The Annals of Statistics
2003, Vol. 31, No. 5, 1491–1516

ENRICHED CONJUGATE AND REFERENCE PRIORS FOR THE WISHART FAMILY ON SYMMETRIC CONES

BY GUIDO CONSONNI[1] AND PIERO VERONESE[2]

University of Pavia and L. Bocconi University

A general Wishart family on a symmetric cone is a natural exponential family (NEF) having a homogeneous quadratic variance function. Using results in the abstract theory of Euclidean Jordan algebras, the structure of conditional reducibility is shown to hold for such a family, and we identify the associated parameterization ϕ and analyze its properties. The enriched standard conjugate family for ϕ and the mean parameter μ are defined and discussed. This family is considerably more flexible than the standard conjugate one. The reference priors for ϕ and μ are obtained and shown to belong to the enriched standard conjugate family; in particular, this allows us to verify that reference posteriors are always proper. The above results extend those available for NEFs having a simple quadratic variance function. Specifications of the theory to the cone of real symmetric and positive-definite matrices are discussed in detail and allow us to perform Bayesian inference on the covariance matrix Σ of a multivariate normal model under the enriched standard conjugate family. In particular, commonly employed Bayes estimates, such as the posterior expectation of Σ and Σ^{-1}, are provided in closed form.

1. Introduction. The usual Wishart distribution is defined on the space of symmetric and positive-definite (s.p.d.) matrices having real entries, which represents a symmetric cone. Symmetric cones may be defined in greater generality. There are five irreducible symmetric cones, namely, the four cones of Hermitian positive-definite matrices on the real line, the complex plane, the set of quaternions, the Cayley algebra, and the cone of revolution in $\mathbb{R}^d$, also called the Lorentz cone, and accordingly five distinct general Wishart distributions may be defined. They appear in Jensen (1988) within the context of linear hypothesis testing; for an application to the complex plane, see Andersen, Højbjerre, Sørensen and Eriksen (1995).

Simple Euclidean Jordan algebras represent the appropriate mathematical framework to embed these extensions; see Faraut and Korányi (1994). Massam (1994) provides a detailed study of the Wishart distribution in this general

---

Received December 2000; revised November 2002.

[1]Supported in part by MIUR Grants 9813041044009 and 2001138888, CNR Grant 98.01329.CT10 and the University of Pavia.

[2]Supported in part by MIUR Grants 9813041044003 and 2001138888, CNR Grant 99.01688CT10 and L. Bocconi University.

AMS 2000 subject classifications. Primary 62E15, 62F15; secondary 60E05.

Key words and phrases. Bayesian inference, conditional reducibility, exponential family, Jordan algebra, noninformative prior, Peirce decomposition.

abstract setting; see also Casalis and Letac (1996) and Massam and Neher (1997). Analogous extensions are provided in Massam and Neher (1998) with reference to the lattice conditional independence model introduced by Andersson and Perlman (1993); see also Andersson and Perlman (1995).

The usual Wishart family defined on the space of real s.p.d. matrices is a natural exponential family (NEF) with a homogeneous quadratic variance function (HQVF). Casalis (1991) has shown that all NEF-HQVFs on $\mathbb{R}^d$ are of Wishart type. Letac (1991) discusses various types of quadratic variance functions in the multivariate setting.

A useful structural property of multivariate NEFs, called conditional reducibility, has been introduced and thoroughly discussed in Consonni and Veronese (2001). In particular, this property allows us to reparameterize the family in terms of a parameter ϕ, say, whose components are variation and likelihood independent. This allows us to construct an enriched standard conjugate family on ϕ which enjoys greater flexibility relative to the standard conjugate one, while remaining closed under i.i.d. sampling. Furthermore, the ϕ-parameterization has proved to be especially effective for the construction of reference priors relative to NEFs having a simple quadratic variance function (SQVF); see Consonni, Veronese and Gutiérrez-Peña (2000). We recall that reference priors represent a useful tool to perform Bayesian inference when prior information is limited and provide a benchmark to assess robustness and sensitivity to prior specifications. For an extensive treatment of reference priors, see Bernardo (1979), Berger and Bernardo (1992), Bernardo and Smith (1994) and Kass and Wasserman (1996).

In this paper, we show that general Wishart families on symmetric cones are conditionally reducible relative to a suitable (Peirce) decomposition of the observables, identify the ϕ-parameterization and study its properties. Next, we construct the enriched standard conjugate family on ϕ and provide sufficient conditions under which the family is proper. Furthermore, we derive the reference prior for both ϕ and the mean parameter and show that they belong to the enriched standard conjugate family. Some results are also proved for the Jeffreys prior.

The above general results are specified to the usual setting of real s.p.d. matrices. They are directly applicable to Bayesian inference on the covariance matrix Σ of a multivariate normal model, where the need for flexible informative priors is well known. As far as reference analysis is concerned, we obtain a novel reference prior on Σ, wherein the order of inferential importance of the parameters is explicated directly in terms of the elements of Σ. Specifically, let $(X_1, \dots, X_l)$ be jointly normal with zero expectation and let σ_{kk} denote the variance of X_k and σ_{kj}, $j = 1, \dots, k-1$, the covariance of (X_k, X_j). Then the order of our reference prior is specified by the blocks $\{\sigma_{k1}, \dots, \sigma_{kk}\}$, $k = 1, \dots, d$. For an alternative reference prior on Σ, wherein the order of inferential importance is dictated by the ordered eigenvalues of Σ, see Yang and Berger (1994).

One advantage of the reference prior we propose is that Bayesian estimators of Σ, based on commonly used loss functions, are available in closed form and thus do not require simulation methods.

The structure of the paper is as follows. In Section 2, we review some basic facts about NEFs and conditionally reducible NEFs. Next, we present the enriched standard conjugate families, as well as the reference priors, for the parameter ϕ.

Section 3 is devoted to the (general) Wishart family on symmetric cones. We present some background notation on cones and Jordan algebras with particular reference to Peirce decomposition. We also provide results on the variance function and Fisher information. The fundamental property of conditional reducibility for a general Wishart family on symmetric cones is established in Theorem 1 and the nature of the ϕ-parameterization elucidated. Specifications to the cone of real s.p.d. matrices are given, and connections with alternative parameterizations for the multivariate normal model are detailed.

In Section 4, we construct an enriched standard conjugate family for the ϕ-parameter of a general Wishart family and identify its structural properties in Theorem 2. Next, we specialize it to the usual setting and show that the resulting family coincides with the generalized inverted Wishart (GIW) proposed in Brown, Le and Zidek (1994). Then we provide closed-form recursive formulas for the computation of the expected value of Σ and Σ^{-1} under a GIW.

Section 5 is dedicated to the construction of reference priors for the general Wishart family related to the ϕ-parameter (Theorem 3) and to the mean parameter (Theorem 4). Finally, the results are specialized to the setting of real s.p.d. matrices and show that the proposed reference prior does not suffer from marginalization paradoxes and enjoys some frequentist coverage properties.

The last section summarizes the main findings and emphasizes that properties relative to NEF-SQVFs on $\mathbb{R}^d$ proved in Consonni, Veronese and Gutiérrez-Peña (2000) are also valid for NEF-HQVFs on symmetric cones, thus establishing a perfect analogy between the two cases.

2. Background.

2.1. *Parameterizations for natural exponential families.* Excellent accounts of exponential family theory are contained in Barndorff-Nielsen (1978), in Brown (1986) and, with a view toward Bayesian applications, in Gutiérrez-Peña and Smith (1997).

Let η be a σ-finite measure on the Borel sets of $\mathbb{R}^d$. Consider the family $\mathcal{F}$ of probability measures on $\mathbb{R}^d$, whose densities with respect to η are of the form

$$p_\theta(x|\theta) = \exp\{\theta^T x - M(\theta)\}, \qquad \theta \in \Theta \subseteq \mathcal{N}, \tag{1}$$

where $M(\theta) = \ln \int \exp\{\theta^T x\}\eta(dx)$ represents the *cumulant transform* of the measure η and θ^T is the transpose of θ.

Let $\mathcal{N} = \{\theta \in \mathbb{R}^d : M(\theta) < \infty\}$ and denote by $\mathcal{N}^\circ$ its interior. If $\Theta = \mathcal{N}^\circ$ and the carrier measure η is not concentrated on an affine subspace of $\mathbb{R}^d$, $\mathcal{F}$ is called a *natural exponential family* (NEF).

An alternative parameterization of the family $\mathcal{F}$, called the *mean parameterization*, is given by $\mu = \mu(\theta) = \nabla M(\theta)$, with $\mu(\cdot)$ a one-to-one transformation from Θ onto $\Omega = \mu(\Theta)$. The function $\partial^2 M(\theta)/(\partial\theta^T \partial\theta)$ represents the variance associated with $\mathcal{F}$. When regarded as a function of μ, it is called the *variance function* and denoted by $V(\mu)$. Together with the mean domain Ω, the variance function characterizes the family $\mathcal{F}$ within the class of all natural exponential families.

An important class of variance functions is represented by the *quadratic variance function* (QVF); see Morris (1982), Letac (1991) and Casalis (1996). A special case is represented by the homogeneous quadratic variance function (HQVF), defined by $V(\mu) = Q(\mu, \mu)$, where the map $Q: \Omega \times \Omega \to \mathcal{M}_d$ is symmetric bilinear. Casalis (1991) has characterized all such families and proved that they are of Wishart type.

Another useful parameterization is obtained when an NEF is *conditionally reducible*; see Consonni and Veronese (2001), which generalizes earlier work on *reducibility* by Bar-Lev, Bshouty, Enis, Letac, Lu and Richards (1994) and Gutiérrez-Peña and Smith (1997). The notion of conditional reducibility is strictly related to that of a *cut*; see Barndorff-Nielsen (1978).

An NEF is conditionally reducible if its density function can be factored into the product of r conditional densities, each belonging to an NEF, for some $r \in \{1, \ldots, d\}$. For $x = (x_1^T, \ldots, x_d^T)^T$, with $x_i \in \mathbb{R}^{d_i}$, $\sum_{i=1}^r d_i = d$, one can thus write

$$
\begin{aligned}
p_\theta(x|\theta) = p_\phi\big(x|\phi(\theta)\big) &= \prod_{k=1}^r p_{\phi_k}\big(x_k|x_{[k-1]}; \phi_k(\theta)\big) \\
&= \prod_{k=1}^r \exp\big\{\phi_k(\theta)^T x_k - M_k\big(\phi_k(\theta); x_{[k-1]}\big)\big\},
\end{aligned} \tag{2}
$$

where $x_{[k]} = (x_1^T, \ldots, x_k^T)^T$, with the understanding that $x_{[0]}$ is void. The vector $\phi = (\phi_1^T, \ldots, \phi_r^T)^T$, called the *cr-parameter*, is a one-to-one function from Θ onto $\phi(\Theta) = \Phi$, say.

Furthermore, it can be shown that

$$
M_k\big(\phi_k; x_{[k-1]}\big) = x_{[k-1]}^T A_{k[k-1]}(\phi_k) + B_k(\phi_k) \tag{3}
$$

for some functions A_{kj} and B_k, with $A_{k[k-1]} = (A_{k1}^T, \ldots, A_{k(k-1)}^T)^T$.

The *cr*-parameter exhibits some useful features. In particular, its components ϕ_k are variation independent, that is, $\Phi = \Phi_1 \times \cdots \times \Phi_r$, with $\phi_k \in \Phi_k, k = 1, \ldots, r$.

Furthermore, from (2) it follows that the Fisher information matrix $H(\phi)$ is block diagonal, that is,

$$
H^\phi(\phi) = \mathrm{Diag}\big\{H_{11}^\phi(\phi_1), \ldots, H_{kk}^\phi(\phi_{[k]}), \ldots, H_{rr}^\phi(\phi)\big\}, \tag{4}
$$

with the kth block only depending on $\phi_{[k]}$. The diagonal-block structure of (4) implies that the ϕ_k's are totally orthogonal; see Cox and Reid (1987).

The above properties turn out to be very useful for the specification of prior distributions, in particular, for extending conjugate families and constructing reference priors.

2.2. *Prior distributions for conditionally reducible NEFs.* Consonni, Veronese and Gutiérrez-Peña (2000) define an enriched standard conjugate family for a conditionally reducible NEF $\mathcal{F}$ (relative to the cr-parameter ϕ), $\mathcal{E}_\phi(\mathcal{F})$, whose densities, with respect to Lebesgue measure, are

$$(5) \qquad \pi_\phi(\phi|t, n') \propto \prod_{k=1}^{r} \exp\{\phi_k^T t_k^k - [t_{[k-1]}^{kT} A_{k[k-1]}(\phi_k) + n'_k B_k(\phi_k)]\},$$

where $t = (t^{1\,T}, \dots, t^{r\,T})^T$, $t^k = (t_1^{k\,T}, \dots, t_k^{k\,T})^T$, $t_j^k \in \mathbb{R}^{d_j}$, $j = 1, \dots, k$, $k = 1, \dots, r$; and $n' = (n'_1, \dots, n'_r)^T$, $n' \in \mathbb{R}^r$.

It is worth noticing that under the enriched standard conjugate family the parameters ϕ_k are stochastically independent. Clearly, $\mathcal{E}_\phi(\mathcal{F})$ can be regarded as the product of r standard conjugate families, each being relative to a conditional distribution in (2); see Diaconis and Ylvisaker (1979). When $t_j^k = t_j$, $j = 1, \dots, k$, $k = 1, \dots, r$, the family is called *simple* enriched standard conjugate and was introduced in Consonni and Veronese (2001).

A further property of the family $\mathcal{E}_\phi(\mathcal{F})$ is "uniqueness." Specifically, for fixed r, if ψ were the cr-parameter of an alternative factorization of the joint density $p_\theta(x)$, then $\mathcal{E}_\psi(\mathcal{F})$ would coincide with the family of priors on ψ induced by $\mathcal{E}_\phi(\mathcal{F})$ via the standard change-of-variable technique.

Starting from the prior family (5), one can obtain the induced enriched standard conjugate family on the mean parameter. To this end, one needs to compute the Jacobian $J_\phi(\mu)$ of the transformation $\phi \to \mu$, which can be written as $J_\phi(\theta) J_\theta(\mu)$. Consonni and Veronese (2001) have shown that $J_\phi(\theta) = 1$. On the other hand,

$$J_\theta(\mu) = \det\left(\frac{\partial\theta}{\partial\mu}\right) = \det\left(\frac{\partial^2 M(\theta)}{\partial\theta^T \partial\theta}\right)^{-1}\Bigg|_{\theta=\theta(\mu)} = \det(V(\mu))^{-1}.$$

One can thus write the density of the induced family as

$$(6) \qquad \pi_\mu^\phi(\mu|t, n') \propto \pi_\phi(\phi(\mu)|t, n')\{\det(V(\mu))\}^{-1}.$$

2.3. *Reference priors.* Consider an arbitrary family of distributions parameterized by $\phi \in \Phi \subseteq \mathbb{R}^d$. A traditional tool for noninformative Bayesian analysis is represented by the Jeffreys prior whose density relative to Lebesgue measure is given by the square root of the determinant of the (expected) Fisher information matrix H^ϕ.

The Jeffreys priors seem to be quite effective in one-dimensional settings. On the other hand, their inferential performance appears much more debatable when several parameters are present. In this case, reference priors appear to be more suitable because of their ability to distinguish between parameters of interest and nuisance parameters; see, for example, Berger and Bernardo (1992) and Bernardo and Smith (1994), Section 5.4.

We assume that ϕ is separated into r groups $(\phi_1, \dots, \phi_r)$, with $\phi_k \in \Phi_k$, arranged in decreasing order of inferential importance. Under some regularity conditions, if (i) $\Phi = \Phi_1 \times \cdots \times \Phi_r$, (ii) $H^\phi(\phi) = \mathrm{Diag}\{H_1^\phi(\phi), \dots, H_r^\phi(\phi)\}$, with $H_k^\phi(\phi)$, and

$$\text{(7)} \quad \text{(iii)} \quad \det\{H_k^\phi(\phi)\} = a_k(\phi_k) b_k(\phi_{[k-1]}, \phi_{k+1}, \dots, \phi_r) \qquad \forall k \in \{1, 2, \dots, r\},$$

for some positive functions $a_k(\cdot)$ and $b_k(\cdot)$, then the density—with respect to Lebesgue measure—of the r-group reference prior on ϕ, relative to the order $(\phi_1, \dots, \phi_r)$, is given by

$$\text{(8)} \qquad {}_{\phi_1,\dots,\phi_r}\pi_\phi(\phi_1, \dots, \phi_r) \propto \prod_{k=1}^{r} a_k(\phi_k)^{1/2}.$$

In addition, the prior in (8) does not depend on the order of the r-groups, that is, ${}_{\phi_{i_1},\dots,\phi_{i_r}}\pi_\phi(\phi_1, \dots, \phi_r) = {}_{\phi_1,\dots,\phi_r}\pi_\phi(\phi_1, \dots, \phi_r)$, for all permutations $(i_1, \dots, i_r)$ of $(1, \dots, r)$. When the ordering of the variables does not matter, we shall simply write $\pi_\phi(\phi_1, \dots, \phi_r)$. For a proof of this result, see Datta and Ghosh (1995). The case in which the information matrix is not block diagonal is discussed in Gutiérrez-Peña and Rueda (2003). Consonni, Veronese and Gutiérrez-Peña (2000) have shown that conditions (ii) and (iii) hold for NEF-SQVFs, with ϕ corresponding to the cr-parameter, and explicitly identify the prior (8).

3. The Wishart family on symmetric cones.

3.1. *Symmetric cones and Euclidean Jordan algebras.* In this section, we briefly recall the principal aspects of symmetric cones and Jordan algebras. The notation and results are taken from Faraut and Korányi (1994) and Massam and Neher (1997), henceforth FK and MN, respectively, to which we refer for further details.

Let V be a finite-dimensional real Euclidean space. An algebra $\mathcal{V}$ over $\mathbb{R}$ is called a *Jordan algebra* if, for all elements x and y in $\mathcal{V}$, the product xy satisfies $xy = yx$ and $x(x^2y) = x^2(xy)$. Note that $\mathcal{V}$ is said to be *Euclidean* if there exists a positive-definite symmetric bilinear form on $\mathcal{V}$ which is associative, that is, if there exists an inner product $(u|v)$ such that $(L(x)u|v) = (u|L(x)v)$, where L is a map on $\mathcal{V}$ such that $L(x)y = xy$.

For x in $\mathcal{V}$, the map $P(x) = 2L(x)^2 - L(x^2)$ is called the *quadratic representation* on $\mathcal{V}$. The algebra is said to be *simple* if it does not contain any nontrivial ideal. We will assume from now on that $\mathcal{V}$ is a simple Euclidean Jordan algebra.

There exists a one-to-one correspondence between symmetric cones and simple Euclidean Jordan algebras; we denote by Ω the cone associated to $\mathcal{V}$. The operator's trace and determinant on $\mathcal{V}$ are analyzed in detail in MN.

If c is an idempotent element of $\mathcal{V}$, the only possible eigenvalues of c are 0, 1/2 and 1, and the corresponding eigenspaces are denoted by $\mathcal{V}(c,0)$, $\mathcal{V}(c,1/2)$ and $\mathcal{V}(c,1)$. The decomposition of $\mathcal{V}$, with respect to c,

$$\mathcal{V} = \mathcal{V}(c,1) \oplus \mathcal{V}(c,1/2) \oplus \mathcal{V}(c,0),$$

is called the Peirce decomposition of $\mathcal{V}$. If $x \in \mathcal{V}$, then x decomposes as $x_1 + x_{12} + x_0$, with $x_i \in \mathcal{V}(c,i)$, $i = 0, 1$, and $x_{12} \in \mathcal{V}(c,1/2)$. If d represents the dimension of $\mathcal{V}$ and l the rank of $\mathcal{V}$, then

$$d = l + g\frac{l(l-1)}{2},$$

where g is called the Peirce invariant. Finally, if c has rank k, then $\mathcal{V}(c,1)$ has rank k; see MN (Section 2.2).

In the remainder of this paper, we will use the following results:

(i) an element $x \in \mathcal{V}$ is invertible if and only if $P(x)$ is invertible and, in this case, $P(x)x^{-1} = x$ (see FK, Proposition 2.3.1);

(ii) $\det(P(x)) = (\det(x))^{2d/l}$ (see FK, Proposition 3.4.2);

(iii) the differential of the map $x \mapsto x^{-1}$ is $-P(x)^{-1} = -P(x^{-1})$ (see FK, Proposition 2.3.3);

(iv) if the scalar product on $\mathcal{V}$ is defined by $(x|y) = \mathrm{tr}(xy)$, then the differential of the map $x \mapsto \log(\det(x))$ is x^{-1} (see FK, Proposition 3.4.2).

A discussion of the five irreducible symmetric cones can be found in FK (Chapter 5) and in Massam (1994) together with the corresponding definitions of trace and determinant. The most commonly used cone is represented by the set of real s.p.d. matrices which is associated to the Euclidean Jordan algebra $\mathrm{Sym}(l,\mathbb{R})$ of $l \times l$ real symmetric matrices with the Jordan product $\frac{1}{2}(xy + yx)$. In this case, $P(x)y = xyx$, while the notion of trace and determinant are well known. Furthermore, the Peirce decomposition corresponding to the idempotent matrix $c_k = \mathrm{Diag}(I_k, 0)$, with I_k the identity matrix of order k, is

$$X = \begin{pmatrix} X_1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & X_{12} \\ X_{12}^T & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & X_0 \end{pmatrix}, \tag{9}$$

where X_1 is the principal submatrix of X of order k.

3.2. *Definitions and basic properties.* Let $\mathcal{V}$ be a simple Euclidean Jordan algebra with rank l, Peirce invariant g and dimension

$$d = l + g\frac{l(l-1)}{2}$$

and denote by Ω the symmetric cone associated to $\mathcal{V}$.

The Wishart family on symmetric cones has been defined in FK (Chapter 6). We shall follow the exponential family notation presented in MN [see also Casalis and Letac (1996)]. The density, relative to Lebesgue measure on Ω, of the (general) Wishart family is given by

$$p(x|\xi) = \exp\{-(\xi, x) + p\log(\det(\xi))\}h_{l,p}(x), \tag{10}$$

with $\xi \in \Omega$, $(\xi, x) = \mathrm{tr}(\xi x)$ and

$$h_{l,p}(x) = \frac{1}{\Gamma_{l,\Omega}(p)}(\det x)^{p-d/l}, \qquad p > \frac{g(l-1)}{2}, \tag{11}$$

where

$$\Gamma_{l,\Omega}(p) = \int_\Omega e^{-\mathrm{tr}(x)}(\det x)^{p-d/l}\, dx. \tag{12}$$

The family (10) is denoted by $W_l(p, \xi)$.

We note that the natural parameter of the NEF associated to (10) is $-\xi$ and the cumulant transform is $M(-\xi) = -p\log(\det(\xi))$. Using condition (iv) of Section 3.1, one deduces that the mean parameter of the family (10) is given by

$$\mu = \nabla\big(p\log(\det(\xi))\big) = p\xi^{-1}. \tag{13}$$

Suppose now that Ω is the space of real s.p.d. $l \times l$ matrices. The dominating measure of the *usual* Wishart distribution, written $W_l^*(w, \Sigma)$, where w denotes the degrees of freedom and $w\Sigma$ its expectation, can be obtained from $h_{l,p}(x)\,dx$, transforming Lebesgue measure dx to the one commonly adopted in this case which operates only on the $l \times (l+1)/2$ distinct elements of a symmetric matrix. As a consequence, one has $p = w/2$. Finally, equating expectations, one obtains $p\xi^{-1} = (w/2)\xi^{-1} = w\Sigma$, whence $\xi = \Sigma^{-1}/2$, thus completing the specialization of the general case to the usual one. Notice, however, that $-\xi$ is *not* the natural parameter in the usual Wishart setting because $-(\xi, x) = -\mathrm{tr}(\xi x) = -\sum_{i=1}^{l} x_{ii}\xi_{ii} - 2\sum_{i<j} x_{ij}\xi_{ij}$ due to the symmetry of x and ξ. The actual natural parameter $\theta = (\theta_{ij})$ has elements $\theta_{ii} = -\xi_{ii}$ and $\theta_{ij} = -2\xi_{ij}$, $i < j$.

The variance of a Wishart family $W_l(p, \xi)$ is equal to $pP(-\xi^{-1})$, where P is the quadratic application on $\mathcal{V}$, because of condition (iii) of Section 3.1. Recalling the definition of the expectation (13), one obtains the variance function

$$V(\mu) = pP\left(-\frac{\mu}{p}\right) = \frac{1}{p}P(\mu), \tag{14}$$

which is a general version of the HQVF described in Section 2.1.

Since for an exponential family the Fisher information relative to the ξ-parameterization H^ξ coincides with $V(\mu(\xi))$, we can write

$$H^\xi(\xi) = V(\mu(\xi)) = \frac{1}{p}P(p\xi^{-1}), \tag{15}$$

because of (13) and (14).

From (14) and condition (iii) of Section 3.1, it follows that

$$\det(V(\mu)) = \det\left(\frac{1}{p}P(\mu)\right) = \left(\frac{1}{p}\right)^l (\det\mu)^{2d/l}. \tag{16}$$

Since $M(-\xi(\mu)) = -p\log(\det(p\mu^{-1}))$ [see (10) and (13)] one can write

$$\det V(\mu) \propto \exp\{-(\xi(\mu), z) - vM(-\xi(\mu))\}, \tag{17}$$

with $z = 0$ and $v = -2d/lp$.

Formula (17) extends the scope of the results presented in Gutiérrez-Peña and Smith (1997), Letac (1997) and Consonni, Veronese and Gutiérrez-Peña (2000) and will be useful to investigate properties of the families of priors defined in Section 4 and the reference prior in Section 5.

3.3. *Conditional reducibility.* It is well known that if the matrix X is distributed like a usual Wishart, then the same property holds for each principal submatrix of X. A similar result is true for the general setting. Let $x \sim W_l(p, \xi)$ and consider the Peirce components of x with respect to the idempotent element c_k of $\mathcal{V}$ having trace $k < l$, which we write as x_1, x_{12} and x_0. A similar decomposition can be specified for ξ. We observe that if $x \in \Omega$ then x_1 and x_0 admit inverse in $\mathcal{V}(c_k, 1)$ and $\mathcal{V}(c_k, 0)$, respectively; see MN (Lemma 4).

From MN (Theorem 7), it follows that $x_1 \sim W_k(p, \phi^k)$, with

$$\phi^k = \xi_1 - P(\xi_{12})\xi_0^{-1} = [(\xi^{-1})_1]^{-1}, \tag{18}$$

where the last equality is based on MN [formula (3.4)]. Notice that $-\phi^k$ is equal to the natural parameter of the marginal distribution of x_1.

The fact that the first Peirce component is still distributed like a Wishart makes an iterative decomposition feasible. In order to do this, we shall need slightly more elaborate notation. More precisely, set $x = x_l$ and denote its Peirce components, relative to the idempotent of rank $l-1$, c_{l-1}, by $x_{l,1}$, $x_{l,12}$ and $x_{l,0}$, where $x_{l,1} \in \mathcal{V}(c_{l-1}, 1)$. We set $x_{l,1} = x_{l-1}$ and proceed by decomposing x_{l-1}, with respect to c_{l-2}, into $x_{l-1,1}$, $x_{l-1,12}$ and $x_{l-1,0}$; again, we set $x_{l-1,1} = x_{l-2}$ and repeat the decomposition until $x_{2,1} = x_1$. Note that from now on x_1 represents the first Peirce component of x_2 and no longer the first Peirce component of an arbitrary decomposition of x as in Section 3.1.

Let us now consider the Peirce decomposition of the parameter indexing the Wishart distribution. Start with $\xi = \phi^l$ and denote its Peirce components by $\phi_{l,1}$, $\phi_{l,12}$ and $\phi_{l,0}$, where $\phi_{l,1} \in \mathcal{V}(c_{l-1}, 1)$. Using (18), the parameter associated to the marginal distribution of x_{l-1} is given by

$$\phi^{l-1} = \phi_{l,1} - P_l(\phi_{l,12})\phi_{l,0}^{-1}, \tag{19}$$

where P_l is the quadratic application on $\mathcal{V}_l = \mathcal{V}$. We now decompose ϕ^{l-1}, with respect to c_{l-2}, into its Peirce components and derive the parameter ϕ^{l-2} corresponding to the marginal distribution of x_{l-2} and so on. In general, we have

$$\phi^k = \phi_{k+1,1} - P_{k+1}(\phi_{k+1,12})\phi_{k+1,0}^{-1} = \big[[(\phi^{k+1})^{-1}]_{k+1,1}\big]^{-1}, \tag{20}$$

where $\phi_{k+1,1}$, $\phi_{k+1,12}$ and $\phi_{k+1,0}$ are the Peirce components of ϕ^{k+1} and P_{k+1} is the quadratic application on the Jordan subalgebra $\mathcal{V}(c_{k+1}, 1)$. Clearly, (18) and (20) are identical, the only difference being that the latter originates from an iterative procedure.

Notice that $\phi^k \in \mathcal{V}(c_k, 1)$ (see MN, page 877) and therefore is variation independent of $\phi_{k+1,12}, \phi_{k+1,0}$. Applying a further Peirce decomposition to ϕ^k, one concludes that $(\phi_{k,12}, \phi_{k,0})$ is variation independent of $\phi_{k+1,12}, \phi_{k+1,0}$.

THEOREM 1. *The family $W_l(p, \xi)$ is l-conditionally reducible. Specifically, there exists a reparameterization $\xi \mapsto \phi$ such that the density of $x \sim W_l(p, \xi)$ can be written as*

$$p\big(x|\xi(\phi)\big) = p(x_1|\phi_1) \prod_{k=2}^{l} p(x_{k,12}, x_{k,0}|x_{k-1}; \phi_k), \tag{21}$$

where $x_{k-1} = x_{k,1}$ and $x_{k,1}, x_{k,12}, x_{k,0}$ are the components of the Peirce decomposition of x_k, with respect to the idempotent element of $\mathcal{V}(c_k, 1)$ having rank $k-1$; furthermore, $\phi = (\phi_1, \dots, \phi_l)$, with $\phi_1 = \phi^1$ and $\phi_k = (\phi_{k,12}, \phi_{k,0})$, $k = 2, \dots, l$, with the ϕ_k's variation independent.

Moreover, $x_1 \sim W_1(p, \phi_1)$ and

$$\begin{aligned} &p(x_{k,12}, x_{k,0}|x_{k-1}; \phi_k) \\ &\quad = \frac{h_{k,p}(x_k)}{h_{k-1,p}(x_k)} \exp\{-(\phi_{k,12}, x_{k,12}) - (\phi_{k,0}, x_{k,0}) - M_k(\phi_k, x_{k-1}, p)\}, \end{aligned} \tag{22}$$

where $M_k(\phi_k, x_{k-1}, p) = (P_k(\phi_{k,12})\phi_{k,0}^{-1}, x_{k-1}) - p \log \det_0(\phi_{k,0})$ and $h_{k,p}$ is defined in accordance with (11) *on the Jordan subalgebra of rank k.*

PROOF. Consider the Peirce components of $x = x_l$, relative to the idempotent having rank $l-1$, and write $p(x_l|\xi) = p(x_{l,12}, x_{l,0}|x_{l-1,1}; \xi)p(x_{l-1,1}|\xi)$, where, as usual, $x_{l-1} = x_{l,1}$. Proceed now by a further factorization of the latter density using the Peirce decomposition of x_{l-1}, relative to the idempotent having rank $l-2$, and so on.

The kth conditional density $p(x_{k,12}, x_{k,0}|x_{k-1}; \phi_k(\xi))$, $k = 2, \dots, l$, is obtained as the ratio of the marginal densities of x_k and x_{k-1}, which are both general Wishart.

Expression (22) follows by simplifying the above ratio using the decomposition provided in MN (Theorem 7) and recalling that

$$\begin{aligned} (\phi^{k-1}, x_{k-1}) &= \big((\phi_{k,1} - P_k(\phi_{k,12})\phi_{k,0}^{-1}), x_{k-1}\big) \\ &= (\phi_{k,1}, x_{k-1}) - \big(P_k(\phi_{k,12})\phi_{k,0}^{-1}, x_{k-1}\big), \end{aligned}$$

where the first equality follows from (20).

Finally, the distribution of x_1 and the structure of the kth conditional density reveal that it belongs to an exponential family, and variation independence of the parameters ϕ_k is a consequence of the observation following (20). □

3.4. *Parameterizations and relations with other work.* The structure of conditional reducibility of the Wishart family exhibited in Theorem 1 defines a reparameterization $\xi \to \phi$ involving a decomposition of the symmetric cone Ω. Generalizing previous work by Andersson and Perlman (1993) on lattice conditional independence models, Massam and Neher (1998), Section 2.3, provides a decomposition of the so-called AP cone $\Omega(\mathcal{K})$, which represents the cone associated to a finite distributive lattice $\mathcal{K}$. When $\mathcal{K}$ is chosen to be $(\varnothing = K_0, K_1, \dots, K_l = \{1, \dots, l\})$ with $(K_0 \subset K_1 \subset \cdots \subset K_{l-1} \subset K_l)$, then $\mathcal{K}$ is a chain, and $\Omega(\mathcal{K}) = \Omega$. In this case, it is possible to establish a comparison between the ϕ-parameterization and the $\mathcal{K}$-parameterization.

For concreteness and to facilitate comparison with previous work, we discuss in some detail the case where Ω is the space of real s.p.d. $l \times l$ matrices so that the general Wishart reduces to the usual Wishart $W_l^*(w, \Sigma)$. In this case, $\xi = \Sigma^{-1}/2$, and adhering to standard matrix notation, we use X instead of x, where X is an s.p.d. matrix of order l, while X_k denotes the principal submatrix of X of order k. The Peirce components of X_k, relative to the idempotent matrix $c_{k-1} = \mathrm{Diag}(I_{k-1}, 0)$, will be identified by $X_{k,1}$, $X_{k,12}$ and $X_{k,0}$ and, in accordance with the general case, we write $X_{k,1} = X_{k-1}$. A similar notation will be used for the decomposition of $\Sigma^{-1}/2$. Since $X_k \sim W_k^*(w, \Sigma_k)$, it follows that $\phi^k = \Sigma_k^{-1}/2$ can be partitioned as

$$\tfrac{1}{2}\Sigma_k^{-1} = \phi^k = \begin{bmatrix} \phi^{k-1} & \phi_{k\backslash k} \\ \phi_{k\backslash k}^T & \phi_{kk} \end{bmatrix}. \tag{23}$$

Recalling that ϕ_k is equal to the last two Peirce components of ϕ^k relative to c_{k-1}, one has in this case $\phi_k = (\phi_{k\backslash k}^T, \phi_{kk})^T$.

Consider now the recursive factorization of the joint density of a normal vector $(Y_1, \dots, Y_l)$ with zero mean and covariance matrix Σ into a product of l univariate conditional normal densities. The distribution of Y_k given $Y_1, \dots, Y_{k-1}$ can be naturally indexed in terms of the "regression" parameter and of the conditional variance

$$\beta_k = \Sigma_{k-1}^{-1}\sigma_{k\backslash k}, \qquad \sigma^2_{k\cdot 1,\dots,k-1} = (\sigma_{kk} - \sigma_{k\backslash k}^T \Sigma_{k-1}^{-1} \sigma_{k\backslash k}), \tag{24}$$

where

$$\Sigma_k = \begin{bmatrix} \Sigma_{k-1} & \sigma_{k\backslash k} \\ \sigma_{k\backslash k}^T & \sigma_{kk} \end{bmatrix}. \tag{25}$$

Expression (24) corresponds to the $\mathcal{K}$-parameterization of Andersson and Perlman (1993) previously described. This reparameterization of Σ had also been used in Shachter and Kenley (1989) and Geiger and Heckerman (1994).

Using (23) and standard results on the inverse of partitioned matrices, one obtains $\phi_{kk} = (1/2)(\sigma^2_{k\cdot 1,\dots,k-1})^{-1}$. Furthermore, one can check that

$$\phi_{k\setminus k}\phi_{kk}^{-1} = -\Sigma_{k-1}^{-1}\sigma_{k\setminus k} = \beta_k, \tag{26}$$

whence $\phi_{k\setminus k} = -\beta_k/(2\sigma^2_{k\cdot 1,\dots,k-1})$.

A further related parameterization refers to the Cholesky decomposition of $\Sigma^{-1} = \triangle^T\triangle$, where $\triangle$ is an upper triangular matrix with positive diagonal entries. It turns out that the kth row of $\triangle$ is $(\tilde{\beta}_k/\sqrt{\sigma^2_{k\cdot 1,\dots,k-1}}, 1/\sqrt{\sigma^2_{k\cdot 1,\dots,k-1}})$, where $\tilde{\beta}_k$ is the "regression" parameter of $Y_k|Y_{k-1},\dots,Y_1$; see Wermuth (1980). Further properties of the matrix $\triangle$ are investigated in Roverato (2000).

4. Enriched standard conjugate priors for the Wishart family. To construct an enriched standard conjugate family on the cr-parameter ϕ, we apply the theory discussed in Section 2.2 to the likelihood function given in Theorem 1. We note that this function corresponds, through sufficiency, either to the case in which we have a random sample from a multivariate normal distribution, with known mean, or, alternatively, to the case in which we have a random sample from a Wishart distribution. Only the interpretation of x and p will change.

DEFINITION 1. Given the Wishart family $W_l(p,\xi)$, consider the reparameterization $\xi \mapsto \phi$, where ϕ is the cr-parameter. Then the enriched standard conjugate family on ϕ has density, relative to Lebesgue measure, given by

$$\begin{aligned}\pi_\phi(\phi|t,p') &\propto \exp\{-(\phi_1,t^1) + p'_1 \log\det{}_1(\phi_1)\}\\ &\quad\times \prod_{k=2}^{l} \exp\{-(\phi_{k,12},t^k_{k,12}) - (\phi_{k,0},t^k_{k,0})\\ &\qquad - (P_k(\phi_{k,12})\phi_{k,0}^{-1},t^k_{k-1}) + p'_k \log\det{}_0(\phi_{k,0})\},\end{aligned} \tag{27}$$

where $t^k_{k-1} = t^k_{k,1}$, $t^k_{k,1}, t^k_{k,12}, t^k_{k,0}$ are the Peirce components of $t^k \in \mathcal{V}(c_k,1)$, $k = 2,\dots,l$, and $p' = (p'_1,\dots,p'_l)^T \in \mathbb{R}^l$.

In accordance with the definition of conjugate families described in Section 2.2, no constraints are imposed on the hyperparameters (t^k, p'_k).

Family (27) specializes to the simple enriched standard conjugate family when each component of the set of hyperparameters $\{t^k, k = 1,\dots,l\}$ is derived from a *unique* element $t \in \mathcal{V}$ by means of a sequence of nested Peirce decompositions analogously to those operated on x. In this case, $t^1 = t_{1,1}$, $t^k = (t_{k,1}, t_{k,12}, t_{k,0})$,

$k = 2, \ldots, l$. This family of priors may prove to be sufficiently flexible in many circumstances, while requiring a much smaller set of prior assignments.

Starting from the prior family defined in (27), one can obtain the induced enriched standard conjugate family on the mean parameter using (6). From the structure of $\det(V(\mu))$ exhibited in (17), it follows that μ and ϕ are conjugate parameterizations; that is, the conjugate family on μ constructed directly from the likelihood function for μ is the same as the induced one. For a discussion of the issue of conjugate parameterization, see Gutiérrez-Peña and Smith (1995).

THEOREM 2. *Let $\phi = (\phi_1, \ldots, \phi_l)$ be distributed according to the enriched standard conjugate family* (27). *The family is proper if $t^k \in \Omega_k$, where Ω_k is the symmetric cone associated to $\mathcal{V}(c_k, 1)$, $k = 1, \ldots, l$, and $p'_k > -d_k/k$. Furthermore*:

(i) $\phi_1 \sim W_1(p'_1 + 1, t^1)$;

(ii) $\phi_{k,12}|\phi_{k,0} \sim N(-\{\phi_{k,0}\, t^k_{k,12}(t^k_{k-1})^{-1}\}, (4L(\phi^{-1}_{k,0})L(t^k_{k-1}))^{-1})$, $k = 2, \ldots, l$, *where* $\{xyz\} = P(x+z)y - P(x)y - P(z)y$;

(iii) $\phi_{k,0} \sim W_1(p'_k + d_k/k, (t^k_{k,0} - P_k(t^k_{k,12})(t^k_{k-1})^{-1}))$, $k = 2, \ldots, l$.

PROOF. (i) Straightforward by inspection of the kernel of the distribution of ϕ_1 in (27) and the definition of the Wishart family (10).

(ii) First of all notice that

$$\begin{aligned}(\phi_{k,12}, t^k_{k,12}) &+ \big(P_k(\phi_{k,12})\phi^{-1}_{k,0}, t^k_{k-1}\big) + \big(\phi_{k,0}, P_k(t^k_{k,12})(t^k_{k-1})^{-1}\big)\\ &= \big(P_k\big(\phi_{k,12} + \{\phi_{k,0}\, t^k_{k,12}(t^k_{k-1})^{-1}\}\big)\phi^{-1}_{k,0}, t^k_{k-1}\big);\end{aligned}$$

see MN [formula (3.39)]. Then the exponent of the kth term in (27), omitting the last component $p'_k \log \det_0(\phi_{k,0})$, can be written as

$$\begin{aligned}&\big(P_k\big(\phi_{k,12} + \{\phi_{k,0}\, t^k_{k,12}(t^k_{k-1})^{-1}\}\big)\phi^{-1}_{k,0}, t^k_{k-1}\big) - \big(\phi_{k,0}, P_k(t^k_{k,12})(t^k_{k-1})^{-1}\big)\\ &\quad = \tfrac{1}{2}\big[\phi_{k,12} + \{\phi_{k,0}(t^k_{k-1})^{-1} t^k_{k,12}\},\\ &\qquad\quad 4L(\phi^{-1}_{k,0})L(t^k_{k-1})\big(\phi_{k,12} + \{\phi_{k,0}(t^k_{k-1})^{-1} t^k_{k,12}\}\big)\big]\\ &\qquad - \big(\phi_{k,0}, P_k(t^k_{k,12})(t^k_{k-1})^{-1}\big).\end{aligned} \tag{28}$$

Recalling the structure of the density of a normal distribution on $\mathcal{V}(c, \frac{1}{2})$ defined in MN (page 897), one recognizes that minus the expression in (28) represents the exponential term of a normal distribution having expectation $-\{\phi_{k,0}(t^k_{k-1})^{-1} t^k_{k,12}\}$ and covariance operator $(4L(\phi^{-1}_{k,0})L(t^k_{k-1}))^{-1}$. To recover the normal density, one needs to multiply (and subsequently divide) by $[\det(4L(\phi^{-1}_{k,0})L(t^k_{k-1}))]^{1/2} = [\det_0(\phi^{-1}_{k,0})]^{(k-1)g/2}[\det_1(t^k_{k-1})]^{g/2}$; see again MN (page 897).

(iii) Setting aside the normal component previously constructed, which represents the conditional density of $\phi_{k,12}|\phi_{k,0}$, the remaining terms must amount to the kernel of the marginal density of $\phi_{k,0}$, which is

$$\exp\{-(\phi_{k,0}, t^k_{k,0}) + \big((t^k_{k-1})^{-1}, P_k(t^k_{k,12})\phi_{k,0}\big)\} \\ \times [\det{}_0(\phi_{k,0})]^{p'_k}[\det{}_0(\phi_{k,0})]^{(k-1)g/2}. \tag{29}$$

Since $((t^k_{k-1})^{-1}, P_k(t^k_{k,12})\phi_{k,0}) = (\phi_{k,0}, P_k(t^k_{k,12})(t^k_{k-1})^{-1})$, (29) becomes

$$\exp\{-\big(\phi_{k,0}, \big(t^k_{k,0} - P_k(t^k_{k,12})(t^k_{k-1})^{-1}\big)\big)\}[\det(\phi_{k,0})]^{p'_k+(k-1)g/2},$$

which corresponds to the kernel of a

$$W_1\bigg(p'_k + \frac{(k-1)g}{2} + 1, t^k_{k,0} - P_k(t^k_{k,12})(t^k_{k-1})^{-1}\bigg).$$

Since $d_k = k + gk(k-1)/2$, the result follows. $\square$

Notice that the hyperparameter $t^k_{k,0} - P_k(t^k_{k,12})(t^k_{k-1})^{-1}$ appearing in (iii) can be written as $[(t^k)^{-1}_0]^{-1}$; see (20).

The prior family (27) is conjugate and therefore the posterior family has the same structure. Updating is trivial under a likelihood expressed by (21), see also (22), namely,

$$\begin{aligned} &t^1 \mapsto (t^1 + x_1), \\ &t^k_{k-1} \mapsto (t^k_{k-1} + x_{k-1}), \quad t^k_{k,12} \mapsto (t^k_{k,12} + x_{k,12}), \quad t^k_{k,0} \mapsto (t^k_{k,0} + x_{k,0}), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad k = 2, \dots, l, \\ &p'_k \mapsto (p'_k + p), \qquad k = 1, \dots, l. \end{aligned} \tag{30}$$

4.1. *Enriched conjugate family on real s.p.d. matrices.* When specializing (27) to the usual case, we shall denote the hyperparameters by T^k, a real symmetric matrix of order k, and w'_k. In this case, Theorem 2 yields the following result.

COROLLARY 1. *The enriched standard conjugate family for the cr-parameter ϕ of a usual Wishart is proper if, for $k = 1, \dots, l$, T^k is a $k \times k$ s.p.d. matrix and $w'_k > -(k+1)$. Furthermore, it is characterized by*

(i) $$\phi_1 \sim Ga\bigg(\frac{w'_1 + 2}{2}, T^1\bigg),$$

(ii) $\phi_{k\backslash k}|\phi_{kk} \sim N_{k-1}\big(-\phi_{kk}(T^k_{k-1})^{-1}t^k_{k\backslash k}, \frac{1}{2}\phi_{kk}(T^k_{k-1})^{-1}\big), \qquad k=2,\ldots,l,$

(iii) $\phi_{kk} \sim Ga\left(\dfrac{w'_k+k+1}{2}, t^k_{kk}-(t^k_{k\backslash k})^T(T^k_{k-1})^{-1}t^k_{k\backslash k}\right), \qquad k=2,\ldots,l,$

where T^k_{k-1} *is the principal matrix of* T^k *having order* $(k-1)$ *and* $((t^k_{k\backslash k})^T, t^k_{kk})^T$ *is the last column of* T^k.

Notice that the term $t^k_{kk}-(t^k_{k\backslash k})^T(T^k_{k-1})^{-1}t^k_{k\backslash k} = 1/[(T^k)^{-1}]_{kk}$, where $[(T^k)^{-1}]_{kk}$ is the (k,k)th element of $(T^k)^{-1}$.

Given a data matrix X, updating of the family described in Corollary 1 mimicks that we obtain in the general case, namely, $T^k \mapsto (T^k + X_k)$ and $w'_k \mapsto (w'_k + w)$.

Brown, Le and Zidek (1994) introduced a generalized inverted Wishart (GIW) family as a prior for a covariance matrix Σ, in order to overcome the well-known deficiencies associated with the standard conjugate family represented by the inverted Wishart.

They first partition the data matrix into two blocks and define the GIW for this case. Next, they extend their procedure in a recursive fashion. Although not immediate, it can be shown that their GIW family coincides with that in Corollary 1, when the number of blocks is equal to l. Accordingly, we shall denote it with the same acronym and write $\Sigma \sim \mathrm{GIW}(\{T^k, w'_k;\ k=1,\ldots,l\})$.

If in Corollary 1 for each $k=1,\ldots,l$, one sets $T^k = T_k$, the principal $k\times k$ matrix of a unique matrix T, one recovers the simple enriched standard conjugate family. If one further takes $w'_k = w'$ for each k, then the GIW family reduces to the standard inverted Wishart. In this case, the distributional results of Corollary 1 agree with those already available in the literature concerning the distribution induced by a standard inverted Wishart prior for Σ on alternative parameterizations; see Section 3.4 and the references therein.

A useful feature of the GIW prior is that it admits closed-form expressions for the first moment of Σ and Σ^{-1}.

PROPOSITION 1. *Let* Σ *be distributed according to a* $\mathrm{GIW}(\{T^k, w'_k > -(k+1), k=1,\ldots,l\})$, *with* T^k *s.p.d. matrices. Then*

$$E(\sigma_{11}) = \frac{T^1}{w'_1}.$$

For $k=2,\ldots,l$,

$$E(\sigma_{k\backslash k}) = E(\Sigma_{k-1})(T^k_{k-1})^{-1}t^k_{k\backslash k},$$

$$E(\sigma_{kk}) = \frac{1}{w'_k+k-1}\frac{1}{((T^k)^{-1})_{kk}}\big\{\mathrm{tr}\big[(E(\Sigma_{k-1})(T^k_{k-1})^{-1}\big]+1\big\} + E(\sigma_{k\backslash k})^T(T^k_{k-1})^{-1}t^k_{k\backslash k},$$

where $\sigma_{11} = \Sigma_1$.

PROOF. The proof of the expectation of σ_{11} is immediate, since $\phi_1 = 1/(2\sigma_{11})$.

We now consider the expectation of $\sigma_{k\backslash k}$ and of σ_{kk}. Partition Σ_k as in (25). Using (26), one obtains

$$\sigma_{k\backslash k} = -(\Sigma_{k-1})\phi_{k\backslash k}(\phi_{kk})^{-1}. \tag{31}$$

Recalling that $\phi_{kk} = 1/(2\sigma^2_{k\cdot 1,\dots,k-1})$ and using (24) and (26), one obtains

$$\sigma_{kk} = \frac{1}{2}\frac{1}{\phi_{kk}} + (\sigma_{k\backslash k})^T(\Sigma_{k-1})^{-1}\sigma_{k\backslash k} = \frac{1}{2}\frac{1}{\phi_{kk}} + \left\{\frac{1}{\phi_{kk}^2}\phi_{k\backslash k}^T(\Sigma_{k-1})\phi_{k\backslash k}\right\}. \tag{32}$$

Taking expectations on both sides of (31) with respect to the distribution on ϕ described in Corollary 1, one obtains

$$E(\sigma_{k\backslash k}) = -E(\Sigma_{k-1})E\{(\phi_{kk})^{-1}E(\phi_{k\backslash k}|\phi_{kk})\} = E(\Sigma_{k-1})(T^k_{k-1})^{-1}t^k_{k\backslash k}.$$

Consider now the expectation of σ_{kk}. From (32), it follows that

$$E(\sigma_{kk}) = \frac{1}{2}E\left(\frac{1}{\phi_{kk}}\right) + E\left\{\frac{1}{\phi_{kk}^2}E\left(\phi_{k\backslash k}^T E(\Sigma_{k-1})\phi_{k\backslash k}|\phi_{kk}\right)\right\}.$$

Since

$$\begin{aligned}
&E\left(\phi_{k\backslash k}^T E(\Sigma_{k-1})\phi_{k\backslash k}|\phi_{kk}\right)\\
&\quad = E\{\operatorname{vec}[(\phi_{k\backslash k}\phi_{k\backslash k}^T)]^T \operatorname{vec}[E(\Sigma_{k-1})|\phi_{kk}]\}\\
&\quad = E\left\{\operatorname{vec}\left[\operatorname{cov}(\phi_{k\backslash k}|\phi_{kk}) + E(\phi_{k\backslash k})\left(E(\phi_{k\backslash k})\right)^T|\phi_{kk}\right]^T\right\}\operatorname{vec}\left(E(\Sigma_{k-1})\right)\\
&\quad = \left[\operatorname{vec}\left[\tfrac{1}{2}\phi_{kk}(T^k_{k-1})^{-1} + \phi_{kk}^2(T^k_{k-1})^{-1}t^k_{k\backslash k}(t^k_{k\backslash k})^T(T^k_{k-1})^{-1}\right]\right]^T\\
&\qquad \times \operatorname{vec}\left(E(\Sigma_{k-1})\right),
\end{aligned}$$

we get

$$\begin{aligned}
E(\sigma_{kk}) = E&\left\{\frac{1}{\phi_{kk}}\left[\frac{1}{2}\left[\operatorname{vec}(T^k_{k-1})^{-1}\right]^T\operatorname{vec}\left(E(\Sigma_{k-1})\right)\right]\right\}\\
&+ \operatorname{vec}\left[(T^k_{k-1})^{-1}t^k_{k\backslash k}(t^k_{k\backslash k})^T(T^k_{k-1})^{-1}\right]^T\operatorname{vec}\left(E(\Sigma_{k-1})\right) + \frac{1}{2}E\left(\frac{1}{\phi_{kk}}\right).
\end{aligned}$$

From

$$E\left(\frac{1}{\phi_{kk}}\right) = 2\frac{t^k_{kk} - (t^k_{k\backslash k})^T(T^k_{k-1})^{-1}t^k_{k\backslash k}}{w'_k + k - 1} = 2\frac{([(T^k)^{-1}]_{kk})^{-1}}{w'_k + k - 1},$$

one obtains

$$
\begin{aligned}
E(\sigma_{kk}) &= \frac{([(T^k)^{-1}]_{kk})^{-1}}{w_k' + k - 1} \operatorname{vec}\big[(T_{k-1}^k)^{-1}\big]^T \operatorname{vec}[E(\Sigma_{k-1})] \\
&\quad + \operatorname{vec}\big[(T_{k-1}^k)^{-1} t_{k\backslash k}^k (t_{k\backslash k}^k)^T (T_{k-1}^k)^{-1}\big]^T \operatorname{vec}[E(\Sigma_{k-1})] \\
&\quad + \frac{([(T^k)^{-1}]_{kk})^{-1}}{w_k' + k - 1} \\
&= \frac{([(T^k)^{-1}]_{kk})^{-1}}{w_k' + k - 1} \operatorname{tr}\big((T_{k-1}^k)^{-1}\big(E(\Sigma_{k-1})\big)\big) \\
&\quad + \operatorname{tr}\big((T_{k-1}^k)^{-1} t_{k\backslash k}^k (t_{k\backslash k}^k)^T (T_{k-1}^k)^{-1}\big(E(\Sigma_{k-1})\big)\big) + \frac{([(T^k)^{-1}]_{kk})^{-1}}{w_k' + k - 1} \\
&= \frac{([(T^k)^{-1}]_{kk})^{-1}}{w_k' + k - 1} \big\{\operatorname{tr}\big((T_{k-1}^k)^{-1}\big(E(\Sigma_{k-1})\big)\big) + 1\big\} \\
&\quad + \operatorname{tr}\big((T_{k-1}^k)^{-1} t_{k\backslash k}^k E(\sigma_{kk})\big),
\end{aligned}
$$

where the second equality sign is based on $\operatorname{vec}(A^T)\operatorname{vec}(B) = \operatorname{tr}(A^T B)$. Since, for vectors x, y and a symmetric matrix A, $\operatorname{tr}(Axy^T) = \operatorname{tr}(y^T Ax) = y^T Ax$, the second term in the last line may be alternatively written as $(E(\sigma_{kk}))^T (T_{k-1}^k)^{-1}(t_{k\backslash k}^k)$. □

We finally consider the Bayes estimator of Σ under the GIW distribution discussed above. One commonly employed loss function [see Yang and Berger (1994) and Daniels and Kass (1999)] is

$$
L_1(\hat{\Sigma}, \Sigma) = \operatorname{tr}(\hat{\Sigma}\Sigma^{-1}) - \log\big(\det(\hat{\Sigma}\Sigma^{-1})\big) - l. \tag{33}
$$

The Bayes estimate of Σ under L_1 is given by $(E(\Sigma^{-1})|\text{data})^{-1}$. Because of conjugacy, to compute this estimate, it is enough to provide a formula for $(E(\Sigma^{-1}))$ under the GIW family. The subsequent inversion can be numerically evaluated.

PROPOSITION 2. *Let Σ be distributed according to a* GIW$(\{T^k, w_k' > -(k+1), k = 1, \ldots, l\})$, *with T^k s.p.d. matrices. Then, for $k = 2, \ldots, l$,*

$$
E(\Sigma_k^{-1}) = \begin{pmatrix} E(\Sigma_{k-1}^{-1} + Q^k) & 2E(\phi_{k\backslash k}) \\ 2E(\phi_{k\backslash k})^T & 2E(\phi_{kk}) \end{pmatrix},
$$

with

$$
E(\Sigma_1^{-1}) = E(\sigma_{11}^{-1}) = 2E(\phi_1) = \frac{w_1' + 2}{T^1}
$$

and

$$2E(\phi_{k\backslash k}) = -(w'_k + k + 1)\big[(T^k)^{-1}\big]_{kk}(T^k_{k-1})^{-1}t^k_{k\backslash k},$$

$$2E(\phi_{kk}) = (w'_k + k + 1)\big[(T^k)^{-1}\big]_{kk},$$

$$E(Q^k) = (T^k_{k-1})^{-1}\big[I_{k-1} + (w'_k + k + 1)\big[(T^k)^{-1}\big]_{kk}t^k_{k\backslash k}(t^k_{k\backslash k})^T(T^k_{k-1})^{-1}\big].$$

PROOF. The expected value of Σ_1^{-1} follows immediately from (i) of Corollary 1.

Recalling the relationship between $(\phi_{k\backslash k}, \phi_{kk})$ and $(\beta_k, \sigma^2_{k\cdot 1,\dots,k-1})$ expressed in (24) and (26), it follows from Shachter and Kenley [(1989), Section 5] that

$$\Sigma_k^{-1} = \begin{pmatrix} \Sigma_{k-1}^{-1} + Q^k & 2\phi_{k\backslash k} \\ 2(\phi_{k\backslash k})^T & 2\phi_{kk} \end{pmatrix},$$

where $Q^k = 2\phi_{k\backslash k}(\phi_{k\backslash k})^T/\phi_{kk}$.

The expected value of ϕ_{kk} follows immediately from (iii) of Corollary 1.

Consider now $E(\phi_{k\backslash k})$. One can write

$$2E(\phi_{k\backslash k}) = 2E\{E(\phi_{k\backslash k}|\phi_{kk})\} = 2E\big\{-\phi_{kk}(T^k_{k-1})^{-1}t^k_{k\backslash k}\big\},$$

where the last equality holds by (ii) of Corollary 1. Taking expectations w.r.t. the distribution of ϕ_{kk} gives the desired result. Consider now $E(Q^k)$. Formally, this involves the expectation of uu^T/v, with u a vector and v a scalar, such that $u|v \sim N(vm, vS)$ and $v \sim Ga(a, b)$. Then $E(uu^T/v) = S + (a/b)mm^T$. Substituting m, S, a and b with the corresponding values obtained from (ii) and (iii) of Corollary 1, the result follows. □

5. Reference priors for the Wishart family. In this section, we discuss the Jeffreys and reference priors and show some useful connections with the enriched standard conjugate family.

Consider first the Jeffreys prior for a general Wishart family, $W_l(p, \xi)$. Recall that the Jeffreys prior for the natural parameter θ of an exponential family (1) is given by

$$\left\{\det\left(\frac{\partial^2 M(\theta)}{\partial\theta^T\partial\theta}\right)\right\}^{1/2} = \big\{\det\big(V(\mu(\theta))\big)\big\}^{1/2};$$

combining this result with the Jacobian of the transformation $\theta \to \mu$ already derived before formula (6), one obtains that the Jeffreys prior on the mean parameter is

$$\pi^J_\mu(\mu) \propto \det(V(\mu))^{-1/2} = (\det\mu)^{-d/l}, \tag{34}$$

where the equality sign holds because of (16).

We now turn to reference priors. Block diagonality of the Fisher information H^{ϕ} follows immediately from the property of l-conditional reducibility of the Wishart family. Furthermore, the kth block H_k^{ϕ} is given by the expected value of the second differential of $M_k(\phi_k; x_{k-1}; p)$. Using (21), one concludes that the marginal distribution of x_{k-1} only depends on $(\phi_1, \dots, \phi_{k-1}) = \phi_{[k-1]}$, and thus H_k^{ϕ} only depends on $\phi_{[k]}$.

LEMMA 1. *Consider a Wishart family and let ϕ denote its cr-parameter defined in Theorem* 1. *If $H_k^{\phi}(\phi_{[k]})$ denotes the kth diagonal block of the Fisher information H^{ϕ}, one has*

$$\det\left(H_k^{\phi}(\phi_{[k]})\right) = a_k(\phi_k) b_k(\phi_{[k-1]}), \qquad k = 2, \dots, l. \tag{35}$$

PROOF. Recall that $x_k \sim W_k(p, \phi^k)$ and consider the Peirce decomposition of x_k and ϕ^k with respect to the idempotent c_{k-1}. Consider the mapping $\phi^k \mapsto (\phi^{k-1}, \phi_{k,12}, \phi_{k,0})$, where ϕ^{k-1} indexes the Wishart marginal distribution of x_{k-1}.

Write $p(x_k | \phi^k(\phi^{k-1}, \phi_k)) = p(x_{k-1} | \phi^{k-1}) p(x_{k,12}, x_{k,0} | x_{k-1}; \phi_k)$. Because of the functional independence between ϕ^{k-1} and ϕ_k, the structure of the associated Fisher information H^{ϕ^{k-1}, ϕ_k} is "block diagonal" with blocks $H^{\phi^{k-1}}(\phi^{k-1})$ and $H_k^{\phi}(\phi_{[k]})$, where the first one is the Fisher information relative to the distribution of x_{k-1}. Because of block diagonality, one has

$$\det\left(H_k^{\phi}(\phi_{[k]})\right) = \frac{\det(H^{\phi^{k-1}, \phi_k}(\phi^{k-1}(\phi_{[k-1]}), \phi_k))}{\det(H^{\phi^{k-1}}(\phi^{k-1}(\phi_{[k-1]})))},$$

where we have explicitly indicated the dependence of the "marginal parameters" ϕ^{k-1} on the cr-parameter $\phi_{[k-1]}$.

To establish (35), it is enough to factor $\det(H^{\phi^{k-1}, \phi_k})(\cdot)$ in terms of a function of ϕ^{k-1} and a function of ϕ_k. Recalling the expression of ϕ^{k-1} provided in (20), it follows from MN (Proposition 1a) that

$$\tau\left(-2\phi_{k,0}^{-1}\phi_{k,12}\right)\phi^k = \phi^{k-1} \oplus \phi_{k,0}, \tag{36}$$

where $\tau(z)$, $z \in \mathcal{V}(c_{k-1}, 1/2)$, denotes the Frobenius transformation. Since $\det(\tau(z)x) = \det x$ [see MN (Lemma 3)] using (36) and MN (Proposition 1b), one obtains

$$\det_1(\phi^k) = \det_1(\phi^{k-1}) \det_0(\phi_{k,0}). \tag{37}$$

Furthermore, the Jacobian of the mapping $\phi^k \mapsto (\phi^{k-1}, \phi_{k,12}, \phi_{k,0})$ is 1 [see MN (proof of Theorem 7)] so that $H^{\phi^{k-1}, \phi_k} = H^{\phi^k}$.

Using (15), the Fisher information relative to ϕ^k can be written as

$$
\begin{aligned}
&\det\big(H^{\phi^{k-1},\phi_k}(\phi^{k-1},\phi_k)\big)\\
&\quad = \det\big(H^{\phi^k}(\phi^k(\phi^{k-1},\phi_k))\big)\\
&\quad = \det\bigg(\frac{1}{p}P_k\big(p\big(\phi^k(\phi^{k-1},\phi_k)\big)^{-1}\big)\bigg) && \text{[by (15)]}\\
&\quad \propto \det\big(\phi^k(\phi^{k-1},\phi_k)\big)^{-2d_k/k} && \text{[by (ii) of Section 3.1]}\\
&\quad = \big(\det_1(\phi^{k-1})\big)^{-2d_k/k}\big(\det_0(\phi_{k,0})\big)^{-2d_k/k} && \text{[by (37)].}
\end{aligned}
$$

Thus, (35) holds with $a_k(\phi_k) = (\det_0(\phi_{k,0}))^{-2d_k/k}$. □

As a consequence, using (8), the following result is established:

THEOREM 3. *Given the Wishart family $W_l(p,\xi)$, consider the reparameterization $\xi \mapsto \phi$, where ϕ is the cr-parameter. Then the l-group reference prior on ϕ is order invariant and has density, relative to Lebesgue measure, given by*

$$
\pi_\phi(\phi) \propto \big(\det_1(\phi_1)\big)^{-1}\prod_{k=2}^{l}\big(\det_0(\phi_{k,0})\big)^{-d_k/k}. \tag{38}
$$

REMARK. The reference prior (38) belongs to the enriched standard conjugate family (27), since it can be obtained by letting t^k, $k=1,\dots,l$, be null, and $p'_k = -d_k/k$, $k=1,\dots,l$.

Recalling the updating rule for the enriched standard conjugate family described in (30), the reference posterior belongs to the family (27) with hyperparameters $(x_1, p-1)$ and, for $k=2,\dots,l$, $\{(x_{k-1}, x_{k,12}, x_{k,0}), p-d_k/k\}$, since the t^k's are null.

From Theorem 2, the reference posterior is always proper since $x_k \in \Omega_k$, and $p - d_k/k > -d_k/k$, that is, $p>0$, $k=1,\dots,l$, a condition trivially satisfied by any Wishart family.

We now consider the reference prior on the mean parameter μ. In general, reference priors on two distinct parameters η and λ, say, are not related. In particular, the reference prior on λ need not coincide with the prior on λ induced by the reference prior on η. However, Yang (1995) and Datta and Ghosh (1996) prove that if λ is grouped and ordered in such a way that the map $\eta \to \lambda$ is block lower triangular, then the prior induced on λ by the reference prior on η coincides with the reference prior on λ. This is indeed the case for the transformation $\phi \to \mu$ as we now show.

Assume that μ is grouped as $(\mu_1, \dots, \mu_l)$, where $\mu_1 = E(x_1|\phi_1)$ and $\mu_k = E(x_{k,12}, x_{k,0}|\phi_k)$, $k = 2, \dots, l$. Using (21), μ_k can be computed as a k-fold conditional expectation, namely,

$$E_{\phi_1}^{x_1}\Big\{E_{\phi_2}^{x_{2,12},x_{2,0}}\{\cdots\{E_{\phi_k}^{x_{k,12},x_{k,0}}(x_{k,12}, x_{k,0}|x_{k-1,1})\}\}\Big\},$$

whence μ_k only depends on $(\phi_1, \dots, \phi_k)$, and thus the map $\phi \mapsto \mu$ is block lower triangular.

THEOREM 4. *Consider the Wishart family $W_l(p, \xi)$. The l-group reference prior on the mean parameter μ, with respect to the order $(\mu_1, \mu_2, \dots, \mu_l)$, has density, relative to Lebesgue measure, given by*

$$(39) \qquad {}_{\mu_1,\mu_2,\dots,\mu_l}\pi_\mu(\mu) \propto (\det(\mu))^{-d/l} \prod_{k=1}^{l-1} \big(\det{}_1(\mu^k)\big)^{-g/2},$$

where $\mu_k = E(x_{k,12}, x_{k,0}|\phi_k)$ and $\mu^k = E(x_k|\phi^k)$.

PROOF. Since the map $\phi \to \mu$ is block lower triangular, the density of the reference prior on μ can be written as

$$\begin{aligned} {}_{\mu_1,\mu_2,\dots,\mu_l}\pi_\mu(\mu) &\propto \pi_\mu^\phi(\mu) \propto \pi_\phi(\phi(\mu))\big(\det(V(\mu))\big)^{-1} && \text{[by (6)]} \\ &\propto (\det(\mu))^{-2d/l} && \text{[by (16)].} \end{aligned}$$

Substituting the expression for the reference prior on ϕ given in (38), one obtains

$$(40) \quad {}_{\mu_1,\mu_2,\dots,\mu_l}\pi_\mu(\mu) \propto \big(\det{}_1\phi_1(\mu)\big)^{-1} \prod_{k=2}^{l} \big(\det{}_0\big(\phi_{k,0}(\mu)\big)\big)^{-d_k/k} (\det \mu)^{-2d/l}.$$

From (37), one has

$$(41) \qquad \det{}_0(\phi_{k,0}) \propto \frac{\det_1(\phi^k)}{\det_1(\phi^{k-1})},$$

and the result follows because $\mu^k = p(\phi^k)^{-1}$, $k = 1, \dots, l$, with, as usual, $\phi^1 = \phi_1$. □

5.1. *Reference priors for real s.p.d. matrices.* In the usual case, a direct specialization of Theorem 3 yields the l-group order-invariant reference prior on ϕ whose density, relative to Lebesgue measure, is

$$(42) \qquad \pi_\phi(\phi) \propto \prod_{k=1}^{l} (\phi_{kk})^{-(k+1)/2}, \qquad \phi_{kk} > 0.$$

We now consider the problem of constructing a reference prior on the covariance matrix Σ, which represents an important issue in applied Bayesian statistics. Yang and Berger (1994) proposed a prior on Σ induced from a reference prior on its eigenvalues. Our general result in Theorem 4 leads to an actual reference prior on Σ since $\mu = w\Sigma$ and $\mu^k = \Sigma_k$.

COROLLARY 2. *Consider the usual Wishart family $W_l^*(w, \Sigma)$. The l-group reference prior on Σ, with respect to the order $\sigma_{11}, \sigma_2, \dots, \sigma_l$, where $\sigma_k = (\sigma_{k\backslash k}^T, \sigma_{kk})^T$, $k = 2, \dots, l$, has density, relative to Lebesgue measure, given by*

$$
{}_{\sigma_{11},\sigma_2,\dots,\sigma_l}\pi_\Sigma(\Sigma) \propto (\det \Sigma)^{-(l+1)/2} \prod_{k=1}^{l-1} (\det \Sigma_k)^{-1/2}. \tag{43}
$$

To appreciate the meaning of (43), assume $(Y_1, \dots, Y_l) \sim N(0, \Sigma)$. In this case, σ_k represents the covariance structure of Y_k with $(Y_1, \dots, Y_{k-1})$ together with the variance of Y_k. As a consequence, the reference prior (43) will be especially appropriate whenever the Y_k's are arranged in such a way that the corresponding covariances σ_k are in decreasing order of inferential importance.

We finally note that the general result of Theorem 2 ensures that the posterior reference for Σ is always proper and is given by a $\mathrm{GIW}(X_k, w - k - 1, k = 1, \dots, l)$. The expectations of Σ_k and $(\Sigma_k)^{-1}$ are analytically available through Propositions 1 and 2, setting $T^k = X_k$, $w_k' = w - k - 1$. Furthermore, the latter implicitly defines the Bayes estimates of Σ under the loss function L_1 given in (33).

We close this section by discussing two useful inferential properties of the reference prior (42), dealing respectively with the marginalization paradoxes and the frequentist validity of Bayesian credible intervals.

Marginalization paradoxes were studied in Dawid, Stone and Zidek (1973) for some typical inferential problems. Consider a real s.p.d. matrix Σ of order l partitioned into four blocks: $\Sigma_{i,j}$, $i, j = 1, 2$, where Σ_{11} is the principal submatrix of order k. Define $\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ and $B_{2|1} = \Sigma_{21}\Sigma_{11}^{-1}$. Dawid, Stone and Zidek [(1973), Appendix 1] provide conditions on the prior for $(\Sigma_{11}, \Sigma_{2|1}, B_{2|1})$ such that the marginalization paradox does not arise when inference is sought on the above parameters.

We show that the reference prior on ϕ in (42), equivalently that on Σ in (43), satisfies these conditions and therefore is paradox free. Consider first Σ_{11}. In this case, the marginalization paradox does not arise, provided the prior density factorizes as $\pi(\Sigma_{11})\pi(\Sigma_{2|1}, B_{2|1})$. Because of conditional reducibility, it is simple to verify that there exists a one-to-one correspondence between $(\phi_1, \dots, \phi_k)$ and Σ_{11}, as well as between $(\phi_{k+1}, \dots, \phi_l)$ and $\Sigma_{2|1}, B_{2|1}$. This ensures that the induced prior on $(\Sigma_{11}, \Sigma_{2|1}, B_{2|1})$ factorizes as required. Next, consider inference on $(\Sigma_{11}, \Sigma_{2|1})$ or on $\Sigma_{2|1}$. In both cases, the paradox does not arise whenever the prior on $B_{2|1}$ is uniform and it can be checked that this condition is again satisfied for the reference prior (42).

With respect to the frequentist validity of inferences based on the reference prior (42), we consider coverage of posterior probability intervals. Specifically, let Π be a prior for $\phi = (\phi_1, \dots, \phi_l)$. Let $(t_1(\phi), \dots, t_s(\phi))$, $s \le l$, be real-valued, twice-continuously differentiable parametric functions of interest. Consider a region T equal to the Cartesian product of s one-sided intervals for each $t_k(\phi)$ such that the posterior probability of T, based on an i.i.d. sample of size n, generated by Π is α. If the confidence level of T, with respect to the sampling distribution, is $\alpha + O(n^{-1})$, then Π is said to be a joint probability matching prior; see Datta (1996). When the above coverage condition is required marginally on each one-sided interval for $t_k(\phi)$, then Π is said to be a simultaneous marginal probability matching prior.

Let $s = l$ and $t_k(\phi) = \phi_k$, $k = 1, \dots, l$. Since the ϕ_k's are globally orthogonal and moreover $H^{\phi}(\phi) = \mathrm{Diag}\{H_1^{\phi}(\phi), \dots, H_r^{\phi}(\phi)\}$, with $\det(H_k^{\phi}(\phi)) = \det(H_k^{\phi}(\phi_{[k]})) = a_k(\phi_k) b_k(\phi_{[k-1]})$, see (35), it follows from Remark 2 in Datta (1996) that the reference prior (42) is both a joint- and a simultaneous marginal-probability matching prior with respect to $(\phi_1, \dots, \phi_l)$.

6. Discussion. In this paper, we have shown that the Wishart family on symmetric cones is conditionally reducible and thus admits a useful reparameterization in terms of orthogonal parameters. Furthermore, we have constructed enriched conjugate and reference priors for some alternative parameterizations. Interesting features of these priors are: (i) the posterior distribution is straightforward to compute and is always proper when derived from a reference prior; (ii) the enriched standard conjugate family exibits great flexibility, which may be very useful in applications; (iii) reference priors are probability matching, that is, they generate posterior credibility intervals that enjoy frequentist validity, and do not suffer from marginalization paradoxes; and (iv) for the usual cone of real s.p.d. matrices, the expression of commonly used Bayes estimators can be provided in closed form.

Recalling that an NEF-HQVF is of Wishart type, it is interesting to emphasize that some structural issues of special relevance to Bayesian inference which hold for NEF-SQVFs [see Consonni, Veronese and Gutiérrez-Peña (2000)] are perfectly mirrored in the HQVF case. The technical reason for this is that $\det(V(\mu))$ can be formally written, both in the SQVF and the HQVF case, as proportional to the likelihood function for μ with fictitious sufficient statistics z and sample size v determined from the variance function $V(\mu)$; see (17) for the HQVF case. Clearly, for conditionally reducible families, $\det(V(\mu))$ can be further factored along the lines of (2) or (5). In particular, for both families:

(a) The parameters θ, ϕ and μ are conjugate with respect to the standard conjugate, and hence the enriched standard conjugate, family since the Jacobians of the transformations $\theta \to \mu$ and $\phi \to \mu$ are proportional to $\det(V(\mu))$.

(b) The Jeffreys prior on the mean parameter μ belongs to the standard conjugate family since it is proportional to $(\det(V(\mu))^{-1}$; see Gutiérrez-Peña and Smith (1997), Section 3.4. Because of (a), the same conclusion holds for the parameters θ and ϕ.

(c) The reference priors on ϕ and μ belong to the enriched standard conjugate family.

The practical usefulness of points (b) and (c) is that one can employ standard results on conjugate families to evaluate aspects of the Jeffreys, and reference, posterior such as propriety and a closed-form expression for the expectation of μ.

Acknowledgments. We are very grateful to one referee for very detailed comments that have greatly improved the presentation of the paper and for suggesting connections with previous work. We also thank Hélène Massam for pointing out a very useful reference when this research was begun.

REFERENCES

ANDERSEN, H. H., HØJBJERRE, M., SØRENSEN, D. and ERIKSEN, P. S. (1995). *Linear and Graphical Models for the Multivariate Complex Normal Distribution. Lecture Notes in Statist.* **101**. Springer, New York.

ANDERSSON, S. A. and PERLMAN, M. D. (1993). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21** 1318–1358.

ANDERSSON, S. A. and PERLMAN, M. D. (1995). Unbiasedness of the likelihood ratio test for lattice conditional independence models. *J. Multivariate Anal.* **53** 1–17.

BAR-LEV, S., BSHOUTY, D., ENIS, P., LETAC, G., LU, I. and RICHARDS, D. (1994). The diagonal multivariate natural exponential families and their classification. *J. Theoret. Probab.* **7** 883–929.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester.

BERGER, J. O. and BERNARDO, J. M. (1992). On the development of reference priors (with discussion). In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 35–60. Clarendon, Oxford.

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 113–147.

BERNARDO, J. M. and SMITH, A. F. M. (1994). *Bayesian Theory*. Wiley, Chichester.

BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families*, *with Applications in Statistical Decision Theory*. IMS, Hayward, CA.

BROWN, P. J., LE, N. D. and ZIDEK, J. M. (1994). Inference for a covariance matrix. In *Aspects of Uncertainty*: *A Tribute to D. V. Lindley* (P. R. Freeman and A. F. M. Smith, eds.) 77–92. Wiley, Chichester.

CASALIS, M. (1991). Les familles exponentielles à variance quadratique homogène sont des lois de Wishart sur un cône symétrique. *C. R. Acad. Sci. Paris Sér. I Math.* **312** 537–540.

CASALIS, M. (1996). The $2d + 4$ simple quadratic natural exponential families on $\mathbb{R}^d$. *Ann. Statist.* **24** 1828–1854.

CASALIS, M. and LETAC, G. (1996). The Lukacs–Olkin–Rubin characterization of Wishart distributions on symmetric cones. *Ann. Statist.* **24** 763–786.

CONSONNI, G. and VERONESE, P. (2001). Conditionally reducible natural exponential families and enriched conjugate priors. *Scand. J. Statist.* **28** 377–406.

CONSONNI, G., VERONESE, P. and GUTIÉRREZ-PEÑA, E. (2000). Reference priors for exponential families with simple quadratic variance function. *J. Multivariate Anal.* To appear. Available at economia.unipv.it/˜gconsonni/papers/RefPrior11.pdf.

COX, D. R. and REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49** 1–39.

DANIELS, M. J. and KASS, R. E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.* **94** 1254–1263.

DATTA, G. S. (1996). On priors providing frequentist validity of Bayesian inference for multiple parametric functions. *Biometrika* **83** 287–298.

DATTA, G. S. and GHOSH, M. (1995). Some remarks on noninformative priors. *J. Amer. Statist. Assoc.* **90** 1357–1363.

DATTA, G. S. and GHOSH, M. (1996). On the invariance of noninformative priors. *Ann. Statist.* **24** 141–159.

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **35** 189–233.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281.

FARAUT, J. and KORÁNYI, A. (1994). *Analysis on Symmetric Cones.* Clarendon, Oxford.

GEIGER, D. and HECKERMAN, D. (1994). Learning Gaussian networks. In *Proc. Tenth Conference on Uncertainty in Artificial Intelligence* 235–243. Morgan Kaufmann, San Francisco.

GUTIÉRREZ-PEÑA, E. and RUEDA, R. (2003). Reference priors for exponential families. *J. Statist. Plann. Inference.* **110** 35–54.

GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1995). Conjugate parameterizations for natural exponential families. *J. Amer. Statist. Assoc.* **90** 1347–1356.

GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1997). Exponential and Bayesian conjugate families: Review and extensions (with discussion). *Test* **6** 1–90.

JENSEN, S. T. (1988). Covariance hypotheses which are linear in both the covariance and the inverse covariance. *Ann. Statist.* **16** 302–322.

KASS, R. E. and WASSERMAN, L. (1996). The selection of prior distributions by formal rules (with discussion). *J. Amer. Statist. Assoc.* **91** 1343–1370.

LETAC, G. (1991). The classification of the natural exponential families by their variance functions. *Proc. 48th Session of the International Statistical Institute* **54** Book 3 1–37.

LETAC, G. (1997). Comment on "Exponential and Bayesian conjugate families: Review and extension," by E. Gutiérrez-Peña and A. F. M. Smith. *Test* **6** 79–81.

MASSAM, H. (1994). An exact decomposition theorem and a unified view of some related distributions for a class of exponential transformation models on symmetric cones. *Ann. Statist.* **22** 369–394.

MASSAM, H. and NEHER, E. (1997). On transformations and determinants of Wishart variables on symmetric cones. *J. Theoret. Probab.* **10** 867–902.

MASSAM, H. and NEHER, E. (1998). Estimation and testing for lattice conditional independence models on Euclidean Jordan algebras. *Ann. Statist.* **26** 1051–1082.

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80.

ROVERATO, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika* **87** 99–112.

SHACHTER, R. and KENLEY, C. (1989). Gaussian influence diagrams. *Management Science* **35** 527–550.

WERMUTH, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Amer. Statist. Assoc.* **75** 963–972.

YANG, R. (1995). Invariance of the reference prior under reparameterization. *Test* **4** 83–94.

YANG, R. and BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *Ann. Statist.* **22** 1195–1211.

DIPARTIMENTO DI ECONOMIA POLITICA
E METODI QUANTITATIVI
UNIVERSITY OF PAVIA
VIA S. FELICE, 5
27100 PAVIA
ITALY
E-MAIL: guido.consonni@unipv.it

ISTITUTO DI METODI QUANTITATIVI
L. BOCCONI UNIVERSITY
VIALE ISONZO, 25
20135 MILANO
ITALY
E-MAIL: piero.veronese@uni-bocconi.it