

We next argue that the quantity between curly brackets can be bounded between two positive constants as long as $\xi_k$, $k \in B$, are sufficiently small, in which case the result follows immediately.

If $\bar{A} \neq \varnothing$, then that quantity converges to $\boldsymbol{y}'\boldsymbol{C}\boldsymbol{y}$, that we claim to be positive. To see that, recall that the Kronecker product of positive definite matrices is still a positive definite matrix. As a consequence, it suffices to show that $\boldsymbol{G}_k$ is positive definite. This follows from (3.14) and the fact that $\boldsymbol{\Sigma}_k^{-1}$ is positive definite (recall that $\nu$ is positive).

If $\bar{A} = \varnothing$, then it suffices to show that $\boldsymbol{y}'(\boldsymbol{C} - \boldsymbol{C}^\star)\boldsymbol{y} > 0$. To see that, note that it can be checked that $\boldsymbol{C} - \boldsymbol{C}^\star$ can be written as

$$\boldsymbol{C}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C})$$

where, from what we have seen above, $\boldsymbol{C}$ is positive definite. The quantity between brackets in the last expression is a projection matrix, and it is known that in this case, cf. Harville (1997, page 262),

$$\boldsymbol{C}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}) = (\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C})'\boldsymbol{C}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C})$$

which shows that

$$\boldsymbol{y}'(\boldsymbol{C} - \boldsymbol{C}^\star)\boldsymbol{y} = [(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C})\boldsymbol{y}]' \ \boldsymbol{C} \ [(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{C}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C})\boldsymbol{y}]$$

and this is positive since $\boldsymbol{C}$ is positive definite.

The proof of part (b) follows essentially from the same type of arguments. It is possible to check that in these circumstances we now have

$$\begin{aligned} L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \propto \prod_{k \in B} \nu(\xi_k)^{(n_k-3+2a)/2} & \prod_{k \in A \cap B} \nu(\xi_k)^{1/2} \ (1 + g(\{\xi_k, k \in B\})) \times \\ & \times \prod_{k \in \bar{B}} \left\{ |\boldsymbol{\Sigma}_k|^{-n_{(k)}/2} (\boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k)^{-1/2} \right\} \times \\ & \times \left\{ \boldsymbol{y}' \left[ \boldsymbol{C} - \prod_{k \in \bar{A} \cap B} \nu(\xi_k) \ \boldsymbol{C}^\star \right] \boldsymbol{y} \right\}^{-(n-3+2a)/2} \end{aligned}$$

where $\boldsymbol{C} \equiv \otimes_{k=1}^p \boldsymbol{C}_k$, $\boldsymbol{C}^\star \equiv \otimes_{k=1}^p \boldsymbol{C}_k^\star$, with

$$\boldsymbol{C}_k = \begin{cases} \boldsymbol{G}_k, & \text{if } k \in B \\ \boldsymbol{\Sigma}_k^{-1}, & \text{if } k \in \bar{B} \end{cases}$$

and

$$\boldsymbol{C}_k^\star = \begin{cases} \boldsymbol{F}_k, & \text{if } k \in \bar{A} \cap B \\ \boldsymbol{H}_k, & \text{if } k \in A \cap B \\ \boldsymbol{\Phi}_k, & \text{if } k \in \bar{B} \ . \end{cases}$$

If $\bar{A} \neq \varnothing$, then the quantity between curly brackets in the last factor converges to $\boldsymbol{y}'\boldsymbol{C}\boldsymbol{y} > 0$. In this case, $h$ can be taken to be the second-to-last factor, which is obviously continuous in the set $\{0 < \delta_k < \xi_k < M_k < +\infty\}$.

If $\bar{A} = \varnothing$, then we must show that $\boldsymbol{C} - \boldsymbol{C}^\star$ is positive definite in the set $\{\delta_k < \xi_k < M_k, k \in \bar{B}\}$. That follows from an argument similar to the one used for part (a), and therefore will be omitted. The product of the last and second-to-last factors is therefore a continuous function on $\{\delta_k < \xi_k < M_k, k \in \bar{B}\}$, which concludes the proof. ■

## C.2 Proofs of Section 3.3

Proof of Proposition 3.4: Using facts I and VII of Appendix A along with (3.7), it is clear that

$$\boldsymbol{W}_k = \boldsymbol{I}_{n_1} \otimes \cdots \otimes \dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Sigma}_k^{-1} \otimes \cdots \otimes \boldsymbol{I}_{n_p} - \boldsymbol{\Sigma}_1 \boldsymbol{\Phi}_1 \otimes \cdots \otimes \dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Phi}_k \otimes \cdots \otimes \boldsymbol{\Sigma}_p \boldsymbol{\Phi}_p \ .$$

Next, note that $\boldsymbol{\Sigma}_k\boldsymbol{\Phi}_k$ is a projection matrix, so that it is idempotent, and its rank, and therefore its trace, is $\operatorname{rank} \boldsymbol{X}_{n_k} = 1$. Also,

$$\begin{aligned}
\boldsymbol{\Phi}_k\boldsymbol{\Sigma}_k\boldsymbol{\Phi}_k &= \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\Sigma}_k\boldsymbol{\Phi}_k)(\boldsymbol{\Sigma}_k\boldsymbol{\Phi}_k) \\
&= \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{\Sigma}_k\boldsymbol{\Phi}_k) \qquad \text{(by idempotence)} \\
&= \boldsymbol{\Phi}_k \ .
\end{aligned}$$

These facts, formula V of Appendix A, some algebra and the known fact that $\operatorname{tr} \boldsymbol{AB} = \operatorname{tr} \boldsymbol{BA}$ whenever the products are possible, allows one to show that

$$\begin{aligned}
\operatorname{tr} \boldsymbol{W}_k &= n_{(k)} \ \operatorname{tr} \dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Sigma}_k^{-1} - \operatorname{tr} \dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Phi}_k \\
\operatorname{tr} \boldsymbol{W}_k^2 &= n_{(k)} \ \operatorname{tr}(\dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Sigma}_k^{-1})^2 + \operatorname{tr}(\dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Phi}_k)^2 - 2\operatorname{tr} \dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Sigma}_k^{-1}\dot{\boldsymbol{\Sigma}}_k\boldsymbol{\Phi}_k \\
\operatorname{tr} \boldsymbol{W}_i\boldsymbol{W}_j &= \prod_{l\neq i,j} n_l \times \operatorname{tr} \dot{\boldsymbol{\Sigma}}_i\boldsymbol{\Sigma}_i^{-1} \ \operatorname{tr} \dot{\boldsymbol{\Sigma}}_j\boldsymbol{\Sigma}_j^{-1} - \operatorname{tr} \dot{\boldsymbol{\Sigma}}_i\boldsymbol{\Phi}_i \ \operatorname{tr} \dot{\boldsymbol{\Sigma}}_j\boldsymbol{\Phi}_j \ ,
\end{aligned}$$

the key point to note being that only $\operatorname{tr} \boldsymbol{W}_i\boldsymbol{W}_j$ depends on more than one parameter. Therefore, we can write (2.7) as

$$I_R(\boldsymbol{\xi}) = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{\omega} \\ \boldsymbol{\omega}' & g_p^R(\xi_p) \end{pmatrix} \tag{C.1}$$

where $\boldsymbol{T}$ is $(p-1)\times(p-1)$ and does not depend on $\xi_p$, $\boldsymbol{\omega}$ is $(p-1)\times 1$, and $g_p^R(\xi_p) = \operatorname{tr} \boldsymbol{W}_p^2$, which depends on $\xi_p$ only. As a consequence of the above block format, we have (cf. Harville, 1997, page 188)

$$\begin{aligned}
|I_R(\boldsymbol{\xi})| &= |\boldsymbol{T}| \ (g_p^R(\xi_p) - \boldsymbol{\omega}'\boldsymbol{T}^{-1}\boldsymbol{\omega}) \\
&\leq |\boldsymbol{T}| \ g_p^R(\xi_p) \ ,
\end{aligned}$$

the last step following from the fact that $\boldsymbol{T}$ is positive definite, since $I_R(\boldsymbol{\xi})$ is, and therefore so is $\boldsymbol{T}^{-1}$. If we repeat this procedure $p$ times, we end up proving (3.23) by defining $\pi_k(\xi_k) = [g_k^R(\xi_k)]^{1/2} = [\operatorname{tr} \boldsymbol{W}_k^2]^{1/2}$.

We now study $\pi_k(\xi_k)$ as $\xi_k \to 0$. It is easy to verify that $\boldsymbol{D}_k\boldsymbol{G}_k$ is idempotent and that $\operatorname{tr} \boldsymbol{D}_k\boldsymbol{G}_k = n_k - 1$. Also, it is possible to show that $\boldsymbol{D}_k\boldsymbol{G}_k\boldsymbol{D}_k\boldsymbol{H}_k = \boldsymbol{D}_k\boldsymbol{H}_k$ and that $\operatorname{tr} \boldsymbol{D}_k\boldsymbol{H}_k = 1$. Note also that, as a consequence of the assumptions we have, as $\xi_k \to 0$,

$$\dot{\boldsymbol{\Sigma}}_k = \nu'(\xi_k)\boldsymbol{D}_k(1 + o(1)) \ . \tag{C.2}$$

These properties, and simple expansions along with (3.14) and (3.17), shows (3.24). Assumption $\mathscr{A}8$ is capital to this result.

Since, as $\xi_k \to \infty$, $\boldsymbol{\Sigma}_k \to \boldsymbol{I}_{n_k}$, the integrability of $\pi_k(\xi_k)$ at infinity follows from Assumption $\mathscr{A}7$. ■

Proof of Proposition 3.5: This proof essentially mimics the one of Proposition 3.4. Using the definition of $\boldsymbol{U}_k$ and facts I and VII of Appendix A, it is easy to check that

$$\boldsymbol{U}_k = \boldsymbol{I}_{n_1} \otimes \cdots \otimes \dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Sigma}_k^{-1} \otimes \cdots \otimes \boldsymbol{I}_{n_p} \ .$$

Next, essentially the same type of arguments as before allows one to write

$$\begin{aligned} \operatorname{tr} \boldsymbol{U}_k &= n_{(k)} \ \operatorname{tr} \dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Sigma}_k^{-1} \\ \operatorname{tr} \boldsymbol{U}_k^2 &= n_{(k)} \ \operatorname{tr} (\dot{\boldsymbol{\Sigma}}_k \boldsymbol{\Sigma}_k^{-1})^2 \\ \operatorname{tr} \boldsymbol{U}_i \boldsymbol{U}_j &= \prod_{l \neq i,j} n_l \times \operatorname{tr} \dot{\boldsymbol{\Sigma}}_i \boldsymbol{\Sigma}_i^{-1} \ \operatorname{tr} \dot{\boldsymbol{\Sigma}}_j \boldsymbol{\Sigma}_j^{-1} \ . \end{aligned}$$

Again, only $\operatorname{tr} \boldsymbol{U}_i \boldsymbol{U}_j$ depends on more than one parameter, and the construction carried out in the proof of Proposition 3.4 can be repeated to show (3.25), where now $\pi_k^{J1}(\xi_k) \propto (\operatorname{tr} \boldsymbol{U}_k^2)^{1/2}$ and $\pi_k^{J2}(\xi_k) \propto \pi_k^{J1}(\xi_k) \ |\boldsymbol{X}_k' \boldsymbol{\Sigma}_k^{-1} \boldsymbol{X}_k|^{1/2}$.

To obtain the behavior of these functions as $\xi_k \to 0$, it suffices to use expansions (3.14), (3.16) and (C.2). It is also necessary to recall that $\boldsymbol{D}_k \boldsymbol{G}_k$ is idempotent and that its trace is $n_k - 1$ (and hence assumption $\mathscr{A}8$).

The integrability at infinity of $\pi_k^{Ji}(\cdot)$ follows from the fact that, as $\xi_k \to \infty$, $\boldsymbol{\Sigma}_k \to \boldsymbol{I}_{n_k}$ and Assumption $\mathscr{A}7$. ■

## C.3 Proofs of Section 3.4

Proof of Theorem 3.6: We will determine conditions under which

$$0 < \int_0^\infty L^I(\boldsymbol{\xi} \mid \boldsymbol{y}) \ \pi(\boldsymbol{\xi}) \ d\xi_1 \cdots \ d\xi_p < \infty$$

holds, which is tantamount to posterior propriety. To simplify the notation, we will write $f(\boldsymbol{\xi}) = L^I(\boldsymbol{\xi} \mid \boldsymbol{y})$. The function $\pi(\boldsymbol{\xi})$ will be either associated with the reference prior or with one of the Jeffreys' priors we considered, but we will carry out the calculations for a general exponent $a$ in the integrated likelihood. Recall that when $\pi = \pi^R$, $a = 1$, while when $\pi = \pi^{Ji}$, $i = 1, 2$, it is $a = 1$ and $a = 3/2$, respectively. Also, in all of these instances of $\pi$, there are functions $\pi_k(\xi_k)$, $k = 1, \ldots, p$ such that

$$\pi(\boldsymbol{\xi}) \leq \prod_{k=1}^p \pi_k(\xi_k)$$

and those have essentially the same behavior across the different choices for $\pi(\boldsymbol{\xi})$, so that a common proof can be sought — cf. propositions 3.4 and 3.5. Note that

$$\int_0^\infty f(\boldsymbol{\xi})\ \pi(\boldsymbol{\xi})\ d\xi_1 \cdots d\xi_p \leq \int_0^\infty f(\boldsymbol{\xi})\ \pi_1(\xi_1)\ d\xi_1\ \cdots \pi_p(\xi_p) d\xi_p\ , \tag{C.3}$$

and therefore we will have posterior property if the right-hand side is finite.

The above integral can be partitioned into a sum of integrals each of which with different regions of integration. If the region is of the form $[\delta_1, +\infty[ \times \cdots \times [\delta_p, +\infty[$, then there are no issues to address since the $\pi_k$ are integrable at infinity and in this set the integrated likelihood is bounded.

Therefore, we need only consider regions of the following type: let $B$ be a nonempty but otherwise arbitrary subset of $\{1, \ldots, n\}$, and define said regions by

i) for $k \in B$, $0 < \xi_k < \delta_k$, where $\delta_k$ can be chosen arbitrarily small; for $k \in \bar{B}$, $\xi_k > M_k$, where $M_k$ can be assumed arbitrarily large;

ii) for $k \in B$, $0 < \xi_k < \delta_k$, where $\delta_k$ can be chosen arbitrarily small; for $k \in \bar{B}$, $\delta_k < \xi_k < M_k$, where $M_k$ is finite but otherwise arbitrary.

Let's consider (i) first. Combining expansions (3.19), (3.24) and (3.26), it is easy to see that

$$f(\boldsymbol{\xi}) \prod_{k=1}^{p} \pi_k(\xi_k) \propto \prod_{k \in A \cap B} |\nu'(\xi_k)|\ \nu(\xi_k)^{(n_k - 2 - 2\alpha_k + 2a)/2} \prod_{k \in \bar{A} \cap B} |\nu'(\xi_k)|\ \nu(\xi_k)^{(n_k - 5 + 2a)/2} \times$$
$$\times \prod_{k \in \bar{B}} \pi_k(\xi_k)\ (1 + g(\boldsymbol{\xi})),$$

where $\alpha_k$ is defined in Proposition 3.5 ($\alpha_k \equiv 1$ in the case of the reference prior).

When we look at situation (ii), formula (3.20), and the same expansions as before for the priors, yield an expression formally identically to the one above multiplied by $h(\xi_k, k \in \bar{B})$.

Since the functions $\pi_k$ are integrable at infinity, and for small enough $\xi$, $\nu(\xi) < 1$, it is clear that posterior propriety will be achieved for all three priors whenever the function

$$|\nu'(\xi)|\ \nu(\xi)^{(n_k - 5 + 2a)/2}$$

is integrable at the origin for every $k$, or, equivalently, when that happens to the function

$$|\nu'(\xi)|\ \nu(\xi)^{(\min n_k - 5 + 2a)/2}\ .$$

From the assumptions, it is easy to conclude that, for small enough $\xi$, $\nu'(\xi) > 0$, so that the integrability at the origin reduces to verifying that

$$\lim_{\xi \to 0} \nu(\xi)^{(\min n_k - 5 + 2a)/2+1} < \infty \ ,$$

which reduces to (3.28) since $\nu = o(1)$ and, by assumption $\mathscr{A}8$, $\min n_k \geq 2$. ■

# References

Bayarri, M. J., Berger, J. O., Higdon, D., Kennedy, M. C., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H. and Tu, J. (2002). A framework for validation of computer models. Tech. Rep. 128, National Institute of Statistical Sciences.

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors (Disc: p49-60). In *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting.*

Berger, J. O., De Oliveira, V., and Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* **396** 1361–1374.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference (C/R p128-147). *Journal of the Royal Statistical Society, Series B, Methodological* **41** 113–128.

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* Wiley.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag Inc.

Kennedy, M. C. and O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87** 1–13.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B, Methodological* **63** 425–464.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer experiments (C/R: p423-435). *Statistical Science* **4** 409–423.

Tong, Y. L. (1990). *The Multivariate Normal Distribution.* Springer-Verlag Inc.

National Institute of Statistical Sciences
Statistical and Applied Mathematical Sciences Institute
19 T.W. Alexander Drive
P.O. Box 14006
Research Triangle Park, NC 27709–4006
USA
E-Mail: rui@niss.org