The Annals of Statistics
2011, Vol. 39, No. 6, 3417–3440
DOI: 10.1214/11-AOS951

MINIMAX ESTIMATION FOR MIXTURES OF WISHART DISTRIBUTIONS

BY L. R. HAFF, P. T. KIM[1], J.-Y. KOO[2] AND D. ST. P. RICHARDS[3]

University of California at San Diego, University of Guelph, Korea University and Penn State University

The space of positive definite symmetric matrices has been studied extensively as a means of understanding dependence in multivariate data along with the accompanying problems in statistical inference. Many books and papers have been written on this subject, and more recently there has been considerable interest in high-dimensional random matrices with particular emphasis on the distribution of certain eigenvalues. With the availability of modern data acquisition capabilities, smoothing or nonparametric techniques are required that go beyond those applicable only to data arising in Euclidean spaces. Accordingly, we present a Fourier method of minimax Wishart mixture density estimation on the space of positive definite symmetric matrices.

1. Introduction. The space of positive definite symmetric matrices has been studied extensively in statistics as a means of understanding dependence in multivariate data along with the accompanying problems in statistical inference. Many books and papers, for example, [7–9, 17, 19, 22] and [23], have been written on this subject, and there has been considerable interest recently in high-dimensional random matrices with particular emphasis on the distribution of certain eigenvalues [11] and on graphical models [15].

In this paper we consider the problem of estimating the mixing density of a continuous mixture of Wishart distributions. We construct a nonparametric estimator of that density and obtain minimax rates of convergence for the estimator. Throughout this work, we adopt, as a guide, results developed for the classical problem of deconvolution density estimation on Euclidean spaces; see, for example, [2, 4, 5, 14, 18] and [26]. Much of the difficulty with the space of positive definite symmetric matrices is due to the fact that mathematical analysis on the space is technically demanding. Helgason [10] and Terras [25] provide much insight and technical innovation, however, and we make extensive use of these methods.

Received August 2009; revised July 2011.

[1]Supported in part by the Natural Sciences and Engineering Research Council of Canada, Grant DG 46204.

[2]Supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0027601).

[3]Supported in part by NSF Grant DMS-07-05210.

MSC2010 subject classifications. Primary 62G20; secondary 65R32.

Key words and phrases. Deconvolution, Harish–Chandra c-function, Helgason–Fourier transform, Laplace–Beltrami operator, optimal rate, Sobolev ellipsoid, stochastic volatility.

We summarize the paper as follows. In Section 2 we discuss and set up the notation for Wishart mixtures. In Section 3 we begin by reviewing the necessary Fourier methods which allow us to construct a nonparametric estimator of the mixing density, and then we provide the estimator. The minimax property of our nonparametric estimator is stated in Section 4 along with supporting results. Section 5 presents simulation results as well as an application to finance examining real financial data. Finally, Sections 6 and 7 present the proofs.

2. Wishart mixtures. Throughout the paper, for any square matrix y, we denote the trace and determinant of y by $\operatorname{tr} y$ and $|y|$, respectively; further, we denote by $\mathbf{I}_m$ the $m \times m$ identity matrix. We will denote by $\mathcal{P}_m$ the space of $m \times m$ positive definite symmetric matrices.

For $s = (s_1, \ldots, s_m) \in \mathbb{C}^m$ with $\operatorname{Re}(s_j + \cdots + s_m) > (j-1)/2$, $j = 1, \ldots, m$, the multivariate gamma function is defined as

$$\Gamma_m(s_1, \ldots, s_m) = \pi^{m(m-1)/4} \prod_{j=1}^{m} \Gamma\left(s_j + \cdots + s_m - \frac{1}{2}(j-1)\right), \tag{2.1}$$

where $\Gamma(\cdot)$ denotes the classical gamma function.

We denote by G the general linear group $\mathrm{GL}(m, \mathbb{R})$ of all $m \times m$ nonsingular real matrices, by K the group $\mathrm{O}(m)$ of $m \times m$ orthogonal matrices and by A the group of diagonal positive definite matrices. The group G acts transitively on $\mathcal{P}_m$ by the action

$$G \times \mathcal{P}_m \to \mathcal{P}_m, \qquad (g, y) \mapsto g' y g, \tag{2.2}$$

$g \in G$, $y \in \mathcal{P}_m$, where g' denotes the transpose of g. Under this group action, the isotropy group of the identity in G is K; hence the homogeneous space $K \setminus G$ can be identified with $\mathcal{P}_m$ by the natural mapping from $K \setminus G \to \mathcal{P}_m$ that sends $Kg \mapsto g'g$. In distinguishing between left and right cosets, we place the quotient operation on the left and right of the group, respectively.

For $y = (y_{ij}) \in \mathcal{P}_m$, define the measure

$$\mathrm{d}_* y = |y|^{-(m+1)/2} \prod_{1 \le i \le j \le m} dy_{ij}.$$

It is well known that the measure $\mathrm{d}_* y$ is invariant under action (2.2). Relative to the dominating measure $\mathrm{d}_* y$, the probability density function of the standard Wishart distribution with N degrees of freedom is

$$w(y) = \frac{1}{2^{Nm/2} \Gamma_m(0, \ldots, 0, N/2)} |y|^{N/2} \exp\left(-\frac{1}{2} \operatorname{tr} y\right), \tag{2.3}$$

$y \in \mathcal{P}_m$. Consequently, for $\sigma \in \mathcal{P}_m$, we note that $\operatorname{tr}(\sigma^{-1/2} y \sigma^{-1/2}) = \operatorname{tr}(\sigma^{-1} y)$ and $|\sigma^{-1/2} y \sigma^{-1/2}| = |\sigma^{-1} y|$. It then follows that, relative to the dominating measure $\mathrm{d}_* y$, the density of the general Wishart distribution, with covariance parameter σ, is $w(\sigma^{-1} y)$, $y \in \mathcal{P}_m$.

Suppose next that σ is a random matrix and, relative to the dominating measure $\mathrm{d}_*\sigma$, has a continuous mixing density, f, that is invariant under the action (2.2). By integration with respect to σ, the continuous Wishart mixture density is given by

$$r(y) = \int_{\mathcal{P}_m} f(\sigma) w(\sigma^{-1} y)\, \mathrm{d}_*\sigma, \tag{2.4}$$

$y \in \mathcal{P}_m$. For the case in which $m = 1$, the standard Wishart density is essentially a chi-square density, in which case (2.4) is a continuous mixture of chi-square densities.

In general, (2.4) is a convolution operation for functions on $\mathcal{P}_m$. We denote by $x^{1/2}$ any matrix with $x^{t/2} x^{1/2} = x$, where $x^{t/2} = (x^{1/2})'$ and denote $x^{-t/2} = (x^{-1/2})^t$. Define $X \circ Z$, the convolution of two random matrices X and Z which are distributed on $\mathcal{P}_m$ by

$$X \circ Z = X^{t/2} Z X^{1/2}$$

and $f_1 * f_2$, the convolution of $f_1 \in L^1(\mathcal{P}_m)$ and $f_2 \in L^1(\mathcal{P}_m)$ by

$$(f_1 * f_2)(y) = \int_{\mathcal{P}_m} f_1(x) f_2(x^{-t/2} y x^{-1/2})\, \mathrm{d}_* x \qquad \text{for } y \in \mathcal{P}_m,$$

where $L^q(\mathcal{P}_m)$ is the space of integrable functions raised to the qth power on $\mathcal{P}_m$ for $q \geq 1$. If X and Z with densities f and w, respectively, are independent, then $Y = X \circ Z$ has the density $r = f * w$ since $w(\sigma^{-t/2} y \sigma^{-1/2}) = w(\sigma^{-1} y)$. Finally, (2.4) can be transformed into a scalar multiplication; see Section 3.2.

3. Fourier analysis on $\mathcal{P}_m$ and estimation of the mixing density. In this section we review the Fourier methods needed to transform the convolution product (2.4) and to construct a nonparametric estimator of the mixing density f.

3.1. *The Helgason–Fourier transform.* For $y \in \mathcal{P}_m$, denote by $|y_j|$ the principal minor of order j, $j = 1, \dots, m$. For $s \in \mathbb{C}^m$, the *power function* $p_s : \mathcal{P}_m \to \mathbb{C}$ is

$$p_s(y) = \prod_{j=1}^{m} |y_j|^{s_j}, \tag{3.1}$$

$y \in \mathcal{P}_m$. Let $\mathrm{d}_* k$ denote the Haar measure on K, normalized to have total volume equal to one; then

$$h_s(y) = \int_K p_s(k' y k)\, \mathrm{d}_* k, \tag{3.2}$$

$y \in \mathcal{P}_m$, is the *zonal spherical function* on $\mathcal{P}_m$. It is well known that the functions h_s are fundamental to harmonic analysis on symmetric spaces [10, 25]. If

$s_1, \ldots, s_m$ are nonnegative integers then, up to a constant factor, (3.2) is an integral formula for the zonal polynomials which arise in many aspects of multivariate statistical analysis [19], pages 231 and 232.

Let $C_c^\infty(\mathcal{P}_m)$ denote the space of infinitely differentiable, compactly supported, complex-valued functions f on $\mathcal{P}_m$; also, let

$$C_c^\infty(\mathcal{P}_m/K) = \{f \in C_c^\infty(\mathcal{P}_m) : f(k'yk) = f(y) \text{ for all } k \in K, y \in \mathcal{P}_m\}.$$

For $s \in \mathbb{C}^m$ and $k \in K$, the *Helgason–Fourier transform* ([25], page 87) of a function $f \in C_c^\infty(\mathcal{P}_m)$ is

$$\mathcal{H}f(s,k) = \int_{\mathcal{P}_m} f(y)\overline{p_s(k'yk)}\, \mathrm{d}_* y, \tag{3.3}$$

where $\overline{p_s(k'yk)}$ denotes complex conjugation.

For the case in which $f \in C_c^\infty(\mathcal{P}_m/K)$, we make the change of variables $y \mapsto k_1' y k_1$ in (3.3), $k_1 \in K$, and integrate with respect to the Haar measure $\mathrm{d}_* k_1$. Applying the invariance of f and formula (3.2), we deduce that $\mathcal{H}f(s,k)$ does not depend on k. Specifically, $\mathcal{H}f(s,k) = \hat{f}(s)$ where

$$\hat{f}(s) = \int_{\mathcal{P}_m} f(y)\overline{h_s(y)}\, \mathrm{d}_* y, \tag{3.4}$$

$s \in \mathbb{C}^m$, is the *zonal spherical transform* of f.

In the case of the standard Wishart density (2.3), which is a K-invariant function, the zonal spherical transform is well known (Muirhead [19], page 248; Terras [25], pages 85 and 86):

$$\hat{w}(s) = \frac{\Gamma_m(s_{m-1}, \ldots, s_1, -(s_1 + \cdots + s_m) + N/2)}{\Gamma_m(0, \ldots, 0, N/2)} h_s\left(\frac{1}{2}\mathbf{I}_m\right).$$

3.2. *The convolution property of the Helgason–Fourier transform.* The following result shows that the convolution operation can be transformed into a scalar multiplication.

PROPOSITION 3.1. *Suppose X and Z with densities f_X and f_Z, respectively, are independent, and Z is K-invariant. Let f_Y be the density of $Y = X \circ Z$. Then*

$$\mathcal{H}f_Y(s,k) = \mathcal{H}f_X(s,k)\hat{f}_Z(s) \qquad \textit{for } s \in \mathbb{C}^m \textit{ and } k \in K.$$

PROOF. Note

$$\mathcal{H}f_Y(s,k) = \mathbb{E}p_{\bar{s}}(k'Yk) = \mathbb{E}p_{\bar{s}}(k'X^{t/2}ZX^{1/2}k).$$

Using the KAN-Iwasawa decomposition of $X^{t/2}k$ (Terras [25], page 20), we have $X^{1/2}k = HU$ for $H \in K$ and U, an upper triangular matrix. Observe

$$\begin{aligned}\mathbb{E}p_{\bar{s}}(k'X^{t/2}ZX^{1/2}k) &= \mathbb{E}_X\{p_{\bar{s}}(U'U)\mathbb{E}_{Z|X}p_{\bar{s}}(H'ZH)\} \\ &= \hat{f}_Z(s)\mathbb{E}_X\{p_{\bar{s}}(U'U)\} \\ &= \mathcal{H}f_X(s,k)\hat{f}_Z(s),\end{aligned}$$

where Proposition 1 of Terras [25], page 39, is used for the first equality. □

3.3. *The inversion formula for the Helgason–Fourier transform.* For $a_1, a_2 \in \mathbb{C}$ with $\operatorname{Re}(a_1), \operatorname{Re}(a_2) > 0$, let

$$B(a_1, a_2) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1 + a_2)}$$

denote the classical beta function. For $s \in \mathbb{C}^m$ such that $\operatorname{Re}(s_i + \cdots + s_j) > -\frac{1}{2}(j - i + 1)$ for all $1 \le i < j \le m - 1$, the *Harish–Chandra c-function* is

$$c_m(s) = \prod_{1 \le i < j \le m-1} \frac{B(1/2, s_i + \cdots + s_j + (j - i + 1)/2)}{B(1/2, (j - i + 1)/2)}. \tag{3.5}$$

Let $\rho \equiv (\frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{4}(1 - m))$, and set

$$\omega_m = \frac{\prod_{j=1}^m \Gamma(j/2)}{(2\pi i)^m \pi^{m(m+1)/4} m!}, \tag{3.6}$$

$$\mathbb{C}^m(\rho) = \{s \in \mathbb{C}^m : \operatorname{Re}(s) = -\rho\} \tag{3.7}$$

and

$$\mathrm{d}_* s = \omega_m |c_m(s)|^{-2} \, \mathrm{d}_* s_1 \cdots \mathrm{d}_* s_m.$$

Let $M = \{\operatorname{diag}(\pm 1, \ldots, \pm 1)\}$ be the set of $m \times m$ diagonal matrices with entries ± 1 on the diagonal; then M is a subgroup of K and is of order 2^m. By factorizing the Haar measure $\mathrm{d}_* k$ on K, it may be shown ([25], page 88) that there exists an invariant measure $\mathrm{d}_* \bar{k}$ on the coset space K/M such that

$$\int_{\bar{k} \in K/M} \mathrm{d}_* \bar{k} = 1.$$

The *inversion formula for the Helgason–Fourier transform* $\mathcal{H}$ in (3.4) is that if $f \in C_c^\infty(\mathcal{P}_m)$, then [10, 25]

$$f(y) = \int_{\mathbb{C}^m(\rho)} \int_{\bar{k} \in K/M} \mathcal{H} f(s, k) p_s(k' y k) \, \mathrm{d}_* \bar{k} \, \mathrm{d}_* s, \tag{3.8}$$

$y \in \mathcal{P}_m$. In particular, if $f \in C_c^\infty(\mathcal{P}_m / K)$, then

$$f(y) = \int_{\mathbb{C}^m(\rho)} \hat{f}(s) h_s(y) \, \mathrm{d}_* s,$$

$y \in \mathcal{P}_m$, and there also holds the *Plancherel formula*,

$$\int_{\mathcal{P}_m} |f(y)|^2 \, \mathrm{d}_* y = \int_{\mathbb{C}^m(\rho)} \int_{K/M} |\mathcal{H} f(s, k)|^2 \, \mathrm{d}_* \bar{k} \, \mathrm{d}_* s. \tag{3.9}$$

We refer to Terras [25], page 87 ff., for full details of the inversion formula and for references to the literature.

3.4. *Eigenvalues, the Laplacian and Sobolev spaces.* For $y = (y_{ij}) \in \mathcal{P}_m$, we define the $m \times m$ matrix of partial derivatives,

$$\frac{\partial}{\partial y} = \left(\frac{1}{2}(1 + \delta_{ij})\frac{\partial}{\partial y_{ij}}\right),$$

where δ_{ij} denotes Kronecker's delta. The Laplacian, Δ, on $\mathcal{P}_m$ can be written ([25], page 106) in terms of the local coordinates y_{ij} as

$$\Delta = -\operatorname{tr}\left(\left(y \frac{\partial}{\partial y}\right)^2\right).$$

The power function p_s in (3.1) is an eigenfunction of Δ (see [19], page 229, [21], page 283, [25], page 49). Indeed, let $r_j = s_j + s_{j+1} + \cdots + s_m + \frac{1}{4}(m - 2j + 1)$, $j = 1, \dots, m$, and define

$$\lambda_s = -(r_1^2 + \cdots + r_m^2) + \tfrac{1}{48} m(m^2 - 1); \tag{3.10}$$

then $\Delta p_s(Y) = \lambda_s p_s(Y)$. Since $\operatorname{Re}(s) = -\rho$ then each r_j, $j = 1, \dots, m$, is purely imaginary; hence, $\lambda_s > 0$, $s \in C_m(\rho)$.

The operator $\mathcal{H}$ changes the effect of invariant differential operators on functions to pointwise multiplication: if $f \in C_c^\infty(\mathcal{P}_m)$, then

$$\mathcal{H}(\Delta f)(s, k) = \lambda_s \mathcal{H} f(s, k),$$

$s \in \mathbb{C}^m$, $k \in K$ ([25], page 88). For $\varphi > 0$, we therefore define the fractional power, $\Delta^{\varphi/2}$, of Δ, as the operator such that

$$\mathcal{H}(\Delta^{\varphi/2} f)(s, k) = \lambda_s^{\varphi/2} \mathcal{H} f(s, k),$$

$f \in C_c^\infty(\mathcal{P}_m)$. Having constructed $\Delta^{\varphi/2}$, we define the *Sobolev class*,

$$\mathcal{F}_\varphi = \{f \in C^\infty(\mathcal{P}_m) : \|\Delta^{\varphi/2} f\|^2 < \infty\},$$

where for $f \in C^\infty(\mathcal{P}_m)$,

$$\|f\| = \left(\int_{\mathcal{P}_m} |f(y)|^2 \,\mathrm{d}_* y\right)^{1/2}$$

denotes the $L^2(\mathcal{P}_m)$-norm with respect to the measure $\mathrm{d}_* y$. For $Q > 0$, we also define the *bounded Sobolev class*,

$$\mathcal{F}_\varphi(Q) = \{f \in C^\infty(\mathcal{P}_m) : \|\Delta^{\varphi/2} f\|^2 < Q\}.$$

4. Main result. In this section we will present the main result. We do so by applying the Helgason–Fourier transform to the mixture density (2.4) so that

$$\mathcal{H}r(s,k) = \mathcal{H}f(s,k)\hat{w}(s), \tag{4.1}$$

$s \in \mathbb{C}^m$, $k \in K$; see Proposition 3.1. Having observed a random sample $Y_1, \dots, Y_n$ from the mixture density, r, in (2.4), we estimate $\mathcal{H}r(s,k)$ by its *empirical Helgason–Fourier transform*,

$$\mathcal{H}_n r(s,k) = \frac{1}{n}\sum_{\ell=1}^{n} \overline{p_s(k'Y_\ell k)}. \tag{4.2}$$

On substituting (4.2) in (4.1), together with the assumption that $\hat{w}(s) \neq 0$, $s \in \mathbb{C}^m$, we obtain

$$\mathcal{H}_n f(s,k) = \frac{\mathcal{H}_n r(s,k)}{\hat{w}(s)},$$

$s \in \mathbb{C}^m$, $k \in K$.

Analogous with classical Euclidean deconvolution, we introduce a smoothing parameter $T = T(n)$ where $T(n) \to \infty$ as $n \to \infty$, and then we apply the inversion formula (3.8) using a spectral cut-off based on the eigenvalues of Δ. First, we introduce the notation

$$\mathbb{C}^m(\rho, T) = \{s \in \mathbb{C}^m(\rho) : \lambda_s < T\},$$

where $\mathbb{C}^m(\rho)$ is defined in (3.7). We now define

$$f_n(y) = \int_{\mathbb{C}^m(\rho,T)} \int_{\bar{k}\in K/M} \frac{\mathcal{H}_n r(s,\bar{k})}{\hat{w}(s)} p_s(\bar{k}' y \bar{k})\, \mathrm{d}_*\bar{k}\, \mathrm{d}_* s, \tag{4.3}$$

$y \in \mathcal{P}_m$, and take this as our nonparametric estimator of f.

We now state the minimax result for the estimator (4.3). Let C denote a generic positive constant independent of n. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we use the notations $a_n \ll b_n$ and $a_n \gg b_n$ to mean $a_n < Cb_n$ and $a_n > Cb_n$, respectively, as $n \to \infty$. Moreover, $a_n \asymp b_n$ means that $a_n \ll b_n$ and $a_n \gg b_n$.

THEOREM 4.1. *Suppose f is a density on $\mathcal{P}_m$ and $N > (m-1)/2$. Then, for the Wishart mixture* (2.4),

$$\sup_{f\in\mathcal{F}_\varphi(Q)} \mathbb{E}\|f_n - f\|^2 \ll (\log n)^{-2\varphi} \tag{4.4}$$

and for any estimator g_n of f,

$$\inf_{g_n} \sup_{f\in\mathcal{F}_\varphi(Q)} \mathbb{E}\|g_n - f\|^2 \gg (\log n)^{-2\varphi}. \tag{4.5}$$

We now provide some comments about this result. In the situation where (2.4) is a finite sum, so that

$$r(y) = \sum_{\ell=1}^{q} f_\ell w(\sigma_\ell^{-1} y) \quad \text{and} \quad \sum_{\ell=1}^{q} f_\ell = 1,$$

we have the finite mixture model. Methods for recovering the mixing coefficients can be covered by the techniques employed in [3]. We note that the continuous mixture model is a generalization of this approach.

It is noted that the condition f is a density and seems to be mild. The upper bound property of (4.4) is established in [13], Theorem 3.3, with $\beta = 1/2$. In the latter, the moment condition

$$\int_{\mathcal{P}_m} |y_1|^{-1} \cdots |y_{m-1}|^{-1} |y|^{(m-1)/2} r(y)\, \mathrm{d}_* y < \infty, \tag{4.6}$$

on the principal minors $|y_1|, \ldots, |y_m|$ of $y \in \mathcal{P}_m$ is assumed. In our theorem, we did not impose this moment condition as condition (4.6) is automatically satisfied. This is pointed out and commented upon below in the proof.

To derive the lower bound for estimating f in the $L^2(\mathcal{P}_m)$-norm, we shall follow the standard Euclidean approach. Thus we choose a pair of functions $f^0, f^n \in \mathcal{F}_\varphi(Q)$, and, with w denoting the Wishart density (2.3), we shall show that, for some constants $C_1, C_2 > 0$,

$$\|f^n - f^0\|^2 \ge C_1 (\log n)^{-2\varphi}$$

and

$$\chi^2(f^0 * w, f^n * w) \le \frac{C_2}{n}, \tag{4.7}$$

where

$$\chi^2(g_1, g_2) = \int_{\mathcal{P}_m} \frac{(g_1(y) - g_2(y))^2}{g_1(y)}\, \mathrm{d}_* y.$$

Precisely, let us suppose we can choose $f^0 \in \mathcal{F}_\varphi(Q)$ and a perturbation $\psi \in \mathcal{F}_\varphi(Q)$, and, for $\delta = \delta_n > 0$, let ψ^δ be a scaling of ψ such that $\|\psi^\delta\| \asymp \delta^{-1/2} \|\psi\|$. Define

$$f^n = f^0 + C_\psi \delta^{-\varphi + 1/2} \psi^\delta.$$

If δ can be chosen so that

$$\chi^2(f^0 * w, f^n * w) \le C n^{-1},$$

then the lower bound rate of convergence is determined by $\delta^{-2\varphi}$. We shall develop such a construction and, moreover, do so in a way such that $\delta \asymp \log n$ as $n \to \infty$.

REMARK 4.2. The profound influence of Charles Stein on covariance estimation originates largely from his Rietz lecture; see [22]. The idea is that for certain loss functions over $\mathcal{P}_m$, the usual estimator of the covariance matrix parameter is inadmissible. Through an unbiased estimation of the risk function over covariance matrices, Stein was able to improve upon the usual estimator by pooling the observed eigenvalues of the sample covariance matrix. Subsequent to this, through a series of papers, improvements were obtained in Haff [7–9]. Other related works include Takemura [24], Lin and Perlman [16] and Loh [17], to name a few.

In this paper, we contribute to the case in which one observes data from a continuous mixture of Wishart distributions, not merely a sample from a single distribution. Therefore, the parameter of interest would be the mixing density of the covariance parameters. And the nonparametric estimator of the mixing density (4.3) is an attractive candidate because of its minimax property. Based on this procedure, one could consider the moment, or mode, of f_n, as a possible estimator of the corresponding population parameters. Alternatively, one could take a nonparametric empirical Bayes approach as in Pensky [20].

5. Numerical aspects and an application to finance. This section presents numerical aspects for the $m = 2$ case with an application to finance.

5.1. *Computation of estimators.* Suppose X and Z are independent with Z having a Wishart distribution. Let $Y = X^{t/2} Z X^{1/2}$ where $X^{1/2}$ is upper triangular. For visualization, we display estimators of the marginal density for D where $X = H' D H$ with $H \in K$ and $D \in A$. Let

$$\hat{r}_n(s) = \frac{1}{n} \sum_{j=1}^{n} \overline{h_s(E_j)} \quad \text{and} \quad \hat{f}_n(s) = \frac{\hat{r}_n(s)}{\hat{w}(s)},$$

where $E_j \in A_+$ denotes the diagonal matrix of eigenvalues of Y_j, $j = 1, \dots, n$. Denote by f^D the density of eigenvalues of X. Then, the estimator for f^D is given by

$$f_n^D(a) = \int_{\mathbb{C}^2(\rho,T)} \operatorname{Re}\{\hat{f}_n(s) h_s(a)\} \, \mathrm{d}_* s \qquad \text{for } a = \operatorname{diag}(a_1, a_2) \in A. \tag{5.1}$$

Consider the computation of $h_s(a)$ when

$$s = -\rho + ib = (-1/2 + ib_1, 1/4 + ib_2)$$

so that $\operatorname{Re}(s) = -\rho$. From pages 90 and 91 of [25], the spherical function is given by

$$h_s(a) = (a_1 a_2)^{i(b_2 + b_1/2)} P_{-1/2 + ib_1}\bigl(\cosh\bigl(\log\bigl(\sqrt{a_1/a_2}\bigr)\bigr)\bigr),$$

where Legendre function $P_{-1/2+it}(x)$ can be computed using `conicalP_0(t, x)` in the `gsl` library in `R`.

The mean integrated squared error (MISE) of f_n^D is defined by

$$\mathrm{MISE}(f_n) = \mathbb{E}\int_A \bigl(f_n^D(a) - f^D(a)\bigr)^2 \, \mathrm{d}_*a.$$

It is reasonable to choose T which minimizes $\mathrm{MISE}(f_n^D)$ or equivalently

$$M(T) = \mathbb{E}\int_A (f_n^D(a))^2 \, \mathrm{d}_*a - 2\mathbb{E}\int_A f_n^D(a) f^D(x) \, \mathrm{d}_*a.$$

One can find an unbiased estimator $M_0(T)$ of $M(T)$; see [12]. We choose $\hat{T}$ by

$$\arg\min_T M_0(T).$$

Monte Carlo approximation is used for integration of (5.1) and $M_0(T)$.

5.2. *Simulation.* Denote by $W_N(\sigma)$ the Wishart density with degrees of freedom N and covariance matrix σ. We generate data as follows. For $j = 1, \ldots, n$:

- generate $Z_j \sim W_{20}(I_2)$;
- generate $X_j \sim f$;
- do a Cholesky decomposition of $X_j = (X_j)^{t/2}(X_j)^{1/2}$, and calculate $Y_j = (X_j)^{t/2} Z_j (X_j)^{1/2}$.

As examples, we consider a unimodal mixing density $W_{15}(2I_2)$ and a bimodal density $0.5W_{15}(2I_2) + 0.5W_{15}(6I_2)$. Figure 1 show the results for the unimodal case whereas Figure 2 show the results for the bimodal case. In each of these plots the domain consists of the two eigenvalues starting with the largest. One can see that the general shapes of the estimators become closer to that of the true density as n increases.

5.3. *Application to stochastic volatility.* Stochastic volatility using the Wishart distribution is of much interest in finance; see, for example, [1] and [6]. In particular, this entails a situation precisely of the form (2.4). Let us apply this to the situation where we are interested in estimating the mixing density.

Although our methods can be applied to a portfolio of many assets, let us restrict ourselves to two assets since this would be the smallest multivariate example. Indeed, let S_j^1 and S_j^2 denote the daily closing stock prices of Samsung Electronics (005930.KS) and LG-display (034220.KS), respectively, traded on the Korea Stock Exchange (KSC) for 2010, where the data can be easily accessed on public financial websites. We will assume as usual that $Q_j^k = \log(S_{j+1}^k / S_j^k)$ follows a bi-variate normal distribution for $k = 1, 2$. We transform the daily data to weekly data and compute the weekly 2×2 covariance matrix Y_i for $i = 1, \ldots, 52$. In case a week has a holiday, we repeat the last previous observation. Under the usual assumptions this would constitute observations from a mixture model (2.4) with a standard Wishart distribution with four degrees of freedom.

Figure 3 plots the mixing density estimator corresponding to the two eigenvalues. One can see that there are two peaks, suggesting a possible bimodal stochastic volatility mixing density in the eigenvalues.

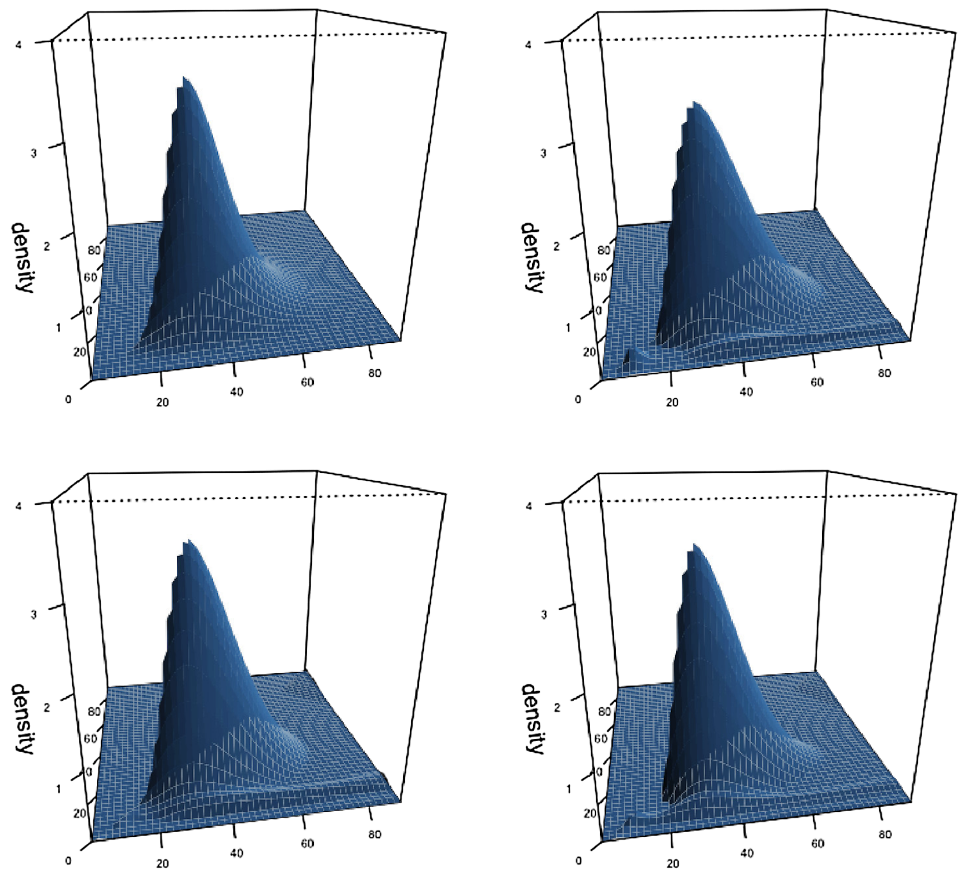

FIG. 1. *Unimodal case*: *upper left displays the true density of* $W_{15}(2I_2)$, *upper right shows an estimate with* $n = 500$, *lower left with* $n = 1{,}000$ *and lower right with* $n = 2{,}000$.

6. Proof of upper bound. The strategy here is, first, to decompose the integrated mean-squared error into its variance and bias components,

$$\begin{aligned}\mathbb{E}\|f_n - f\|^2 &= \mathbb{E}\|(f_n - \mathbb{E}f_n) + (\mathbb{E}f_n - f)\|^2 \\ &= \mathbb{E}\|f_n - \mathbb{E}f_n\|^2 + \|\mathbb{E}f_n - f\|^2,\end{aligned} \tag{6.1}$$

and, last, to estimate each component separately using estimates based on the Plancherel formula and the inversion formula for the Helgason–Fourier transform.

6.1. *The integrated bias.*

LEMMA 6.1. *Suppose that* $f \in \mathcal{F}_\varphi(Q)$ *and* $\varphi > \dim \mathcal{P}_m/2$. *Then*

$$\|\mathbb{E}f_n - f\|^2 \ll T^{-\varphi}.$$

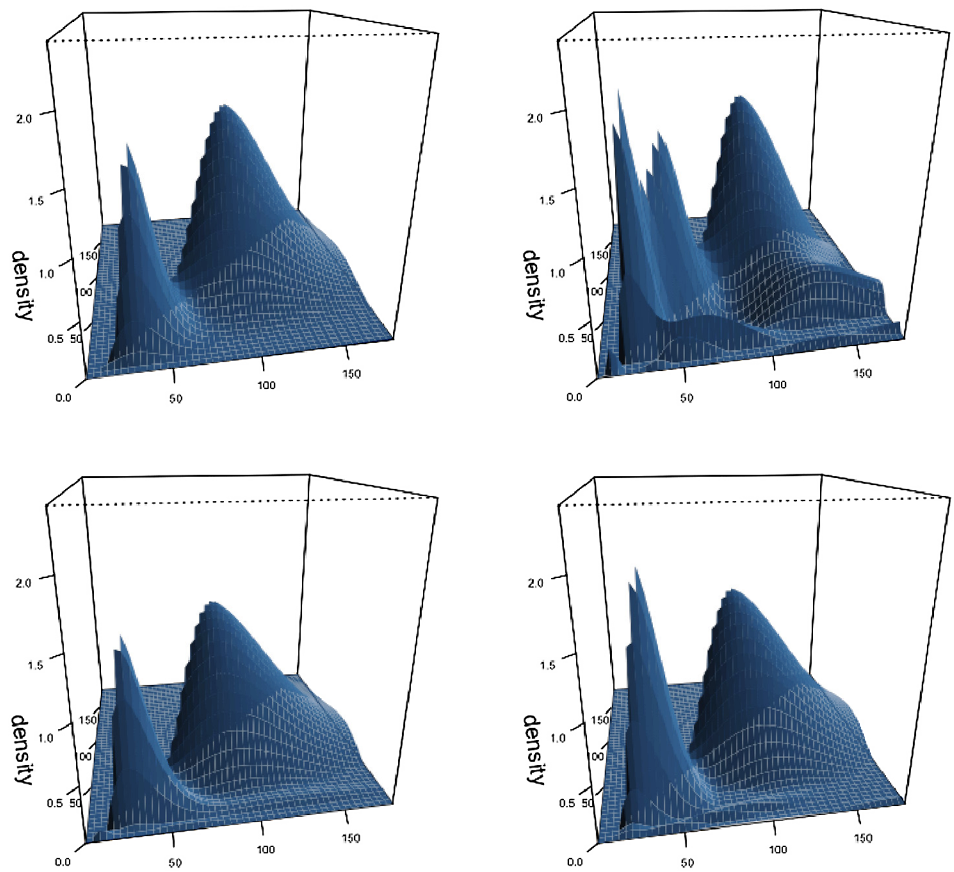

FIG. 2. *Bimodal case*: *upper left displays the true density* $0.5W_{15}(2I_2) + 0.5W_{15}(6I_2)$, *upper right shows an estimate with* $n = 500$, *lower left with* $n = 1{,}000$ *and lower right with* $n = 2{,}000$.

PROOF. We have for $x \in \mathcal{P}_m$

$$\begin{aligned} \mathbb{E} f_n(x) - f(x) &= \int_{\mathbb{C}^m(\rho,T)} \int_{\bar{k}\in K/M} \mathcal{H}f(s,\bar{k}) p_s(\bar{k}' x \bar{k})\, \mathrm{d}_*\bar{k}\, \mathrm{d}_* s \\ &\quad - \int_{\mathbb{C}^m(\rho)} \int_{\bar{k}\in K/M} \mathcal{H}f(s,\bar{k}) p_s(\bar{k}' x \bar{k})\, \mathrm{d}_*\bar{k}\, \mathrm{d}_* s \\ &= -\int_{\lambda_s > T, \mathrm{Re}(s) = -\rho} \int_{\bar{k}\in K/M} \mathcal{H}f(s,\bar{k}) p_s(\bar{k}' x \bar{k})\, \mathrm{d}_*\bar{k}\, \mathrm{d}_* s. \end{aligned} \tag{6.2}$$

Applying the Plancherel formula, we obtain

$$\|\mathbb{E} f_n - f\|^2 = \int_{\lambda_s > T, \mathrm{Re}(s) = -\rho} \int_{K/M} |\mathcal{H}f(s,\bar{k})|^2\, \mathrm{d}_*\bar{k}\, \mathrm{d}_* s.$$

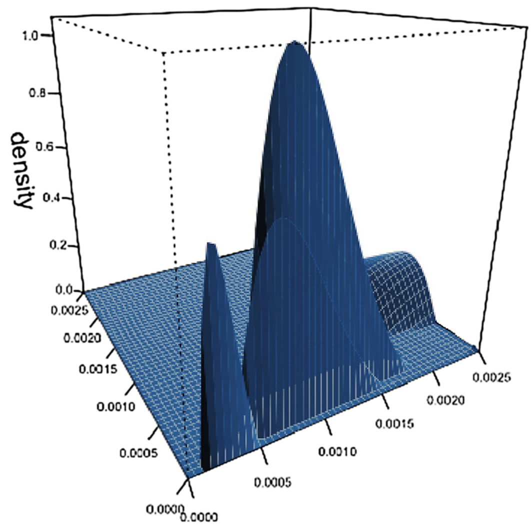

FIG. 3. *A density estimator for the weekly covariance matrix of stock prices.*

Consequently,

$$
\begin{aligned}
\|\mathbb{E}f_n - f\|^2 &= \int_{\lambda_s \geq T, \mathrm{Re}(s)=-\rho} \int_{K/M} |\mathcal{H}f(s,\bar{k})|^2 \,\mathrm{d}_*\bar{k}\,\mathrm{d}_*s \\
&\leq T^{-\varphi} \int_{\lambda_s \geq T, \mathrm{Re}(s)=-\rho} \int_{K/M} \lambda_s^{\varphi} |\mathcal{H}f(s,\bar{k})|^2 \,\mathrm{d}_*\bar{k}\,\mathrm{d}_*s \qquad (6.3)\\
&\leq T^{-\varphi} \int_{\mathbb{C}^m(\rho)} \int_{K/M} \lambda_s^{\varphi} |\mathcal{H}f(s,\bar{k})|^2 \,\mathrm{d}_*\bar{k}\,\mathrm{d}_*s,
\end{aligned}
$$

where we use the fact that

$$
\lambda_s^{\varphi} |\mathcal{H}f(s,k)|^2 \equiv |\lambda_s^{\varphi/2}\mathcal{H}f(s,k)|^2 = |\mathcal{H}(\Delta^{\varphi/2}f)(s,k)|^2.
$$

Therefore

$$
\begin{aligned}
\|\mathbb{E}f_n - f\|^2 &\leq T^{-\varphi} \int_{\mathbb{C}^m(\rho)} \int_{\bar{k}\in K/M} |\mathcal{H}(\Delta^{\varphi/2}f)(s,k)|^2 \,\mathrm{d}_*\bar{k}\,\mathrm{d}_*s \\
&= T^{-\varphi} \int_{\mathcal{P}_m} |\Delta^{\varphi/2}f(w)|^2 \,\mathrm{d}_*w,
\end{aligned}
$$

where the equality follows from the Plancherel formula. By assumption, $f \in \mathcal{F}_{\varphi}(Q)$, the latter integral is bounded above by Q, so we obtain

$$
\|\mathbb{E}f_n - f\|^2 \leq QT^{-\varphi},
$$

and the proof is complete. □

6.2. *The integrated variance.* To obtain bounds for the integrated variance, several preliminary calculations are needed. In particular, we begin with the variance calculation of the empirical Helgason–Fourier transform, which has similarities to the usual empirical characteristic function.

LEMMA 6.2. *For $s \in \mathbb{C}^m(\rho)$ and $k \in K/M$,*

$$\mathbb{E}|\mathcal{H}_n r(s,k) - \mathbb{E}\mathcal{H}_n r(s,k)|^2 = \frac{1}{n}\big(|\mathcal{H}r(-2\rho,k)|^2 - |\mathcal{H}r(s,k)|^2\big).$$

PROOF. By (4.2),

$$\begin{aligned} |\mathcal{H}_n r(s,k)|^2 &= \overline{\mathcal{H}_n r(s,k)}\mathcal{H}_n r(s,k) \\ &= \frac{1}{n^2}\sum_{j,\ell=1}^{n} \overline{p_s(k'Y_jk)}\,p_s(kY_\ell k) \qquad (6.4) \\ &= \frac{1}{n^2}\left\{\sum_{j=1}^{n}|p_s(k'Y_jk)|^2 + \sum_{j\neq\ell}\overline{p_s(k'Y_jk)}\,p_s(k'Y_\ell k)\right\}. \end{aligned}$$

Observe also that

$$\begin{aligned} p_s(w)\overline{p_s(w)} &= |w_1|^{s_1}\cdots|w_m|^{s_m}\overline{|w_1|^{s_1}\cdots|w_m|^{s_m}} \\ &= |w_1|^{2\,\mathrm{Re}(s_1)}\cdots|w_m|^{2\,\mathrm{Re}(s_m)} \\ &= p_{-2\rho}(w) \end{aligned}$$

since $\mathrm{Re}(s) = -\rho$. Applying this result to (6.4) and taking expectations, we obtain

$$\begin{aligned} \mathbb{E}|\mathcal{H}_n r(s,k)|^2 &= \frac{1}{n^2}\mathbb{E}\left\{\sum_{j=1}^{n}|p_s(k'Y_jk)|^2 + \sum_{j\neq\ell}\overline{p_s(k'Y_jk)}\,p_s(k'Y_\ell k)\right\} \\ &= \frac{1}{n^2}\left\{\sum_{j=1}^{n}\mathbb{E}p_{-2\rho}(k'Y_jk) + \sum_{j\neq\ell}\overline{\mathbb{E}p_s(k'Y_jk)}\,\mathbb{E}p_s(k'Y_\ell k)\right\} \\ &= \frac{1}{n}\mathcal{H}r(-2\rho,k) + \frac{n-1}{n}|\mathcal{H}r(s,k)|^2, \end{aligned}$$

where the last equality follows from the fact that $Y_1,\ldots,Y_n$ are independent and identically distributed as Y, and because $\mathbb{E}p_s(k'Yk) = \mathcal{H}r(s,k)$. □

Following Terras [25], pages 34 and 35, let

$$A^+ = \{a = \mathrm{diag}(a_1,\ldots,a_m) \in A : a_1 > \cdots > a_m\}$$

denote the positive Weyl chamber in A. For $a = \mathrm{diag}(a_1,\ldots,a_m) \in A$, let $\mathrm{d}a = \prod_{j=1}^m a_j^{-1}da_j$ and set

$$\gamma(a) = \prod_{j=1}^{m} a_j^{-(m-1)/2}\prod_{1\le i<j\le m}|a_i - a_j|,$$

and define the normalizing constant b_m by $b_m^{-1} = \pi^{-(m^2+m)/4} \prod_{j=1}^m j\Gamma(j/2)$. Denote $d_* a = b_m \gamma(a)\, da$.

LEMMA 6.3. *As* $T \to \infty$,

$$\mathbb{E}\|f_n - \mathbb{E}f_n\|^2 \ll \sup_{s\in\mathbb{C}^m(\rho)} |\hat{w}_Z(s)|^{-2} \frac{T^{\dim \mathcal{P}_m/2}}{n}.$$

PROOF. By the Plancherel formula,

$$\begin{aligned}
&\mathbb{E}\|f_n - \mathbb{E}f_n\|^2 \\
&\quad = \int_{\mathbb{C}^m(\rho,T)} \int_{K/M} \mathbb{E}|\mathcal{H}_n f(s,k) - \mathbb{E}\mathcal{H}_n f(s,k)|^2 \, d\bar{k}\, d_* s \\
&\quad = \int_{\mathbb{C}^m(\rho,T)} \int_{K/M} \mathbb{E}|\mathcal{H}_n r(s,k) - \mathbb{E}\mathcal{H}_n r(s,k)|^2 \, d\bar{k} |\hat{w}_Z(s)|^{-2} \, d_* s \\
&\quad \le \frac{1}{n} \sup_{\lambda_s<T, \mathrm{Re}(s)=-\rho} |\hat{w}_Z(s)|^{-2} \int_{K/M} |\mathcal{H}r(-2\rho,\bar{k})|^2 \, d\bar{k} \int_{\mathbb{C}^m(\rho,T)} d_* s \\
&\quad \ll \sup_{\lambda_s<T, \mathrm{Re}(s)=-\rho} |\hat{w}_Z(s)|^{-2} \frac{T^{\dim \mathcal{P}_m/2}}{n}
\end{aligned}$$

as $T \to \infty$.

Choose $a \in A^+$. Observe that

$$p_{-2\rho}(a) = a_1^{-(m-1)+(m-1)/2} \cdots a_{m-1}^{-1+(m-1)/2} a_m^{(m-1)/2} \le 1. \tag{6.5}$$

Since $p_{-2\rho}(k'ak)$ is a continuous function of k on a compact set K, $p_{-2\rho}(k'ak)$ is uniformly bounded on K such $|p_{-2\rho}(k'ak)| \le C$ on K. Since f is a density so that r is also a density, we have

$$\begin{aligned}
|\mathcal{H}r(-2\rho, \mathbf{I}_m)| &= \left| \int_{\mathcal{P}_m} r(y) \overline{p_{-2\rho}(y)} \, d_* y \right| \\
&= \int_{A^+} \int_K r(k'ak) \overline{p_{-2\rho}(k'ak)} \, d_* a\, d_* k \\
&\le \int_{A^+} \int_K r(k'ak) |\overline{p_{-2\rho}(k'ak)}| \, d_* a\, d_* k \\
&\le C \int_{A^+} \int_K r(k'ak) \, d_* a\, d_* k \\
&= C \int_{\mathcal{P}_m} r(y) \, d_* y \\
&= C.
\end{aligned}$$

Hence, it follows from continuity and the compactness of K/M

$$\int_{K/M} |\mathcal{H}r(-2\rho, \bar{k})|^2 \, \mathrm{d}\bar{k} < \infty,$$

which has been used in the above calculation.

In addition, we use the fact that, as $T \to \infty$,

$$\sup_{\mathbb{C}^m(\rho,T)} |c_m(s)|^{-2} \ll T^{m(m-1)/4},$$

a result which follows from Proposition 7.2 of Helgason [10], page 450. □

The proof of the upper bound can now be obtained by applying Lemmas 6.1 and 6.3 to (6.1) and setting $T \asymp (\log n)^2$.

7. Proof of lower bound. We need to provide some detailed calculations, and the essence of the proof is contained for the case $m = 2$; hence we will keep this assumption for the remainder of this paper. The generalization to $m > 2$ may be obtained by using higher order hyperbolic spherical coordinates. In this section, we assume that ψ is a K-invariant function defined on $\mathcal{P}_2$.

7.1. *Convolution and Helagson–Fourier transform in polar coordinate.* For $y \in \mathcal{P}_2$, let $y = k'ak$ with $a = \operatorname{diag}(a_1, a_2) \in A^+$, $k \in K$ so that $\psi(y) = \psi(a)$. Let

$$a = D_{u_1} e^{u_2}$$

with $D_z = \operatorname{diag}(e^z, e^{-z})$ for $z \in \mathbb{R}$, and write

$$\psi(u) = \psi(D_{u_1} e^{u_2}).$$

By a change of variables,

$$\int_{\mathcal{P}_2} \psi(y) \, \mathrm{d}_* y = \int_{\mathcal{D}} \psi(u) \, \mathrm{d}_* u, \tag{7.1}$$

where

$$\mathcal{D} = \{u : u_1 \in \mathbb{R}^+ \text{ and } u_2 \in \mathbb{R}\}$$

and

$$\mathrm{d}_* u = 4\pi \sinh u_1 \, \mathrm{d}u_1 \, \mathrm{d}u_2.$$

Denote $k_\theta = \left[\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right]$. The next lemma is straightforward so we shall omit the proof.

LEMMA 7.1. *For $u_1 \in \mathbb{R}^+$, $v_1 \in \mathbb{R}^+$ and $\theta \in [0, 2\pi]$, the matrix equation*

$$k_\xi D_R k'_\xi = D_{-u_1/2} k_\theta D_{v_1} k'_\theta D_{-u_1/2}$$

has a solution $R^* = R^*(u_1, v_1, \theta)$ and $\xi^* = \xi^*(u_1, v_1, \theta)$. Further, $\cosh R^*$ *has minimum and maximum values* $\cosh(u_1 - v_1)$ *and* $\cosh(u_1 + v_1)$, *respectively, and R^* and ξ^* can be defined uniquely.*

In general, if both f and g are K-invariant functions on $\mathcal{P}_m$, then $f * g$ is also K-invariant and $f * g = g * f$. Hence, $\psi * w$ is K-invariant and $\psi * w = w * \psi$ due to K-invariance of ψ and w. From this and Lemma 7.1, we can define

$$\Psi_{u_1 v_1}(z) = \frac{1}{2\pi} \int_0^{2\pi} \psi(R^*(\theta, u_1, v_1), z)\,\mathrm{d}\theta$$

for $z \in \mathbb{R}$, $u_1 \in \mathbb{R}^+$, $v_1 \in \mathbb{R}^+$. Denoting by W the distribution function corresponding to the standard Wishart density $w(u)$ with respect to the measure $\mathrm{d}_* u$, we have that for $v \in D$,

$$(\psi * w)(v) = \int_{\mathcal{D}} \Psi_{u_1 v_1}(v_2 - u_2)\,\mathrm{d}W(u). \tag{7.2}$$

The Laplacian for K-invariant functions in polar coordinate is given by $\Delta = \Delta_{u_1} + \Delta_{u_2}$, where

$$\Delta_{u_1} = -\coth(u_1)\frac{\partial}{\partial u_1} - \frac{\partial^2}{\partial u_1^2}, \qquad \Delta_{u_2} = -\frac{\partial^2}{\partial u_2^2},$$

and the spherical function is given by

$$h_s(u) = P_{s_1}(\cosh u_1)e^{(s_1+2s_2)u_2} \qquad \text{for } u \in \mathcal{D} \tag{7.3}$$

with P_s, the Legendre function; see Terras [25]. It can be seen that

$$\Delta_{u_1} h_s = -\tfrac{1}{2}s(s+1)h_s, \qquad \Delta_{u_2} h_s = -\tfrac{1}{2}(s_1 + 2s_2)^2 h_s,$$

so that $\Delta h_s = \lambda_s h_s$ with $\lambda_s = -\frac{1}{2}\{(s_1 + 2s_2)^2 + s(s+1)\}$.

The Helgason–Fourier transform of ψ is given by

$$\hat{\psi}(s) = \int_{\mathcal{D}} \psi(u) P_{\bar{s}_1}(\cosh u_1) e^{(\overline{s_1+2s_2})u_2}\,\mathrm{d}_* u \tag{7.4}$$

from (7.1) and (7.3). Suppose ψ is separable so that $\psi(u) = \psi_1(u_1)\psi_2(u_2)$. Then, (7.4) implies

$$\hat{\psi}(s) = \mathcal{M}\psi_1(\bar{s}_1)\mathcal{L}\psi_2(s_1 + 2s_2), \tag{7.5}$$

where $\mathcal{L}$ and $\mathcal{M}$ denote, respectively, the Laplace transform and the Mehler–Fock transform; see Terras [25].

7.2. *χ^2-divergence*. Choose ψ_1 as the perturbation function as in Fan [5]. Then, one can construct $\psi(u)$ satisfying the following conditions:

(P1) ψ is K-invariant and separable with $\psi(u) = \psi_1(u_1)\psi_2(u_2)$ for $u \in D$.
(P2) $\mathcal{L}\psi_2(it) = 0$ for $t \notin [1, 2]$.
(P3) $\psi \in \mathcal{F}_\varphi(Q)$.
(P4) $\psi_1(u_1) = O(\cosh^{-m_0} u_1)$ and $\psi_2(u_2) = O(e^{-m_0|u_2|})$, where $0 < \xi < 1$ and $m_0\xi > (N-1)(1-\xi)/2$.

(P5) $\int_{\mathcal{D}} \psi(u)\,\mathrm{d}_* u = 0$.

Let p_b be a density on $\mathbb{R}$ such that p_b is sufficiently smooth and satisfies $p_b(u_2) = c_b \exp(-b|u_2|)$, $|u_2| \geq c_0$, where c_b is a normalizing constant. Define the function

$$f^0(u) = C_b(\cosh u_1)^{-b} p_b(u_2),$$

where $C_b = c_b(b-1)/(2\pi)$ for $b > 1$.

For a function $g : \mathcal{P}_2 \to \mathbb{R}$ and $\delta > 0$, define

$$g^\delta(y) = g\bigl(|y|^{(\delta-1)/2} y\bigr) \qquad \text{for } y \in \mathcal{P}_2$$

so that

$$\psi(u) = \psi(u_1, \delta u_2) \qquad \text{for } u \in \mathcal{D}.$$

PROPOSITION 7.2. *Suppose* (P1)–(P5) *hold. For a pair of densities*

$$f^0 \quad \textit{and} \quad f^n = f^0 + C_\psi \delta^{-\varphi+1/2} \psi^\delta,$$

the χ^2*-divergence between* $g^0 = f^0 * w$ *and* $g^n = f^n * w$ *satisfies*

$$\chi^2(g^0, g^n) \leq C/n$$

provided that $b < \frac{1}{2}\min(3\pi, (N-1)(1-\xi) - 1)$.

The fact that $(\log n)^{-2\varphi}$ is a lower bound follows from Proposition 7.2 whose proof follows from a sequence of lemmas below.

7.3. *Perturbing function.* Denote $(q_1, q_2) = (s_1, s_1 + 2s_2)$ and $\beta_j = \operatorname{Im}(q_j)$ for $j = 1, 2$. Note that $q_2 = i\beta_2$ for $s \in C^2(\rho)$.

LEMMA 7.3. *Suppose* (P1) *holds. Then* $\|\psi^\delta\| = \delta^{-1}\|\psi\|$.

PROOF. By (7.5) and the change of variable $u_2 \mapsto \delta u_2$, we obtain

$$\hat{\psi}^\delta(s) = \mathcal{M}\psi_1(\bar{q}_1)\{\delta^{-1}\mathcal{L}\psi_2(\bar{q}_2/\delta)\}.$$

The desired result follows from the Plancherel formula (3.9) and change of variable $s \mapsto q$. □

LEMMA 7.4. *Suppose* (P1), (P2) *and* (P3) *hold. Then there exists a positive constant* C_ψ *such that* $C_\psi \delta^{-\varphi+1/2}\psi^\delta \in \mathcal{F}_\varphi(Q)$.

PROOF. Suppose $s \in C^2(\rho) \cap \mathcal{S}$. Observe that

$$\lambda_s = -\tfrac{1}{2}\{s_1(s_1+1) + (s_1+2s_2)^2\} = \tfrac{1}{2}\bigl(\beta_1^2 + \beta_2^2 + \tfrac{1}{4}\bigr)$$

and that for $\delta \geq 1$,

$$\tfrac{1}{2}(\beta_1^2 + (\delta\beta_2)^2 + \tfrac{1}{4}) \leq \delta^2 \lambda_s.$$

Now, the Plancherel formula (3.9) and change of variable $s \mapsto q$ gives

$$\|\Delta^{\varphi/2} \psi^\delta\|^2 \leq C\delta^{2\varphi-1} \|\Delta^{\varphi/2}\psi\|^2.$$

A suitable choice of C_ψ gives the desired result. □

LEMMA 7.5. *Under* (P1) *and* (P2), $\|\psi^\delta * w\|^2 \leq C\delta e^{-3\pi\delta}\|\psi\|^2$.

PROOF. For $s \in C^2(\rho) \cap \mathcal{S}$,

$$|\hat{w}(s)|^2 \leq Ce^{-\pi(\alpha_1+2\alpha_2)} = Ce^{-2\pi\beta_2} \leq Ce^{-3\pi}.$$

The desired result follows from the inequality $e^{-\pi(s_1+2s_2)\delta} \leq Ce^{-3\pi\delta}$, the Plancherel formula (3.9) and change of variable $s \mapsto q$. □

7.4. *Tail behavior.*

LEMMA 7.6. *For* $c_1 \geq 0$ *and* $c_2, c_3 \in \mathbb{R}$,

$$\int_{\{u_1 > c_1, c_2 < u_2 < c_3\}} \mathrm{d}W(u) = 4\pi \int_{\{c_2 < u_2 < c_3\}} \exp\{(N-1)u_2 - e^{u_2}\cosh c_1\}\,\mathrm{d}u_2.$$

PROOF. Change of variable gives the desired result. □

If U has the distribution function W, then $(U_1, \delta U_2)$ has the distribution function W_δ with density $\delta^{-1}w(u_1, u_2/\delta)$. By (7.2) and a change of variables,

$$(\psi^\delta * w)(v) = \int_{\mathcal{D}} \Psi_{u_1 v_1}(\delta v_2 - u_2)\,\mathrm{d}W_\delta(u). \tag{7.6}$$

LEMMA 7.7. *We have* $g^0 = f^0 * w$ *is* K*-invariant, and as* $v_1 \to \infty$ *and* $|v_2| \to \infty$

$$g^0(v) \geq Ce^{-b(v_1+|v_2|)}.$$

PROOF. Choose a constant $C_{1/2}$ such that

$$\int_{\{0 \leq u_1 < C_{1/2}, |u_2| < C_{1/2}\}} \mathrm{d}W(u) \geq 1/2$$

and $p_b(u_2) = c_b \exp(-b|u_2|)$ for $|u_2| \geq C_{1/2}$. The desired result follows from (7.2) and Lemma 7.1. □

LEMMA 7.8. *Define*

$$\eta_\delta(v) = \int_{\mathcal{D}} \cosh^{-m_0}(u_1 - v_1) e^{-m_0|v_2-u_2|}\, \mathrm{d}W_\delta(u).$$

Then, there exist M and a constant C such that when $v_1 \geq M$ and $|v_2|/\delta \geq M$,

$$\eta_\delta(v) \leq C \exp\{-\tfrac{1}{2}(N-1)(1-\xi)(v_1 + |v_2|/\delta)\}$$

for all $\delta \geq 1$ provided that $0 < \xi < 1$ and $m_0\xi > (N-1)(1-\xi)/2$.

PROOF. Denote

$$\begin{aligned}
J_{11} &= \{u : |u_1 - v_1| \leq \xi v_1, |v_2 - \delta u_2| \leq \xi |v_2|\},\\
J_{12} &= \{u : |u_1 - v_1| \leq \xi v_1, |v_2 - \delta u_2| > \xi |v_2|\},\\
J_{21} &= \{u : |u_1 - v_1| > \xi v_1, |v_2 - \delta u_2| \leq \xi |v_2|\},\\
J_{22} &= \{u : |u_1 - v_1| > \xi v_1, |v_2 - \delta u_2| > \xi |v_2|\}
\end{aligned}$$

and for $i, j = 1, 2$,

$$I_{ij} = \int_{J_{ij}} \cosh^{-m_0}(u_1 - v_1) e^{-m_0|v_2-u_2|}\, \mathrm{d}W_\delta(u).$$

Consider I_{11}. Suppose $v_2 < 0$. If $|u_1 - v_1| \leq \xi v_1$ and $|v_2 - \delta u_2| \leq \xi |v_2|$, then

$$(1-\xi)(v_1 + |v_2|/\delta) \leq u_1 - u_2 \leq (1+\xi)(v_1 + |v_2|/\delta).$$

Let $\zeta = (1-\xi)(v_1 + |v_2|/\delta)$ and $\lambda = e^\zeta$. Note that

$$I_{11} \leq C\biggl(\int_{\{u_1 \geq 0, u_2 < -\zeta\}} \mathrm{d}W(u) + \int_{\{u_1 \geq u_2 + \zeta, -\zeta \leq u_2 < 0\}} \mathrm{d}W(u)\biggr) := I_1^- + I_2^-.$$

Since $e^{u_2}\cosh(u_2 + \zeta) = \frac{1}{2}(\lambda e^{2u_2} + \lambda^{-1})$, then we have

$$I_1^- = 4\pi \int_{\{0 \leq t < 1/\lambda\}} t^{N-2} e^{-t}\, \mathrm{d}t = O(\lambda^{-(N-1)})$$

and by Lemma 7.6 with $c_1 = u_2 + \zeta$, $c_2 = -\zeta$ and $c_3 = 0$

$$I_2^- = C \int_{\{-\zeta \leq u_2 < 0\}} e^{(N-1)u_2} \exp(-e^{u_2}\cosh(u_2 + \zeta))\, \mathrm{d}u_2 = O(\lambda^{-(N-1)/2}).$$

Hence, if $v_2 < 0$, then $I_{11} = O(\lambda^{-(N-1)/2})$ as $v_1 + |v_2|/\delta \to \infty$. Suppose $v_2 \geq 0$. If $|v_2 - \delta u_2| \leq \xi |v_2|$ and $|u_1 - v_1| \leq \xi v_1$, then

$$(1-\xi)(v_1 + |v_2|/\delta) \leq u_1 + u_2 \leq (1+\xi)(v_1 + |v_2|/\delta).$$

Observe that

$$I_{11} \le C\biggl(\int_{\{u_1\ge 0, u_2\ge\zeta\}} \mathrm{d}W(u) + \int_{\{u_1\ge\zeta-u_2, 0\le u_2<\zeta\}} \mathrm{d}W(u)\biggr) := I_1^+ + I_2^+.$$

Since $e^{u_2/2}\cosh(\zeta-u_2) = \frac{1}{2}(\lambda+\lambda^{-1}e^{2u_2})$, then we have

$$I_1^+ = 4\pi\int_{\{t\ge\lambda\}} t^{N-2}e^{-t}\,\mathrm{d}t = O(e^{-\lambda})$$

and by Lemma 7.6, with $c_1 = u_2-\zeta$, $c_2 = 0$ and $c_3 = \zeta$,

$$I_2^+ = C\int_{\{0\le u_2<\zeta\}} e^{(N-1)u_2}\exp\bigl(-e^{u_2}\cosh(\zeta-u_2)\bigr)\,\mathrm{d}u_2 = O(e^{-\lambda}).$$

Hence, for $v_2\in\mathbb{R}$,

$$I_{11} = O\bigl(\lambda^{-(N-1)/2}\bigr) = O\bigl(\exp\bigl\{-\tfrac{1}{2}(N-1)(1-\xi)(v_1+|v_2|/\delta)\bigr\}\bigr)$$

as $v_1+|v_2|/\delta\to\infty$.

Consider I_{21}. Let $\zeta = (1-\xi)|v_2|/\delta$ and $\lambda = e^{\zeta}$. Suppose $v_2<0$. On J_{21}, $u_2\le -(1-\xi)|v_2|/\delta = -\zeta$ on J_{21}. Note that $\cosh(\xi v_1)^{-m_0} \le (\frac{1}{2}e^{\xi v_1})^{-m_0} \le 2^{-m_0}e^{-(1/2)(N-1)(1-\xi)v_1}$. If follows from these and Lemma 7.6 that

$$I_{21} = O\bigl(\exp\bigl\{-\tfrac{1}{2}(N-1)(1-\xi)(v_1+|v_2|/\delta)\bigr\}\bigr),$$

and if $v_2\ge 0$,

$$I_{21} = O\bigl(\exp\bigl\{-\tfrac{1}{2}(N-1)(1-\xi)(v_1+|v_2|/\delta)\bigr\}\bigr).$$

Observe that $m_0\xi|v_2| \ge m_0\xi|v_2|/\delta \ge \frac{1}{2}(N-1)(1-\xi)|v_2|/\delta$. It follows from this and Lemma 7.6 that

$$\begin{aligned}I_{12} &\le Ce^{-m_0\xi|v_2|}\int_{u_2\in\mathbb{R}}\int_{u_1>(1-\xi)v_1}\mathrm{d}W(u)\\ &= O\biggl(\exp\biggl\{-\frac{1}{2}(N-1)(1-\xi)(v_1+|v_2|/\delta)\biggr\}\biggr)\end{aligned}$$

and

$$\begin{aligned}I_{22} &\le Ce^{-m_0\xi|v_2|}(\cosh(\xi v_1))^{-m_0}\int_{\mathcal{D}}\mathrm{d}W(u)\\ &= O\biggl(\exp\biggl\{-\frac{1}{2}(N-1)(1-\xi)(v_1+|v_2|/\delta)\biggr\}\biggr).\end{aligned}$$

This completes the proof of Lemma 7.8. □

7.5. *Proof of Proposition* 7.2. From (P5), one can choose C_ψ sufficiently small such that f^n is a density. By (7.1) and (7.6),

$$\begin{aligned}\chi^2(g^0, g^n) &= C_\psi^2 \delta^{-2\varphi+1} \int_{\mathcal{D}} \frac{\{(\psi^\delta * w)(v)\}^2}{g^0(v)} \,\mathrm{d}_* v \\ &= C_\psi^2 \delta^{-2\varphi} \int_{\mathcal{D}} \frac{\{\int_{\mathcal{D}} \Psi_{u_1 v_1}(v_2 - u_2)\,\mathrm{d}W_\delta(u)\}^2}{g^0(v_1, v_2/\delta)} \,\mathrm{d}_* v.\end{aligned}$$

Let

$$J_{11} = \{v : 0 \le v_1 \le \delta, |v_2|/\delta \le \delta\}, \qquad J_{12} = \{v : 0 \le v_1 \le \delta, |v_2|/\delta > \delta\},$$
$$J_{21} = \{v : v_1 > \delta, |v_2|/\delta \le \delta\}, \qquad J_{22} = \{v : v_1 > \delta, |v_2|/\delta > \delta\},$$

and for $i, j = 1, 2$,

$$I_{ij} = \int_{J_{ij}} \frac{\{\int_{\mathcal{D}} \Psi_{u_1 v_1}(v_2 - u_2)\,\mathrm{d}W_\delta(u)\}^2}{g^0(v_1, v_2/\delta)} \,\mathrm{d}_* v.$$

By (7.6), Lemmas 7.5, 7.7 and the Plancherel formula (3.9), we obtain

$$\begin{aligned}I_{11} &\le C e^{2b\delta} \int_{\mathcal{D}} \left\{ \int_{\mathcal{D}} \Psi_{u_1 v_1}(v_2 - u_2)\,\mathrm{d}W_\delta(u) \right\}^2 \mathrm{d}_* v \\ &= C e^{2b\delta} \delta \|\psi^\delta * w\|^2 \\ &\le C\delta^2 e^{(2b - 3\pi)\delta}.\end{aligned}$$

Lemma 7.1 and (P4) imply

$$\int_{\mathcal{D}} \Psi_{u_1 v_1}(v_2 - u_2)\,\mathrm{d}W_\delta(u) \le C\eta_\delta(v). \tag{7.7}$$

Let $c_1 = (N - 1)(1 - \xi) - b$. It follows from (7.7), Lemmas 7.7 and 7.8 that

$$I_{12} = O\big(\delta e^{-(c_1 - b)\delta}\big), \qquad I_{21} = O\big(e^{-(c_1 - b - 1)\delta}\big), \qquad I_{22} = O\big(\delta e^{-(2c_1 - 1)\delta}\big).$$

Now letting $\varepsilon = \min(3\pi - 2b, (N - 1)(1 - \xi) - 2b - 1) > 0$ and combining the above bounds, we obtain

$$\chi^2(g^0, g^n) \le C_\psi^2 \delta^{-2\varphi}(I_{11} + I_{12} + I_{21} + I_{22}) \le C\delta^{-2\varphi+2} e^{-\varepsilon\delta}.$$

Choosing $\delta = \varepsilon/\log n$, we have the desired result.

REFERENCES

[1] ASAI, M., MCALEER, M. and YU, J. (2006). Multivariate stochastic volatility: A review. *Econometric Rev.* **25** 145–175. MR2256285

[2] BUTUCEA, C. and TSYBAKOV, A. B. (2008). Sharp optimality in density deconvolution with dominating bias, I, II. *Theory Probab. Appl.* **52** 24–39, 237–249.

[3] CHEN, J. and TAN, X. (2009). Inference for multivariate normal mixtures. *J. Multivariate Anal.* **100** 1367–1383. MR2514135

[4] DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531. MR1224414

[5] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

[6] GOURIEROUX, C. and SUFANA, R. (2010). Derivative pricing with Wishart multivariate stochastic volatility. *J. Bus. Econom. Statist.* **28** 438–451. MR2723611

[7] HAFF, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264–1276. MR0550149

[8] HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597. MR0568722

[9] HAFF, L. R. (1991). The variational form of certain Bayes estimators. *Ann. Statist.* **19** 1163–1190. MR1126320

[10] HELGASON, S. (1978). *Differential Geometry, Lie Groups, and Symmetric Spaces. Pure and Applied Mathematics* **80**. Academic Press, New York. MR0514561

[11] JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

[12] KIM, K. R., KIM, P. T., KOO, J. Y., OH, J. and ZHU, H. (2011). An analysis of diffusion tensor image data based on mixture of Wisharts. Preprint, Korea Univ.

[13] KIM, P. T. and RICHARDS, D. S. P. (2010). Deconvolution density estimation on spaces of positive definite symmetric matrices. In *Festschrift for T. P. Hettmansperger* (D. Hunter et al., eds.) 147–168. World Scientific, Singapore.

[14] KOO, J.-Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.* **21** 590–599. MR1232506

[15] LETAC, G. and MASSAM, H. (2007). Wishart distributions for decomposable graphs. *Ann. Statist.* **35** 1278–1323. MR2341706

[16] LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analysis VI* (*Pittsburgh*, *PA*, 1983) (P. R. Krishnaiah, ed.) 411–429. North-Holland, Amsterdam. MR0822310

[17] LOH, W.-L. (1991). Estimating covariance matrices. *Ann. Statist.* **19** 283–296. MR1091851

[18] MAIR, B. A. and RUYMGAART, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56** 1424–1444. MR1409127

[19] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

[20] PENSKY, M. (1999). Nonparametric empirical Bayes estimation of the matrix parameter of the Wishart distribution. *J. Multivariate Anal.* **69** 242–260. MR1703374

[21] RICHARDS, D. S. P. (1985). Applications of invariant differential operators to multivariate distribution theory. *SIAM J. Appl. Math.* **45** 280–288. MR0781108

[22] STEIN, C. (1975). Estimation of a covariance matrix. Rietz lecture, 1975 Annual Meeting of the IMS, Atlanta, GA.

[23] STEĬN, Č. (1977). Lectures on the theory of estimation of many parameters. In *Studies in the Statistical Theory of Estimation*, *Part I* (I. A. Ibragimov and M. S. Nikulin, eds.). *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*) **74** 4–65, 146, 148 (in Russian). Steklov Mathematical Institute, Moscow. [English transl.: *J. Soviet Math.* **34** (1986) 1373–1403.]

[24] TAKEMURA, A. (1984). An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population. *Tsukuba J. Math.* **8** 367–376. MR0767967

[25] TERRAS, A. (1988). *Harmonic Analysis on Symmetric Spaces and Applications. II*. Springer, Berlin. MR0955271

[26] ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist*. **18** 806–831. MR1056338

L. R. HAFF
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA AT SAN DIEGO
LA JOLLA, CALIFORNIA 92093
USA
E-MAIL: haff@ucsd.edu

P. T. KIM
DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF GUELPH
GUELPH, ONTARIO N1G 2W1
CANADA
E-MAIL: pkim@uoguelph.ca

J.-Y. KOO
DEPARTMENT OF STATISTICS
KOREA UNIVERSITY
ANAM-DONG SUNGBUK-KU
SEOUL, 136-701
KOREA
E-MAIL: jykoo@korea.ac.kr

D. ST. P. RICHARDS
DEPARTMENT OF STATISTICS
PENN STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802
USA
E-MAIL: richards@stat.psu.edu