Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ORIGINAL ARTICLE

On the Expectations of Equivariant Matrix-valued Functions of Wishart and Inverse Wishart Matrices

Grant Hillier, PhD[1] | **Raymond M. Kan, PhD**[2]

[1]Economics Department, University of Southampton, Southampton, United Kingdom

[2]Rotman School of Management, University of Toronto, Ontario, Canada

Correspondence
Raymond Kan, Rotman School of Management, University of Toronto, 105 St. George Street, Ontario, Canada, M5S 3E6.
Email: Raymond.Kan@rotman.utoronto.ca

Abstract

Many matrix-valued functions of an $m \times m$ Wishart matrix W, $F_k(W)$, say, are homogeneous of degree k in W, and are equivariant under the conjugate action of the orthogonal group $\mathcal{O}(m)$, i.e., $F_k(HWH^{\mathrm{T}}) = HF_k(W)H^{\mathrm{T}}$, $H \in \mathcal{O}(m)$. It is easy to see that the expectation of such a function is itself homogeneous of degree k in Σ, the covariance matrix, and are also equivariant under the action of $\mathcal{O}(m)$ on Σ. The space of such homogeneous, equivariant, matrix-valued functions is spanned by elements of the type $W^r p_\lambda(W)$, where $r \in \{0, \ldots, k\}$ and, for each r, λ varies over the partitions of $k-r$, and $p_\lambda(W)$ denotes the power-sum symmetric function indexed by λ. In the analogous case where W is replaced by W^{-1}, these elements are replaced by $W^{-r} p_\lambda(W^{-1})$. In this paper we derive recurrence relations and analytical expressions for the expectations of such functions. Our results provide highly efficient methods for the computation of all such moments.

KEYWORDS
equivariant function, inverse Wishart distribution, recurrence relation, Wishart distribution

1 | INTRODUCTION

We are concerned throughout most this paper with functions of a real symmetric matrix W having a Wishart distribution on $n \geq m$ degrees of freedom and $m \times m$ positive definite symmetric covariance matrix Σ. We denote this setup by $W \sim \mathcal{W}_m(n, \Sigma)$. Many of the results can be extended to different versions of the Wishart distribution – for example, the complex Wishart case – the latter being treated briefly in Section 7 here.

Our interest is in the properties of the distribution, such as the expectations $E[W^k]$, i.e., the moments of W, or expectations of functions like $E[f(W)W^r]$, where $f(W)$ is a scalar function of W. Examples of the second type arise in evaluating the properties of estimators for Σ (see Haff (1979), for instance). The first of these expectations is evidently a matrix-valued function of Σ, and possibly also of n and m, say $\Psi_k(\Sigma)$. The second is also a matrix-valued function of Σ, say $\Psi_{r,f}(\Sigma)$. Similar expressions that may be of interest involve the inverse Wishart matrix W^{-1}, rather than W itself.

In the case of the moments, it is not difficult to confirm that $\Psi_k(\Sigma)$ is symmetric, homogeneous of degree k in Σ, and equivariant under the conjugate action of the orthogonal group $\mathcal{O}(m)$. That is, for all $H \in \mathcal{O}(m)$, $\Psi_k(H\Sigma H^{\mathrm{T}}) = H\Psi_k(\Sigma)H^{\mathrm{T}}$ (see Hillier & Kan (2021)). In problems of the second type the scalar function $f(W)$ may itself be invariant under the action of $\mathcal{O}(m)$, i.e., $f(HWH^{\mathrm{T}}) = f(W)$ for all $H \in \mathcal{O}(m)$. For instance, $f(W)$ may be a polynomial in $\mathrm{tr}(W)$. In that case, $\Psi_{r,f}(\Sigma)$ will again be equivariant under $\mathcal{O}(m)$, and homogeneous of degree $r + \mathrm{degree}(f)$. It is functions of this type that are the focus in this paper.

Thus, in this paper we will be concerned with the space of matrix-valued functions on $\mathcal{P}(m)$, the set of $m \times m$ symmetric matrices, that are homogeneous of degree k and equivariant under the action of $\mathcal{O}(m)$. Specifically, we describe, and provide methods for evaluating, the expectations of the generators of this space, thus providing an efficient method for dealing with arbitrary matrix functions in this class.

The rest of the paper is organized as follows. In Section 2 we introduce the notation, the matrix-valued functions of interest, and preliminary results that will be needed later. In Section 3 we review existing approaches for computing some special cases of the terms of interest. Section 4 presents two new recurrence relations for the objects of interest, one for functions of W itself, the other for the analogous functions of W^{-1}. Based on these new recurrence relations, Sections 5 and 6 develop efficient methods for obtaining explicit expressions for terms of both types. Section 7 discusses the analogues of the earlier results for the complex Wishart case. Section 8 concludes.

2 | NOTATION AND PRELIMINARIES

We first consider the set of functions of W, each homogeneous of degree k in W:

$$\{W^r p_\lambda(W);\ r = 0, 1, \ldots, k,\ \lambda \vdash k - r\}. \tag{1}$$

Here, $\lambda \vdash k$ means that $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ is a partition of an integer k into non-negative integers satisfying $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$ and $|\lambda| = \sum_{i=1}^{k} \lambda_i = k$. The number of nonzero parts of λ is denoted by $\ell(\lambda)$. The power-sum symmetric function $p_\lambda(W)$ of W corresponding to a partition λ is defined as

$$p_\lambda(W) = \prod_{i=1}^{\ell(\lambda)} p_{\lambda_i}(W), \tag{2}$$

where $p_i(W) = \mathrm{tr}(W^i)$. Exactly analogous constructions apply with W replaced by W^{-1}. Thus, we shall be interested in evaluating the following expectations:

$$\gamma_{r,\lambda}(\Sigma) = E[W^r p_\lambda(W)], \tag{3}$$

and

$$\tilde{\gamma}_{r,\lambda}(\Sigma^{-1}) = E[W^{-r} p_\lambda(W^{-1})], \tag{4}$$

for all $r = 0, 1, \ldots, k$ and all partitions $\lambda \vdash k - r$. In fact, we shall omit the term $r = 0$ in both cases, since these are proportional to the identity matrix I_m. The expectations $E[p_\lambda(W)]$ and $E[p_\lambda(W^{-1})]$ for $\lambda \vdash k$ can be evaluated separately. The power-sum symmetric functions that appear in these expressions can be replaced by any other homogeneous symmetric functions invariant under the conjugate action of $\mathcal{O}(m)$. For example, elementary symmetric functions, zonal polynomials, etc. The results depend on the choice, of course, and the power-sums do seem the most convenient at the moment. For later use we note the following well-known results: $E[W] = n\Sigma$, and, for $n - m - 1 > 0$, $E[W^{-1}] = \Sigma^{-1}/\tilde{n}$, where $\tilde{n} = n - m - 1$.

Remark 1. For the case of the $\gamma_{r,\lambda}(\Sigma)$, it is easy to see that the expectations exist for all k. In the case of the $\tilde{\gamma}_{r,\lambda}(\Sigma^{-1})$, though, the expectation for given k may not exist unless n is sufficiently large. A necessary and sufficient condition for the existence of all such expectations in this case is that $n - m + 1 > 2k$. To avoid distracting from the main message we will assume throughout that this condition is satisfied.

We assemble the scalars $p_\lambda(W)$ when $\lambda \vdash k - r$ into a vector $p^{(k-r)}(W)$, with λ arranged in reverse lexicographical order. The length of $p^{(i)}(W)$ is $\pi(i)$, where $\pi(i)$ denotes the number of partitions of i (by convention, $\pi(0) = 1$). We will also denote by $d_k = \sum_{i=0}^{k-1} \pi(i)$ the aggregate number of partitions of the integers below k. We then assemble all the terms for powers $r = 1, \ldots, k$ of W, in descending order, into a matrix

$$\mathcal{L}_k(W) = \begin{bmatrix} W^k \\ p^{(1)}(W) \otimes W^{k-1} \\ p^{(2)}(W) \otimes W^{k-2} \\ \vdots \\ p^{(k-1)}(W) \otimes W \end{bmatrix} \tag{5}$$

of dimension $md_k \times m$. Here $\otimes$ denotes the Kronecker product. The identical construction is defined with W replaced by W^{-1}, and denoted by $\mathcal{L}_k(W^{-1})$. For example, for $k = 1, 2, 3$ we have

$$\mathcal{L}_1(W) = W, \qquad \mathcal{L}_2(W) = \begin{bmatrix} W^2 \\ \mathrm{tr}(W)W \end{bmatrix}, \qquad \mathcal{L}_3(W) = \begin{bmatrix} W^3 \\ \mathrm{tr}(W)W^2 \\ \mathrm{tr}(W^2)W \\ \mathrm{tr}(W)^2 W \end{bmatrix}. \tag{6}$$

As observed in Hillier & Kan (2021) for the case $E[W^k] = \Psi_k(\Sigma)$, the elements of $\mathcal{L}_k(W)$ span the space of matrices that are homogeneous of degree k and equivariant under the action of $\mathcal{O}(m)$ (see, for instance, Procesi (1976) (where the equivariants under the group action are called "concomitants"), or Letac & Massam (2004)).

Now, it is easy to check that both $\gamma_{r,\lambda}(\Sigma) = E[W^r p_\lambda(W)]$ and $\tilde{\gamma}_{r,\lambda}(\Sigma^{-1}) = E[W^{-r} p_\lambda(W^{-1})]$ are homogeneous of degree k in their arguments, and equivariant under the action of $\mathcal{O}(m)$. It follows that, for each pair (r, λ), $\gamma_{r,\lambda}(\Sigma)$ is a linear combination of the elements of $\mathcal{L}_k(\Sigma)$, i.e., we can write

$$E[W^r p_\lambda(W)] = \sum_{i=1}^{k} \left[\sum_{\rho \vdash k-i} c_{\lambda,\rho} p_\rho(\Sigma) \right] \Sigma^i, \tag{7}$$

$$E[W^{-r} p_\lambda(W^{-1})] = \sum_{i=1}^{k} \left[\sum_{\rho \vdash k-i} \tilde{c}_{\lambda,\rho} p_\rho(\Sigma^{-1}) \right] \Sigma^{-i}, \tag{8}$$

where $c_{\lambda,\rho}$ and $\tilde{c}_{\lambda,\rho}$ are constants that do not depend on Σ. Assembling all these expressions for the entire matrices $\mathcal{L}_k(W)$ and $\mathcal{L}_k(W^{-1})$, we have:

Proposition 1. *There are $d_k \times d_k$ matrices of constants $\mathcal{C}_k$ and $\tilde{\mathcal{C}}_k$ such that*

$$\mathcal{Q}_k \equiv E[\mathcal{L}_k(W)] = (\mathcal{C}_k \otimes I_m)\mathcal{L}_k(\Sigma), \tag{9}$$

$$\tilde{\mathcal{Q}}_k \equiv E[\mathcal{L}_k(W^{-1})] = (\tilde{\mathcal{C}}_k \otimes I_m)\mathcal{L}_k(\Sigma^{-1}). \tag{10}$$

Thus, the only unknowns in the expressions for the expectations of the matrices $\mathcal{L}_k(W)$ and $\mathcal{L}_k(W^{-1})$ are the (rather large) $d_k \times d_k$ matrices of constants $\mathcal{C}_k$ and $\tilde{\mathcal{C}}_k$. In Hillier & Kan (2021), we dealt with the top-order term in $\mathcal{L}_k(W)$, $E[W^k]$, and obtained the coefficients in the first row of $\mathcal{C}_k$ for $k \leq 10$ using an algorithm based on the properties of products of quadratic forms. Our task in this paper will be to analyse the structure and properties of the complete matrices $\mathcal{C}_k$ and $\tilde{\mathcal{C}}_k$, and to provide efficient methods for computing them. Our results are very much more direct, and vastly more efficient, than those used in the earlier paper. Before undertaking this we introduce a little more notation.

If the matrices $\mathcal{C}_k$ and $\tilde{\mathcal{C}}_k$ in Proposition 1 are partitioned as

$$\mathcal{C}_k = [\mathcal{C}_{k,0},\ \mathcal{C}_{k,1}, \ldots, \mathcal{C}_{k,k-1}], \tag{11}$$

$$\tilde{\mathcal{C}}_k = [\tilde{\mathcal{C}}_{k,0},\ \tilde{\mathcal{C}}_{k,1}, \ldots, \tilde{\mathcal{C}}_{k,k-1}], \tag{12}$$

with $\mathcal{C}_{k,i}$ and $\tilde{\mathcal{C}}_{k,i}$ of dimension $d_k \times \pi(i)$, we have alternative expressions for $\mathcal{Q}_k$ and $\tilde{\mathcal{Q}}_k$:

Proposition 2.

$$\mathcal{Q}_k = \sum_{i=1}^{k} c_{k,i} \otimes \Sigma^i, \tag{13}$$

$$\tilde{\mathcal{Q}}_k = \sum_{i=1}^{k} \tilde{c}_{k,i} \otimes \Sigma^{-i}, \tag{14}$$

where

$$c_{k,i} = \mathcal{C}_{k,k-i} p^{(k-i)}(\Sigma), \quad i = 1, \ldots, k, \tag{15}$$

$$\tilde{c}_{k,i} = \tilde{\mathcal{C}}_{k,k-i} p^{(k-i)}(\Sigma^{-1}), \quad i = 1, \ldots, k. \tag{16}$$

We give explicit expressions for the vectors $c_{k,i}$ and $\tilde{c}_{k,i}$, and the matrices $\mathcal{C}_{k,i}$ and $\tilde{\mathcal{C}}_{k,i}$ in Sections 5 and 6 below.

Taking the trace of each $m \times m$ sub-matrix in $\mathcal{L}_k(W)$ produces elements like $p_r(W)p_\lambda(W)$. These are not distinct. In fact, there are $\pi(k)$ distinct elements of this type, each corresponding to a partition κ of k, and, for $k > 2$, $\pi(k)$ is very much smaller than d_k. For example, $\pi(7) = 15$, while $d_7 = 30$, and $\pi(10) = 42$ while $d_{10} = 97$. The distinct elements of this type are the elements of the vector $p^{(k)}(W)$, the $\pi(k) \times 1$ vector of power-sum symmetric functions $p_\kappa(W)$, with partitions $\kappa \vdash k$ arranged in reverse lexicographical order:

$$p^{(k)}(W) = \begin{bmatrix} p_{(k)}(W) \\ p_{(k-1,1)}(W) \\ \vdots \\ p_{(1^k)}(W) \end{bmatrix}. \tag{17}$$

We define the vector q_k as the expectation of $p^{(k)}(W)$:

$$q_k = E[p^{(k)}(W)]. \tag{18}$$

Correspondingly, we define $\tilde{q}_k = E[p^{(k)}(W^{-1})]$. We note that, if $C^{(k)}(W)$ is the analogously defined vector of zonal polynomials $C_\kappa(W)$ indexed by partitions of $\kappa \vdash k$, it is well known that there is a transition matrix $\mathcal{D}_k$, say, satisfying

$$p^{(k)}(W) = \mathcal{D}_k C^{(k)}(W), \tag{19}$$

where $\mathcal{D}_k$ is a $\pi(k) \times \pi(k)$ matrix of constants. James (1961) presented a method for computing the coefficients of the matrix $\mathcal{D}_k^{-1}$. He also showed that the coefficients of $\mathcal{D}_k^{-1}$ satisfy an orthogonality condition which guarantees that $\mathcal{D}_k^{-1}$ is invertible. The same relation holds with W replaced by W^{-1}. Since the expectations of the elements of $C^{(k)}(W)$ and $C^{(k)}(W^{-1})$ are known (and quite simple), this relation provides one expression for q_k and $\tilde{q}_k$. These are discussed in the next section.

3 | REVIEW OF EXISTING RESULTS

As just mentioned, equation (19), and its analogue for W^{-1}, provide explicit expressions for the expectations $E[p^{(k)}(W)]$ and $E[p^{(k)}(W^{-1})]$ in terms of the transition matrix $\mathcal{D}_k$, because the expectations of the vectors $C^{(k)}(W)$ and $C^{(k)}(W^{-1})$ are known to be of the form

$$E[C^{(k)}(W)] = \mathcal{A}_k C^{(k)}(\Sigma), \tag{20}$$

and

$$E[C^{(k)}(W^{-1})] = \tilde{\mathcal{A}}_k C^{(k)}(\Sigma^{-1}), \tag{21}$$

where $\mathcal{A}_k$ and $\tilde{\mathcal{A}}_k$ are $\pi(k) \times \pi(k)$ diagonal matrices with diagonal elements $2^k(n/2)_\kappa$ and $1/[(-2)^k(-\tilde{n}/2)_\kappa]$, respectively, with the partitions $\kappa \vdash k$ arranged in reverse lexicographical order. Note that the first result is due to Constantine (1963) and the second result is due to Khatri (1966). Here,

$$(a)_\kappa = \prod_{i=1}^{\ell(\kappa)} \left(a - \frac{i-1}{2}\right)_{\kappa_i}, \tag{22}$$

where $(c)_r = c(c+1)\cdots(c+r-1)$ is the Pochhammer symbol, and $\tilde{n} = n - m - 1$. These results yield the simple expression (see Hayakawa & Kikuchi (1979))

$$q_k = E[p^{(k)}(W)] = \mathcal{H}_k p^{(k)}(\Sigma), \tag{23}$$

where $\mathcal{H}_k = \mathcal{D}_k \mathcal{A}_k \mathcal{D}_k^{-1}$, and (see Watamori (1990))

$$\tilde{q}_k = E[p^{(k)}(W^{-1})] = \tilde{\mathcal{H}}_k p^{(k)}(\Sigma^{-1}), \tag{24}$$

with $\tilde{\mathcal{H}}_k = \mathcal{D}_k \tilde{\mathcal{A}}_k \mathcal{D}_k^{-1}$. These results provide one set of expressions for the expectations $E[p_\kappa(W)]$ and $E[p_\kappa(W^{-1})]$, when $\kappa \vdash k$ (i.e., the elements of q_k and $\tilde{q}_k$). We will give alternative, simpler, expressions later.

While these two expressions are elegant, their usefulness is limited by the practical challenge of computing the transition matrix $\mathcal{D}_k$ (and its inverse). For $k \leq 6$, $\mathcal{D}_k$ is tabulated in Mathai, Provost & Hayakawa (1995) (pp.187–188). Currently, the most efficient approach of computing $\mathcal{D}_k^{-1}$ involves first expressing the zonal polynomials in terms of monomial symmetric functions (see James (1968)) and then express the monomial symmetric functions in terms of power-sum symmetric functions (see David & Kendall (1955)). However, this method is prohibitively time-consuming even for moderately large k. For example, with $k = 20$, Gutiérrez, Rodriguez & Saeź (2000) spent about 8 days to compute a transition matrix that is crucial for the computation of $\mathcal{D}_{20}^{-1}$, and that was only done in double precision. Recent advances in computer technology and improved algorithms have made it possible to compute $\mathcal{D}_k$ faster. Nevertheless, it is still very time-consuming to compute $\mathcal{D}_k$ and $\mathcal{D}_k^{-1}$ for large k.

Note that the transition matrix $\mathcal{D}_k$ is universal – i.e., it does not depend on n, m, or Σ. Hence, the elements of $\mathcal{H}_k$ and $\tilde{\mathcal{H}}_k$ are polynomials in n and rational polynomials in $\tilde{n}$, respectively. Once $\mathcal{H}_k$ and $\tilde{\mathcal{H}}_k$ are computed, they can be used to compute $E[p^{(k)}(W)]$ and $E[p^{(k)}(W^{-1})]$ for any Σ. We will find similar properties in the results that follow.

Next, in an approach that is similar to ours, Letac & Massam (2004) define an $m \times m$ matrix for a given $\kappa \vdash k$:

$$L_\kappa(W) = \sum_{i=1}^{\ell(\kappa)} \kappa_i W^{\kappa_i} p_{\kappa_{(i)}}(W), \tag{25}$$

where $\kappa_{(i)} = (\kappa_1, \cdots, \kappa_{i-1}, \kappa_{i+1}, \ldots, \kappa_m)$, i.e., κ with its i-th element removed. These are then assembled into a large $m\pi(k) \times m$ matrix $L^{(k)}(W)$, with, as usual, the matrix elements $L_\kappa(W)$ arranged in reverse lexicographical order of $\kappa \vdash k$. Note that the matrix $L^{(k)}(W)$ contains many fewer sub-matrices than our $\mathcal{L}_k(W)$. Theorem 4 of Letac & Massam (2004) proves that, in our notation,

$$E[L^{(k)}(W)] = (\mathcal{H}_k \otimes I_m) L^{(k)}(\Sigma), \tag{26}$$

$$E[L^{(k)}(W^{-1})] = (\tilde{\mathcal{H}}_k \otimes I_m) L^{(k)}(\Sigma^{-1}), \tag{27}$$

In particular, the first sub-matrix of $L^{(k)}(W)$ is $L_{(k)}(W) = kW^k$, so $\mathcal{H}_k$ provides an analytical expression for $E[W^k]$. Similarly, we can compute $E[L_{(k)}(W^{-1})] = kE[W^{-k}]$ using $\tilde{\mathcal{H}}_k$ and $L^{(k)}(\Sigma^{-1})$. However, for κ with two or more distinct parts, $L_\kappa(W)$ is a polynomial in W with at least two terms, so the results in Letac & Massam (2004) give the expectations of sums of terms like ours, but cannot provide those of the constituents. For example, their method delivers an expression for $E[2W^2\mathrm{tr}(W) + W\mathrm{tr}(W^2)]$ but cannot give us the expressions for $E[W^2\mathrm{tr}(W)]$ and $E[W\mathrm{tr}(W^2)]$. Our results provide the many terms in the expectations of $\mathcal{L}_k(W)$ and $\mathcal{L}_k(W^{-1})$ that are not available from the results in Letac & Massam (2004).

In this paper, we will develop recurrence relations for the expectations of the terms in $\mathcal{L}_k(W)$. Similar recurrence relations for the case $\Sigma = I_m$ are already available from Graczyk & Vostrikova (2006), Pielaszkiewicz, von Rosen & Singull (2017), and Pielaszkiewicz & Holgersson (2020). However, those results do not yield recurrence relations for the case of general Σ. More importantly, our recurrence relations in fact allow us to obtain explicit expressions for the elements of $\mathcal{Q}_k$ and $\tilde{\mathcal{Q}}_k$. These results reveal the structure of these moments, as well as delivering a very efficient method for either obtaining analytical expressions, or for numerically evaluating them, even for very high values of k. While methods of obtaining analytical expressions for the elements of $\mathcal{Q}_k$ and $\tilde{\mathcal{Q}}_k$ have hitherto not been available for the general case, one can numerically evaluate them. This is because the terms in $\mathcal{L}_k(W)$ can be written sums of various products of elements of W, which are quadratic forms in normal random vectors. So, the problem boils down to computing expectations of products of multiple quadratic forms in normal random vectors. Both explicit expressions and recurrence relations for such objects are available in the literature. For example, explicit expressions are available from Lu & Richard (2001), Graczyk, Letac & Massam (2005), and Redelmeier (2011), and recurrence relations are available from Hillier & Kan (2021). However, it is prohibitively expensive to perform this exercise when k is large.

4 | RECURRENCE RELATIONS

We now present the main results of the paper – recurrence relations for the elements of $\mathcal{Q}_k$ and $\tilde{\mathcal{Q}}_k$. These two recurrence relations are obtained by applying a matrix version of an identity for a Wishart matrix that was given in Haff (1981). The complete proof of Theorem 1 is given in Appendix A. We first give a more general result, then specialize it to the matrices $\mathcal{L}_k(W)$ and $\mathcal{L}_k(W^{-1})$.

Theorem 1. *For $r \geq 0$ and $\lambda \vdash l$, we have the following recurrence relations for $E[W^r p_\lambda(W)]$ and $E[W^{-r} p_\lambda(W^{-1})]$:*

$$E[W^{r+1} p_\lambda(W)] = (n + r)\Sigma E[W^r p_\lambda(W)] + \sum_{j=1}^{r} \Sigma E[W^{r-j} p_j(W) p_\lambda(W)] + 2 \sum_{i=1}^{\ell(\lambda)} \lambda_i \Sigma E\left[W^{r+\lambda_i} p_{\lambda_{(i)}}(W)\right], \tag{28}$$

$$\begin{aligned}\Sigma^{-1} E[W^{-r} p_\lambda(W^{-1})] = {} & (\tilde{n} - r) E[W^{-(r+1)} p_\lambda(W^{-1})] - \sum_{j=1}^{r} E[W^{-r-1+j} p_j(W^{-1}) p_\lambda(W^{-1})] \\ & - 2 \sum_{i=1}^{\ell(\lambda)} \lambda_i E[W^{-r-1-\lambda_i} p_{\lambda_{(i)}}(W^{-1})].\end{aligned} \tag{29}$$

These formulae hold for general r and l, but for the term on the left hand side of (28) to be an element of $\mathcal{L}_{k+1}(W)$, it must be of the form $W^{r+1}p_\lambda(W)$ for some $r \in \{0,\ldots,k\}$, and with λ a partition of $k-r$. To see that equation (28) is indeed a recurrence relation, consider the individual terms in order. First, the term on the left is of degree $k+1$, and by construction is the expectation of one term in the sub-matrix $p^{(k-r)}(W)\otimes W^{r+1}$ of $\mathcal{L}_{k+1}(W)$. On the right, each term is of degree k. The first term is one of the elements of $\mathcal{Q}_k$, that indexed by λ in the sub-matrix corresponding to $p^{(k-r)}(W)\otimes W^r$. In the second term, the term in the sum for $j=r$ is a multiple of I_m, with coefficient $E[p_r(W)p_\lambda(W)]$, which is an element of q_k. This can obviously be obtained from $\mathcal{Q}_k$, but this term is not a sub-matrix of $\mathcal{Q}_k$. Each of the remaining terms in the sum corresponds to an element of the sub-matrix $p^{(k-r+j)}(W)\otimes W^{r-j}$ of $\mathcal{L}_k(W)$, so that term is also a sub-matrix of $\mathcal{Q}_k$. Finally, the i-th term in the last sum is an element of the sub-matrix $p^{(k-r-\lambda_i)}(W)\otimes W^{r+\lambda_i}$ of $\mathcal{L}_k(W)$. Thus, $\mathcal{Q}_{k+1}$ can indeed be constructed from $\mathcal{Q}_k$ and q_k.

Turning now to equation (29), for the left hand side to be an element of $\mathcal{L}_k(W^{-1})$, we need $r \in \{1,\ldots,k\}$ and $\lambda \vdash k-r$. In this case, all terms on the right come from $\mathcal{L}_{k+1}(W^{-1})$. Thus, although a recursion, equation (29) is of a different character from (28) – the reverse of what is required. We will show in Section 6 below that this recursion can be inverted to provide a recurrence relation that is in the correct direction.

It should be emphasized that although Haff's identity typically assumes Σ is nonsingular and $n > m+1$, (28) continues to hold when Σ is singular and $n \leq m+1$. This is because each element of $W^{r+1}p_\lambda(W)$ is a product of $k+1$ elements of W, and the elements of W can be written as quadratic forms in mn normal random variables that have mean zero and a covariance matrix $I_n \otimes \Sigma$. As shown in Don (1979), the explicit expression for the expectation of a product of quadratic forms in normal random variables is the same regardless of whether Σ is singular or nonsingular, and the expression holds for any n. As a result, the explicit expression of $E[W^{r+1}p_\lambda(W)]$ and its recurrence relation continues to hold when Σ is singular or $n \leq m+1$.

Two special cases of Theorem 1 are of interest. The first is for the top-order term $E[W^{r+1}]$, in which case the final term in each recursion drops out, and we have

$$E[W^{r+1}] = (n+r)\Sigma E[W^r] + \sum_{j=1}^{r} \Sigma E[W^{r-j}p_j(W)], \tag{30}$$

$$\Sigma^{-1}E[W^{-r}] = (\tilde{n}-r)E[W^{-(r+1)}] - \sum_{j=1}^{r} E[W^{-r-1+j}p_j(W^{-1})]. \tag{31}$$

Equation (30) produces the coefficients for the moments of a Wishart matrix dealt with in Hillier & Kan (2021). The second special case is the case $r=0$, when the second terms drop out and we have

$$E[p_\lambda(W)W] = nE[p_\lambda(W)]\Sigma + 2\sum_{i=1}^{\ell(\lambda)} \lambda_i \Sigma E\left[W^{\lambda_i}p_{\lambda_{(i)}}(W)\right], \tag{32}$$

$$E[p_\lambda(W^{-1})]\Sigma^{-1} = \tilde{n}E[W^{-1}p_\lambda(W^{-1})] - 2\sum_{i=1}^{\ell(\lambda)} \lambda_i E[W^{-\lambda_i-1}p_{\lambda_{(i)}}(W^{-1})]. \tag{33}$$

Now, the recursion for the elements of $\mathcal{Q}_k$ can be written in a more compact notation, as follows:

Proposition 3. *There is a $d_{k+1}\times d_{k+1}$ matrix D_k of coefficients, with linear terms in n on the diagonal, and constants off the diagonal, satisfying*

$$\mathcal{Q}_{k+1} = (D_k \otimes \Sigma)\begin{bmatrix} \mathcal{Q}_k \\ q_k \otimes I_m \end{bmatrix} = (D_{k,a}\otimes\Sigma)\mathcal{Q}_k + D_{k,b}q_k \otimes \Sigma, \tag{34}$$

where $D_{k,a}$ contains the first d_k columns of D_k (which relate to $\mathcal{Q}_k$), and $D_{k,b}$ the last $\pi(k)$ columns of D_k (which relate to q_k).

The diagonal terms in D_k come from the first term on the right hand side in equation (28). The matrix D_k is quite sparse because each sub-matrix in $\mathcal{Q}_{k+1}$ depends on just a few sub-matrices of $\mathcal{Q}_k$, and a single element of q_k. Further details of how D_k is constructed from the recursion (28) are given in Appendix B. In the next section we will show that the matrix of coefficients $\mathcal{C}_k$ is completely determined by the matrices $D_1, D_2, \ldots, D_{k-1}$.

Example 1. For $k = 1$ to 3, we have

$$D_1 = \begin{bmatrix} n+1 & 1 \\ 2 & n \end{bmatrix}, \tag{35}$$

$$D_2 = \begin{bmatrix} n+2 & 1 & 1 & 0 \\ 2 & n+1 & 0 & 1 \\ 4 & 0 & n & 0 \\ 0 & 4 & 0 & n \end{bmatrix}, \tag{36}$$

$$D_3 = \begin{bmatrix} n+3 & 1 & 1 & 0 & 1 & 0 & 0 \\ 2 & n+2 & 0 & 1 & 0 & 1 & 0 \\ 4 & 0 & n+1 & 0 & 0 & 1 & 0 \\ 0 & 4 & 0 & n+1 & 0 & 0 & 1 \\ 6 & 0 & 0 & 0 & n & 0 & 0 \\ 0 & 4 & 2 & 0 & 0 & n & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 & n \end{bmatrix}. \tag{37}$$

Remark 2. Notice that the row sums of D_k are constant and all equal to $n + 2k$. That is because equation (34) holds for all m, including $m = 1$. In that case every term on the left is $(2\sigma^2)^{k+1}(n/2)_{k+1}$, because W/σ^2 is a χ^2_n variate, while on the right each term is the sum of the elements in a row of D_k multiplied by $\sigma^2(2\sigma^2)^k(n/2)_k$. Hence, each row sum of D_k is equal to

$$\frac{2(n/2)_{k+1}}{(n/2)_k} = n + 2k.$$

Since all elements of D_k are nonnegative, and its row sums are equal, D_k is a (generalized) row-stochastic matrix.

Remark 3. When $\Sigma = I_m$, taking the trace on both sides of (28) and (29) gives

$$\begin{aligned} E[p_{r+1}(W)p_\lambda(W)] &= (n+r)E[p_r(W)p_\lambda(W)] + \sum_{j=1}^{r} E[p_j(W)p_{r-j}(W)p_\lambda(W)] \\ &\quad + 2\sum_{i=1}^{\ell(\lambda)} \lambda_i E\left[p_{r+\lambda_i}(W)p_{\lambda_{(i)}}(W)\right], \end{aligned} \tag{38}$$

$$\begin{aligned} E[p_r(W^{-1})p_\lambda(W^{-1})] &= (\tilde{n}-r)E[p_{r+1}(W^{-1})p_\lambda(W^{-1})] - \sum_{j=1}^{r} E[p_j(W^{-1})p_{r+1-j}(W^{-1})p_\lambda(W^{-1})] \\ &\quad - 2\sum_{i=1}^{\ell(\lambda)} \lambda_i E[p_{r+1+\lambda_i}(W^{-1})p_{\lambda_{(i)}}(W^{-1})]. \end{aligned} \tag{39}$$

These are recurrence relations for $E[p_\kappa(W)]$ and $E[p_\kappa(W^{-1})]$ in the case $\Sigma = I_m$. Equation (38) is the same as the recurrence relation given in Theorem 3.1 of Pielaszkiewicz, von Rosen & Singull (2017). Similarly, (39) is the same as the recurrence relation as given in Theorem 3.1 of Pielaszkiewicz & Holgersson (2020). For the case of $\Sigma = I_m$, Cunden, Mezzadri, Simm & Vivo (2016) showed that there exists a double recurrence (on both k and m) relation for $E[p_k(W)]$ and $E[p_k(W^{-1})]$. However, for general Σ, taking the trace on both sides of (28) and (29) does not lead to recurrence relations for $E[p_\kappa(W)]$ and $E[p_\kappa(W^{-1})]$.

Remark 4. Although we focus on the case that $\lambda_i > 0$, the recurrence relations (28) and (29) actually hold even when some or all of the λ_i's are negative. So our recurrence relation also holds for mixed moments of Wishart and inverse Wishart distributions. For example, setting $r = 0$ and $\lambda = (-1)$ in (28) and (29) gives us

$$E[W\mathrm{tr}(W^{-1})] = n\Sigma E[\mathrm{tr}(W^{-1})] - 2\Sigma E[W^{-1}] = \frac{n\mathrm{tr}(\Sigma^{-1})\Sigma - 2I_m}{\tilde{n}}, \tag{40}$$

$$\begin{aligned} \Sigma^{-1}E[\mathrm{tr}(W)] &= \tilde{n}E[W^{-1}\mathrm{tr}(W)] + 2E[W^0] \\ \Rightarrow E[W^{-1}\mathrm{tr}(W)] &= \frac{n\mathrm{tr}(\Sigma)\Sigma^{-1} - 2I_m}{\tilde{n}}, \end{aligned} \tag{41}$$

where we use the fact $E[W] = n\Sigma$ and $E[W^{-1}] = \Sigma^{-1}/\tilde{n}$.

5 | EXPLICIT EXPRESSIONS FOR MOMENTS INVOLVING W

Although the recurrence relation in (28) allows us to compute the expectations of the elements of $\mathcal{L}_k(W)$, it is generally not computationally efficient to use a recursive algorithm for this task. Instead, one would like to obtain an analytical expression for $\mathcal{Q}_k$ to reduce the burden of computation, and equation (34) provides the means to obtaining such an expression. We will show in this section that the matrices $D_1, D_2, \ldots, D_{k-1}$ in equation (34) completely determine the coefficient matrix $\mathcal{C}_k$. The first step in establishing this claim is:

Proposition 4. *Partition D_k as $D_k = [D_{k,a},\ D_{k,b}]$, as in Proposition 3. The coefficients $c_{k,i}$ in the expansion*

$$\mathcal{Q}_k = \sum_{i=1}^{k} c_{k,i} \otimes \Sigma^i \tag{42}$$

in Proposition 2 are given by

$$c_{k,i} = \left(\prod_{j=1}^{i-1} D_{k-j,a}\right) D_{k-i,b}q_{k-i} \quad \text{for } i = 1, \ldots, k-1, \tag{43}$$

$$c_{k,k} = n\prod_{i=1}^{k-1} D_{k-i,a}, \tag{44}$$

and the matrices $\mathcal{C}_{k,i}$ in Proposition 1 are given by $\mathcal{C}_{k,0} = c_{k,k}$, and

$$\mathcal{C}_{k,s} = \left(\prod_{j=1}^{k-s} D_{k-j,a}\right) D_{s,b}\mathcal{H}_s, \quad \text{for } s = 1, \ldots, k-1. \tag{45}$$

Proof. Back-substitution in equation (34) gives, since $\mathcal{Q}_1 = n\Sigma$,

$$\begin{aligned}\mathcal{Q}_k &= n\prod_{i=1}^{k-1} D_{k-i,a} \otimes \Sigma^k + (D_{k-1,a}D_{k-2,a}\cdots D_{2,a}D_{1,b}q_1) \otimes \Sigma^{k-1} + (D_{k-1,a}D_{k-2,a}\cdots D_{3,a}D_{2,b}q_2) \otimes \Sigma^{k-2} + \cdots \\ &\quad + D_{k-1,a}D_{k-2,b}q_{k-2} \otimes \Sigma^2 + D_{k-1,b}q_{k-1} \otimes \Sigma \\ &= \sum_{i=1}^{k} c_{k,i} \otimes \Sigma^i. \end{aligned} \tag{46}$$

This establishes the formulae for the $c_{k,i}$ in terms of the q_r. The formulae for the $C_{k,s}$ in terms of the $\mathcal{H}_s$ follow from equation (23). □

The result in Proposition 4 expresses the coefficient matrix $\mathcal{C}_k$ in terms of the matrices D_r and $\mathcal{H}_r$ for $r = 1, \ldots, k-1$. The final step in showing that the D_r matrices completely determine $\mathcal{C}_k$ is to show that the matrices $\mathcal{H}_r$ can be written in terms of the D_r and the $\mathcal{C}_r$. We have:

Proposition 5. *Partition D_k as*

$$D_k = \begin{bmatrix} D_{k,11} & D_{k,12} \\ D_{k,21} & D_{k,22} \end{bmatrix}, \tag{47}$$

where $D_{k,11}$ is of dimension $d_k \times d_k$. The matrices $\mathcal{H}_k$ defined below equation (23) and $\mathcal{C}_k$ defined in Proposition 1 are related by:

$$\mathcal{H}_k = \frac{1}{2k} D_{k,21}\mathcal{C}_k D_{k,12}. \tag{48}$$

Proof. Let $\bar{\mathcal{Q}}_k$ and $\bar{\mathcal{L}}_k$ denote the $d_k \times 1$ vectors containing the traces of the $m \times m$ sub-matrices in $\mathcal{Q}_k$ and $\mathcal{L}_k$, respectively. The properties of D_k discussed above and in Appendix B imply that $\bar{\mathcal{Q}}_k = D_{k,12}q_k$ and, likewise, $\bar{\mathcal{L}}_k(\Sigma) = D_{k,12}p^{(k)}(\Sigma)$. Hence,

$$\bar{\mathcal{Q}}_k = \mathcal{C}_k\bar{\mathcal{L}}_k(\Sigma) = \mathcal{C}_k D_{k,12}p^{(k)}(\Sigma) = D_{k,12}q_k. \tag{49}$$

Premultiplying by $D_{k,21}$ and using $D_{k,21}D_{k,12} = 2kI_{\pi(k)}$, we have

$$q_k = (2k)^{-1} D_{k,21} \mathcal{C}_k D_{k,12} p^{(k)}(\Sigma) \tag{50}$$

On the other hand, $q_k = \mathcal{H}_k p^{(k)}(\Sigma)$. But, two polynomials that agree on a continuum of values of their argument must be identical, which gives the result. □

Remark 5. Proposition 5 provides an alternative and more efficient way of computing the matrix $\mathcal{H}_k$ that relates q_k to $p^{(k)}(\Sigma)$. This is because this formula does not require the transition matrix $\mathcal{D}_k$, which involves fractions and is very time consuming to construct for large k. In contrast, the $\mathcal{C}_k$ and the D_k matrices involve only integers, and are easily computed by matrix multiplication.

We have therefore established:

Theorem 2. *The component matrices of $\mathcal{C}_k$ are given by $\mathcal{C}_{k,0} = c_{k,k}$, and*

$$\mathcal{C}_{k,s} = (2s)^{-1} \left(\prod_{j=1}^{k-s} D_{k-j,a} \right) D_{s,b} D_{s,21} \mathcal{C}_s D_{s,12}, \quad \text{for } s = 1, \ldots, k-1, \tag{51}$$

with the initial condition $C_1 = n$.

The following corollary is immediate, and gives simple recurrence relations for the coefficient vectors $c_{k,i}$ and the matrices $\mathcal{C}_k$:

Corollary 1. *The coefficient vectors $c_{k,i}$ satisfy the update equations*

$$c_{k+1,1} = D_{k,b} q_k, \tag{52}$$

$$c_{k+1,i} = D_{k,a} c_{k,i-1} \quad \text{for } i = 2, \ldots, k+1, \tag{53}$$

and the coefficient matrices $\mathcal{C}_k$ satisfy the update equation

$$\mathcal{C}_{k+1} = [D_{k,a} C_k,\ D_{k,b} \mathcal{H}_k] = [D_{k,a} C_k,\ (2k)^{-1} D_{k,b} D_{k,21} \mathcal{C}_k D_{k,12}], \tag{54}$$

with the initial condition $C_1 = n$.

Example 2. To illustrate these results we now compute the coefficient matrices $\mathcal{C}_k$ for $k = 1$ to 4, obtained by using the matrices D_1 to D_3 given in (35)–(37). For $k = 1$, we have $\mathcal{Q}_1 = E[W] = \mathcal{C}_1 \Sigma$ with $\mathcal{C}_1 = n$. For $k = 2$, we have $\pi(2) = d_2 = 2$,

$$\mathcal{Q}_2 = \begin{bmatrix} E[W^2] \\ E[W\text{tr}(W)] \end{bmatrix} = (\mathcal{C}_2 \otimes I_m) \begin{bmatrix} \Sigma^2 \\ \Sigma\text{tr}(\Sigma) \end{bmatrix}, \tag{55}$$

and, since $\mathcal{C}_1 = \mathcal{H}_1 = n$,

$$\mathcal{C}_2 = nD_1 = \begin{bmatrix} n(n+1) & n \\ 2n & n^2 \end{bmatrix}. \tag{56}$$

For $k = 3$, we have $\pi(3) = 3$, $d_3 = 4$,

$$\mathcal{Q}_3 = \begin{bmatrix} E[W^3] \\ E[W^2\text{tr}(W)] \\ E[W\text{tr}(W^2)] \\ E[W\text{tr}(W)^2] \end{bmatrix} = (\mathcal{C}_3 \otimes I_m) \begin{bmatrix} \Sigma^3 \\ \Sigma^2\text{tr}(\Sigma) \\ \Sigma\text{tr}(\Sigma^2)] \\ \Sigma\text{tr}(\Sigma)^2 \end{bmatrix}, \tag{57}$$

and

$$\mathcal{C}_3 = \left[\begin{bmatrix} n+2 & 1 \\ 2 & n+1 \\ 4 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} n(n+1) & n \\ 2n & n^2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ n & 0 \\ 0 & n \end{bmatrix} \begin{bmatrix} n(n+1) & n \\ 2n & n^2 \end{bmatrix}\right]$$
$$= \begin{bmatrix} n^3+3n^2+4n & 2n^2+2n & n^2+n & n \\ 4n^2+4n & n^3+n^2+2n & 2n & n^2 \\ 4n^2+4n & 4n & n^3+n^2 & n^2 \\ 8n & 4n^2 & 2n^2 & n^3 \end{bmatrix}. \tag{58}$$

For $k = 4$, we have $\pi(4) = 5$, $d_4 = 7$,

$$Q_4 = \begin{bmatrix} E[W^4] \\ E[W^3\mathrm{tr}(W)] \\ E[W^2\mathrm{tr}(W^2)] \\ E[W^2\mathrm{tr}(W)^2] \\ E[W\mathrm{tr}(W^3)] \\ E[W\mathrm{tr}(W^2)\mathrm{tr}(W)] \\ E[W\mathrm{tr}(W)^3] \end{bmatrix} = (\mathcal{C}_4 \otimes I_m) \begin{bmatrix} \Sigma^4 \\ \Sigma^3\mathrm{tr}(\Sigma) \\ \Sigma^2\mathrm{tr}(\Sigma^2) \\ \Sigma^2\mathrm{tr}(\Sigma)^2 \\ \Sigma\mathrm{tr}(\Sigma^3) \\ \Sigma\mathrm{tr}(\Sigma^2)\mathrm{tr}(\Sigma) \\ \Sigma\mathrm{tr}(\Sigma)^3 \end{bmatrix}, \tag{59}$$

and similar computations using D_3 and $\mathcal{C}_3$ produces $\mathcal{C}_4 = [\mathcal{C}_{4,a},\ \mathcal{C}_{4,b}]$, with

$$\mathcal{C}_{4a} = \begin{bmatrix} n^4+6n^3+21n^2+20n & 3n^3+9n^2+12n & 2n^3+5n^2+5n \\ 6n^3+18n^2+24n & n^4+3n^3+12n^2+8n & 6n^2+6n \\ 8n^3+20n^2+20n & 12n^2+12n & n^4+2n^3+5n^2+4n \\ 24n^2+24n & 8n^3+8n^2+8n & 2n^3+2n^2+8n \\ 6n^3+18n^2+24n & 12n^2+12n & 6n^2+6n \\ 24n^2+24n & 4n^3+4n^2+16n & 2n^3+2n^2+8n \\ 48n & 24n^2 & 12n^2 \end{bmatrix}, \tag{60}$$

$$\mathcal{C}_{4b} = \begin{bmatrix} 3n^2+3n & n^3+3n^2+4n & 3n^2+3n & n \\ 2n^3+2n^2+2n & 4n^2+4n & n^3+n^2+4n & n^2 \\ n^3+n^2+4n & 4n^2+4n & n^3+n^2+4n & n^2 \\ n^4+n^3+4n^2 & 8n & 6n^2 & n^3 \\ 6n & n^4+3n^3+4n^2 & 3n^3+3n^2 & n^2 \\ 6n^2 & 4n^3+4n^2 & n^4+n^3+4n^2 & n^3 \\ 6n^3 & 8n^2 & 6n^3 & n^4 \end{bmatrix}. \tag{61}$$

The terms in the first row of $\mathcal{C}_3$, and of $\mathcal{C}_4$, both agree with those in Table 1 of Hillier & Kan (2021), where the coefficients defining $E[W^k]$ for $k = 5$ may also be found. Those values were obtained by a completely different, and much less direct, method. It is striking that the coefficient matrix $\mathcal{C}_k$ depends on n, but not on the dimension m of W.

Remark 6. Note that again the row sums in the matrix $\mathcal{C}_k$ are constant, in this case equal to $2^k(n/2)_k$. That is again because equation (9) holds for all m, including $m = 1$, and in that case each term on the left is $(2\sigma^2)^k(n/2)_k$, while every term on the right is the sum of the elements in a row of $\mathcal{C}_k$ multiplied by $(\sigma^2)^k$. Again, therefore, $\mathcal{C}_k$ is a generalized row-stochastic matrix.

Remark 7. Like the transition matrices $\mathcal{D}_k$, the coefficients of the polynomials in the elements of $\mathcal{C}_k$ are universal. They do not depend on Σ, so need only be computed once. In fact, our program only produces the coefficients in the polynomials, and researchers can use the coefficients to compute $\mathcal{Q}_k$ and q_k for any n and Σ.

Remark 8. The matrix $\mathcal{C}_k$ allows us to derive an unbiased estimator of $\mathcal{L}_k(\Sigma)$. This is because from (9), we have

$$E[(\mathcal{C}_k^{-1} \otimes I_m)\mathcal{L}_k(W)] = \mathcal{L}_k(\Sigma). \tag{62}$$

For example, when $k = 2$, this gives us the following unbiased estimators of Σ^2 and $\mathrm{tr}(\Sigma)\Sigma$:

$$E\left[\frac{nW^2 - \mathrm{tr}(W)W}{n(n^2+n-2)}\right] = \Sigma^2, \tag{63}$$

$$E\left[\frac{(n+1)\mathrm{tr}(W)W - 2W^2}{n(n^2+n-2)}\right] = \mathrm{tr}(\Sigma)\Sigma. \tag{64}$$

We have written a set of Matlab programs to create the D_k matrices, as well as to construct $\mathcal{H}_k$ and $\mathcal{C}_k$ for arbitrary k. Our programs are extremely efficient and they are available at `https://www-2.rotman.utoronto.ca/~kan/research.htm`. In the following table, we compare the speed of computing $\mathcal{H}_k$ and $\mathcal{C}_k$ with existing methods. For $\mathcal{H}_k$, we compare our method with the formula $\mathcal{D}_k\mathcal{A}_k\mathcal{D}_k^{-1}$. For $\mathcal{C}_k$, we compare our method with the algorithm used in Hillier & Kan (2021), after modifying it to deal with $E[W^r p_\lambda(W)]$. The comparison is performed on a PC with a Ryzen 5950X CPU. As we can see from Table 1, our new method provides a spectacular speed improvement over the existing methods, especially for large k. In fact, for the computation of $\mathcal{C}_k$, the algorithm based on Hillier & Kan (2021) cannot produce an answer for $k > 10$ even after running the programs for a few days.

Table 1 about here

6 | EXPLICIT EXPRESSIONS FOR MOMENTS INVOLVING W^{-1}

In Section 4, we observed that since the recurrence relation (29) relates sub-matrices of $\tilde{\mathcal{Q}}_k$ to sub-matrices of $\tilde{\mathcal{Q}}_{k+1}$, it cannot be directly used to obtain $\tilde{\mathcal{Q}}_k$. However, if we can find a way to invert this recurrence relation, the resulting recurrence will be a forward recurrence, as for $\mathcal{Q}_k$ in the previous section. To this end, we first observe the following: In equation (29), on the left the terms for $r = 1, \ldots, k$ correspond to sub-matrices of $\tilde{\mathcal{Q}}_k$, and for $r = 0$ they correspond to elements of $\tilde{q}_k$. Also, all terms on the right are sub-matrices of $\tilde{\mathcal{Q}}_{k+1}$, and, importantly, all sub-matrices of $\tilde{\mathcal{Q}}_{k+1}$ appear. Thus, stacking the terms on both sides of the recurrence relations in (29) we have:

$$\begin{bmatrix} (I_{d_k} \otimes \Sigma^{-1})\tilde{\mathcal{Q}}_k \\ \tilde{q}_k \otimes \Sigma^{-1} \end{bmatrix} = (D_{-k} \otimes I_m)\tilde{\mathcal{Q}}_{k+1}, \tag{65}$$

where D_{-k} is a $d_{k+1} \times d_{k+1}$ matrix of coefficients that are obtained from the recurrence relation in (29) . For $k = 1$ to 3, we have

$$D_{-1} = \begin{bmatrix} \tilde{n}-1 & -1 \\ -2 & \tilde{n} \end{bmatrix}, \tag{66}$$

$$D_{-2} = \begin{bmatrix} \tilde{n}-2 & -1 & -1 & 0 \\ -2 & \tilde{n}-1 & 0 & -1 \\ -4 & 0 & \tilde{n} & 0 \\ 0 & -4 & 0 & \tilde{n} \end{bmatrix}, \tag{67}$$

$$D_{-3} = \begin{bmatrix} \tilde{n}-3 & -1 & -1 & 0 & -1 & 0 & 0 \\ -2 & \tilde{n}-2 & 0 & -1 & 0 & -1 & 0 \\ -4 & 0 & \tilde{n}-1 & 0 & 0 & -1 & 0 \\ 0 & -4 & 0 & \tilde{n}-1 & 0 & 0 & -1 \\ -6 & 0 & 0 & 0 & \tilde{n}-1 & 0 & 0 \\ 0 & -4 & -2 & 0 & 0 & \tilde{n} & 0 \\ 0 & 0 & 0 & -6 & 0 & 0 & \tilde{n} \end{bmatrix}. \tag{68}$$

Note that D_{-k} looks exactly the same as D_k except that we need to replace n in D_k with $\tilde{n}$ and change the sign of all the constants in D_k. The row-sums of D_{-k} are again constant, and equal to $\tilde{n} - 2k$. Provided – as we assume – $\tilde{n} > 2k$, D_{-k} is a diagonally dominant matrix, so by the Levy-Desplanques theorem, D_{-k} is nonsingular. We denote its inverse by $\tilde{D}_k = D_{-k}^{-1}$. Multiplying both sides of equation (65) by $\tilde{D}_k \otimes I_m$, and simplifying slightly, we obtain the required recurrence:

Proposition 6. *The matrices $\tilde{\mathcal{Q}}_k = E[\mathcal{L}_k(W^{-1})]$ satisfy the recurrence relation*

$$\tilde{\mathcal{Q}}_{k+1} = (\tilde{D}_k \otimes I_m)\begin{bmatrix} (I_{d_k} \otimes \Sigma^{-1})\tilde{\mathcal{Q}}_k \\ \tilde{q}_k \otimes \Sigma^{-1} \end{bmatrix} = (\tilde{D}_k \otimes \Sigma^{-1})\begin{bmatrix} \tilde{\mathcal{Q}}_k \\ \tilde{q}_k \otimes I_m \end{bmatrix}. \tag{69}$$

Thus, after this step we have, as before, a recurrence relation that expresses each sub-matrix in $\tilde{\mathcal{Q}}_{k+1}$ as a linear combination of sub-matrices of $\tilde{\mathcal{Q}}_k$ pre-multiplied by Σ^{-1}, and various elements of $\tilde{q}_k$ multiplied by Σ^{-1}. For $k = 1$ and 2, we have

$$\tilde{D}_1 = \frac{\begin{bmatrix} \tilde{n} & 1 \\ 2 & \tilde{n}-1 \end{bmatrix}}{(\tilde{n}+1)(\tilde{n}-2)}, \tag{70}$$

$$\tilde{D}_2 = \frac{\begin{bmatrix} \tilde{n}^3-\tilde{n}^2-4\tilde{n} & \tilde{n}^2 & \tilde{n}^2-\tilde{n}-4 & \tilde{n} \\ 2\tilde{n}^2 & \tilde{n}^3-2\tilde{n}^2-4\tilde{n} & 2\tilde{n} & \tilde{n}^2-2\tilde{n}-4 \\ 4\tilde{n}^2-4\tilde{n}-16 & 4\tilde{n} & \tilde{n}^3-3\tilde{n}^2-4\tilde{n}+8 & 4 \\ 8\tilde{n} & 4\tilde{n}^2-8\tilde{n}-16 & 8 & \tilde{n}^3-3\tilde{n}^2-4\tilde{n}+4 \end{bmatrix}}{(\tilde{n}+1)(\tilde{n}-4)(\tilde{n}^2-4)}. \tag{71}$$

We do not report $\tilde{D}_3$ here because the expression is too long to fit on the page. Unlike D_{-k}, $\tilde{D}_k$ is a dense matrix, and its elements are no longer linear in $\tilde{n}$. Instead, they are rational polynomials in $\tilde{n}$, with the polynomials in the numerator and denominator of order $d_{k+1}-1$ and d_{k+1}, respectively.

From this point the steps involved are identical to those given earlier, and are omitted. The key results are:

Theorem 3. *With $\tilde{D}_k$ partitioned as $\tilde{D}_k = [\tilde{D}_{k,a},\ \tilde{D}_{k,b}]$, where $\tilde{D}_{k,a}$ is the first d_k columns of $\tilde{D}_k$ and $\tilde{D}_{k,b}$ is the last $\pi(k)$ columns of $\tilde{D}_k$, we have*

$$\tilde{\mathcal{Q}}_k = \sum_{i=1}^{k} \tilde{c}_{k,i} \otimes \Sigma^{-i}, \tag{72}$$

with

$$\tilde{c}_{k,i} = \left(\prod_{j=1}^{i-1} \tilde{D}_{k-j,a}\right) \tilde{D}_{k-i,b}\tilde{q}_{k-i} \quad \text{for } i = 1,\ldots,k-1, \tag{73}$$

$$\tilde{c}_{k,k} = \frac{1}{\tilde{n}} \prod_{i=1}^{k-1} \tilde{D}_{k-i,a}. \tag{74}$$

Similar to Corollary 1, we have

Corollary 2. *The coefficient vectors $\tilde{c}_{k,i}$ satisfy the recurrence relations,*

$$\tilde{c}_{k+1,1} = \tilde{D}_{k,b}\tilde{q}_k, \tag{75}$$

$$\tilde{c}_{k+1,i} = \tilde{D}_{k,a}\tilde{c}_{k,i-1} \quad \text{for } i = 2,\ldots,k+1, \tag{76}$$

and the coefficient matrices $\tilde{\mathcal{C}}_k$ satisfy the update equation

$$\tilde{\mathcal{C}}_{k+1} = [\tilde{D}_{k,a}\tilde{\mathcal{C}}_k,\ \tilde{D}_{k,b}\tilde{\mathcal{H}}_k] = [\tilde{D}_{k,a}\tilde{\mathcal{C}}_k,\ (2k)^{-1}\tilde{D}_{k,b}D_{-k,21}\tilde{\mathcal{C}}_kD_{-k,12}], \tag{77}$$

with the initial condition of $\tilde{\mathcal{C}}_1 = \tilde{n}^{-1}$, and $D_{-k,12}$ and $D_{-k,21}$ are two sub-matrices of D_{-k}, similarly defined as the corresponding sub-matrices of D_k as in Proposition 5.

Example 3. For $k = 1$ and $\tilde{n} > 0$, $\tilde{\mathcal{C}}_1 = 1/\tilde{n}$. The coefficient matrices $\tilde{\mathcal{C}}_k$ for $k = 2$ to 4 are as follows: For $k = 2$, we have for $\tilde{n} > 2$

$$\tilde{\mathcal{C}}_2 = \frac{1}{\tilde{n}(\tilde{n}-2)(\tilde{n}+1)} \begin{bmatrix} \tilde{n} & 1 \\ 2 & \tilde{n}-1 \end{bmatrix}. \tag{78}$$

For $k = 3$, we have for $\tilde{n} > 4$,

$$\tilde{\mathcal{C}}_3 = \frac{1}{\tilde{n}(\tilde{n}^2-4)(\tilde{n}-4)(\tilde{n}+1)} \begin{bmatrix} \tilde{n}^2 & 2\tilde{n} & \tilde{n} & 2 \\ 4\tilde{n} & \tilde{n}^2-2\tilde{n} & 4 & \tilde{n}-2 \\ 4\tilde{n} & 8 & \tilde{n}^2-2\tilde{n}-4 & \tilde{n}-2 \\ 16 & 4\tilde{n}-8 & 2\tilde{n}-4 & \tilde{n}^2-3\tilde{n}-2 \end{bmatrix}. \tag{79}$$

For $k = 4$, we have for $\tilde{n} > 6$, $\tilde{\mathcal{C}}_4 = [\tilde{\mathcal{C}}_{4,a},\ \tilde{\mathcal{C}}_{4,b}]/\Delta_4$ with $\Delta_4 = (\tilde{n}-2)_6(\tilde{n}-4)(\tilde{n}-6)$ and

$$\tilde{\mathcal{C}}_{4a} = \begin{bmatrix} \tilde{n}^4-\tilde{n}^3+2\tilde{n}^2 & 3\tilde{n}^3-3\tilde{n}^2+6\tilde{n} & 2\tilde{n}^3-3\tilde{n}^2-6\tilde{n} & 5\tilde{n}^2-6\tilde{n} \\ 6\tilde{n}^3-6\tilde{n}^2+12\tilde{n} & \tilde{n}^4-4\tilde{n}^3+3\tilde{n}^2+36 & 12\tilde{n}^2-18\tilde{n}-36 & 2\tilde{n}^3-8\tilde{n}^2 \\ 8\tilde{n}^3-12\tilde{n}^2-24\tilde{n} & 24\tilde{n}^2-36\tilde{n}-72 & \tilde{n}^4-5\tilde{n}^3-6\tilde{n}^2+36\tilde{n}+72 & \tilde{n}^3-5\tilde{n}^2+18\tilde{n} \\ 40\tilde{n}^2-48\tilde{n} & 8\tilde{n}^3-32\tilde{n}^2 & 2\tilde{n}^3-10\tilde{n}^2+36\tilde{n} & \tilde{n}^4-6\tilde{n}^3+3\tilde{n}^2+6\tilde{n} \\ 6\tilde{n}^3-6\tilde{n}^2+12\tilde{n} & 18\tilde{n}^2-18\tilde{n}+36 & 12\tilde{n}^2-18\tilde{n}-36 & 30\tilde{n}-36 \\ 40\tilde{n}^2-48\tilde{n} & 4\tilde{n}^3-16\tilde{n}^2+60\tilde{n}-72 & 2\tilde{n}^3-10\tilde{n}^2+36\tilde{n} & 10\tilde{n}^2-42\tilde{n}+36 \\ 240\tilde{n}-288 & 48\tilde{n}^2-192\tilde{n} & 12\tilde{n}^2-60\tilde{n}+216 & 6\tilde{n}^3-36\tilde{n}^2+18\tilde{n}+36 \end{bmatrix}, \tag{80}$$

$$\tilde{\mathcal{C}}_{4b} = \begin{bmatrix} \tilde{n}^3-\tilde{n}^2+2\tilde{n} & 5\tilde{n}^2-6\tilde{n} & 5\tilde{n}-6 \\ 6\tilde{n}^2-6\tilde{n}+12 & \tilde{n}^3-4\tilde{n}^2+15\tilde{n}-18 & 2\tilde{n}^2-8\tilde{n} \\ 8\tilde{n}^2-12\tilde{n}-24 & \tilde{n}^3-5\tilde{n}^2+18\tilde{n} & \tilde{n}^2-5\tilde{n}+18 \\ 40\tilde{n}-48 & 10\tilde{n}^2-42\tilde{n}+36 & \tilde{n}^3-6\tilde{n}^2+3\tilde{n}+6 \\ \tilde{n}^4-4\tilde{n}^3-9\tilde{n}^2+12\tilde{n}+12 & 3\tilde{n}^3-12\tilde{n}^2-15\tilde{n}+18 & 2\tilde{n}^2-8\tilde{n} \\ 4\tilde{n}^3-16\tilde{n}^2-20\tilde{n}+24 & \tilde{n}^4-6\tilde{n}^3+3\tilde{n}^2+6\tilde{n} & \tilde{n}^3-6\tilde{n}^2+3\tilde{n}+6 \\ 16\tilde{n}^2-64\tilde{n} & 6\tilde{n}^3-36\tilde{n}^2+18\tilde{n}+36 & \tilde{n}^4-7\tilde{n}^3+\tilde{n}^2+35\tilde{n}-6 \end{bmatrix}. \tag{81}$$

Remark 9. From a computational point of view, it is not ideal to compute $\tilde{D}_k$, $\tilde{\mathcal{C}}_k$ and $\tilde{\mathcal{H}}_k$, especially for the purpose of obtaining analytical formulae for $\tilde{\mathcal{Q}}_k$ and $\tilde{q}_k$. This is because elements in $\tilde{D}_k$, $\tilde{\mathcal{C}}_k$ and $\tilde{\mathcal{H}}_k$ are all rational polynomials in $\tilde{n}$ and their orders grow rapidly as k increases. Symbolic computation of sums of products of such terms is very slow, especially for large k. In order to overcome this problem, we work instead with D_{-k}, $\tilde{\mathcal{H}}_k^{-1}$ and $\tilde{\mathcal{C}}_k^{-1}$, because these matrices have elements that are only k-th order polynomials in $\tilde{n}$. Similarly to $\mathcal{C}_k$, it can be readily shown that $\tilde{\mathcal{C}}_k^{-1}$ satisfies the following updating formula:

$$\tilde{\mathcal{C}}_{k+1}^{-1} = [D_{-k,a}\tilde{\mathcal{C}}_k^{-1},\ D_{-k,b}\tilde{\mathcal{H}}_k^{-1}] = [D_{-k,a}\tilde{\mathcal{C}}_k^{-1},\ (2k)^{-1}D_{-k,b}D_{-k,21}\tilde{\mathcal{C}}_k^{-1}D_{-k,12}], \tag{82}$$

where $D_{-k,a}$ and $D_{-k,b}$ are the first d_k and last $\pi(k)$ columns of D_{-k}, respectively. Using the boundary condition $\tilde{\mathcal{C}}_1^{-1} = \tilde{n}$, we can use this updating formula to obtain $\tilde{\mathcal{C}}_k^{-1}$. Once we obtain $\tilde{\mathcal{C}}_k^{-1}$, we only need one matrix inversion to obtain $\tilde{\mathcal{C}}_k$ symbolically. Alternatively, if we just need a numerical answer for $\tilde{\mathcal{Q}}_k$ or $\tilde{q}_k$, we can simply numerically invert $\tilde{\mathcal{C}}_k^{-1}$ or $\tilde{\mathcal{H}}_k^{-1}$ to accomplish that.

Remark 10. The matrix $\tilde{\mathcal{C}}_k$ allows us to derive an unbiased estimator of $\mathcal{L}_k(\Sigma^{-1})$. This is because from (10), we have

$$E[(\tilde{\mathcal{C}}_k^{-1} \otimes I_m)\mathcal{L}_k(W^{-1})] = \mathcal{L}_k(\Sigma^{-1}). \tag{83}$$

For example, when $k = 2$, this gives us the following unbiased estimators of Σ^{-2} and $\mathrm{tr}(\Sigma^{-1})\Sigma^{-1}$:

$$E\left[\tilde{n}(\tilde{n}-1)W^{-2} - \tilde{n}\mathrm{tr}(W^{-1})W^{-1}\right] = \Sigma^{-2}, \tag{84}$$

$$E\left[\tilde{n}^2\mathrm{tr}(W^{-1})W^{-1} - 2\tilde{n}W^{-2}\right] = \mathrm{tr}(\Sigma^{-1})\Sigma^{-1}. \tag{85}$$

7 | THE COMPLEX CASE

In this section we discuss the generalization of our results for the real Wishart distribution to the complex Wishart distribution. Let $W \sim \mathcal{W}_m^c(n, \Sigma)$ follow a complex Wishart distribution with n degrees of freedom and a positive definite Hermitian covariance matrix of Σ. The complex Wishart distribution was introduced by Goodman (1963). For the special case of $W \sim \mathcal{W}_m^c(n, I_m)$, Haagerup & Thorbjørnsen (2003) provided a recurrence relation of $E[p_k(W)]$, and the recurrence relation was extended to deal with $E[p_k(W^{-1})]$ by Cunden, Mezzadri, Simm & Vivo (2016). For the complex Wishart with general Σ, we have slightly different recurrence relations as compared with those for the real Wishart case in Theorem 1. These are given in the following Theorem:

Theorem 4. *Suppose $r \geq 0$ and let $\hat{n} = n - m$. For $W \sim \mathcal{W}_m^c(n, \Sigma)$, we have*

$$E[W^{r+1}p_\lambda(W)] = n\Sigma E[W^r p_\lambda(W)] + \sum_{j=1}^{r} \Sigma E[W^{r-j} p_j(W) p_\lambda(W)] + \sum_{i=1}^{\ell(\lambda)} \lambda_i \Sigma E\left[W^{r+\lambda_i} p_{\lambda_{(i)}}(W)\right], \quad (86)$$

$$\Sigma^{-1}E[W^{-r}p_\lambda(W^{-1})] = \hat{n}E[W^{-(r+1)}p_\lambda(W^{-1})] - \sum_{j=1}^{r} E[W^{-r-1+j} p_j(W^{-1}) p_\lambda(W^{-1})] - \sum_{i=1}^{\ell(\lambda)} \lambda_i E[W^{-r-1-\lambda_i} p_{\lambda_{(i)}}(W^{-1})]. \quad (87)$$

The boundary conditions are $E[W] = n\Sigma$ and $E[W^{-1}] = \Sigma^{-1}/\hat{n}$.

The proof of Theorem 4 is given in Appendix A. Compared with the real case, the matrices D_k and D_{-k} for the complex case are slightly different. On the upper left block of D_k and D_{-k} for the complex Wishart case, the diagonal elements are n or $\hat{n}$, rather than $n + r$ or $\tilde{n} - r$. On the lower left block, all the elements of D_k and D_{-k} are reduced by half for the complex case. Other than these two changes, everything else stays the same. In particular, we have

$$\mathcal{Q}_k = (\mathcal{C}_k \otimes I_m)\mathcal{L}_k(\Sigma), \quad (88)$$

$$\tilde{\mathcal{Q}}_k = (\tilde{\mathcal{C}}_k \otimes I_m)\mathcal{L}_k(\Sigma^{-1}), \quad (89)$$

where $\mathcal{C}_k$ and $\tilde{\mathcal{C}}_k$ satisfy the following update equations

$$\mathcal{C}_{k+1} = [D_{k,a}\mathcal{C}_k,\ k^{-1}D_{k,b}D_{k,21}\mathcal{C}_k D_{k,12}], \quad (90)$$

$$\tilde{\mathcal{C}}_{k+1} = [\tilde{D}_{k,a}\tilde{\mathcal{C}}_k,\ k^{-1}\tilde{D}_{k,b}D_{-k,21}\tilde{\mathcal{C}}_k D_{-k,12}], \quad (91)$$

with the initial conditions $C_1 = n$ and $\tilde{\mathcal{C}}_1 = \hat{n}^{-1}$. Note that unlike Corollaries 1 and 2, we have k^{-1} instead of $(2k)^{-1}$ in the updating equations. This is because for the complex case, $D_{k,21}D_{k,12} = kI_{\pi(k)}$ instead of $(2k)I_{\pi(k)}$ as in the real case.

8 | CONCLUDING COMMENTS

This paper provides analytical results, and highly efficient computational methods, for evaluating the expectations of a complete set of generators for the space of equivariant matrix-valued functions of a Wishart matrix, or of the inverse of such a matrix. These results are based on new recurrence relations for moments of the type $E[W^r p_\lambda(W)]$ and $E[W^{-r} p_\lambda(W^{-1})]$ when $W \sim \mathcal{W}_m(n, \Sigma)$. Thus, the paper provides a structure for analysing the properties of any matrix-valued function in this class. The paper provides results for the (many) moments that could not be dealt with by Letac & Massam (2004), and the corresponding results for the complex case are also given.

The challenge now is to develop analogous results for the case of a noncentral Wishart matrix, or its inverse. It is not immediately obvious how to do that, but the results in Hillier & Kan (2021) indicate that precisely the same coefficients seem to appear in the noncentral case as they do here. So, generalization should be possible.

ACKNOWLEDGMENTS

We thank the Associate Editor, two referees, and participants at the 29th International Workshop on Matrices and Statistics for helpful comments.

REFERENCES

Constantine A. G. (1963). Some non-central distribution problems in multivariate analysis. *Annals of Mathematical Statistics*, *34*(4), 1270–1285.

Cunden F. D., Mezzadri F., Simm N., & Vivo P. (2016). Large-*N* expansion for the time-delay matrix of ballistic chaotic cavities. *Journal of Mathematical Physics*, *57* 111901.

David F. N., & Kendall, M. G. (1955). Tables of symmetric functions. Part V. *Biometrika*, *42*(1/2), 223–242.

Don F. J. H. (1979). The expectation of products of quadratic forms in normal variables. *Statistica Neerlandica*, *33*, 73–79.

Graczyk P., Letac G., & Massam H. (2005). The hyperoctahedral group, symmetric Group representations and the moments of the real Wishart distribution. *Journal of Theoretical Probability* *18*(1), 1–42.

Graczyk P., & Vostrikova L. (2006). The moments of Wishart process via Itô calculus. *Teor. Veroyatnost. i Primenen*, *51*(4), 732–751.

Gutiérrez R., Rodriguez J., & Saeź J. (2000) Approximation of hypergeometric functions with matricial argument through their development in series of zonal polynomials. *Electronic Transactions on Numerical Analysis*, *11*, 121–130.

Goodman N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Annals of Mathematical Statistics*, *34*(1), 152–177.

Haff, L. R. (1979). An identity for the Wishart distribution with applications. *Journal of Multivariate Analysis*, *9*(4), 531–544.

Haff, L. R. (1981). Further identities for the Wishart distribution with application with applications in regression. *Canadian Journal of Statistics*, *9*(2), 215–224.

Hayakawa T., & Kikuchi Y. (1979). The moments of a function of traces of a matrix with a multivariate symmetric normal distribution. *South African Statistical Journal*, *13*, 71–82.

Hillier G., & Kan R. (2021). Moments of a Wishart matrix, *Journal of Quantitative Economics*, *19*(Suppl 1), 141–162.

James A. T. (1961). The distributions of noncentral means with known covariance. *Annals of Mathematical Statistics*, *32*(3), 874–882.

James A. T. (1968). Calculation of zonal polynomial coefficients by use of the Laplace-Beltrami operator. *Annals of Mathematical Statistics*, *39*(5), 1711–1718.

Khatri C. G. (1966). On certain distribution problems based on positive definite quadratic functions in normal vectors. *Annals of Mathematical Statistics*, *37*(2), 468–479.

Konno Y. (2010). Estimation of multivariate complex normal covariance matrices under an invariant quadratic loss. *Communications in Statistics—Theory and Methods*, *39*(8–9), 1490–1497.

Letac G., & Massam H. (2004). All invariant moments of the Wishart distribution. *Scandinavian Journal of Statistics*, *31*, 295–318.

Lu I., & Richard D. S.P. (2001). MacMahaon's master theorem, representation theory, and moments of Wishart distributions. *Advances in Applied Mathematics*, *27*, 531–547.

Mathai A. M., Provost S. B., & Hayakawa T. (1995). *Bilinear forms and zonal polynomials*. Springer, New York.

Pielaszkiewicz J., & Holgersson T. (2020). Mixtures of traces of Wishart and inverse Wishart matrices. *Communications in Statistics—Theory and Methods*, *50*(21), 5084–5100.

Pielaszkiewicz J., von Rosen D., & Singull M. (2017). On $E\left[\prod_{i=0}^{k}\mathrm{Tr}(W^{m_i})\right]$, where $W \sim \mathcal{W}_p(I,n)$. *Communications in Statistics—Theory and Methods*, *46*(6), 2990–3005.

Procesi C. (1976). The invariant theory of $n \times n$ matrices. *Advances in Mathematics*, *19*, 306–381.

Redelmeier C. E. I. (2011), Genus expansion for real Wishart matrices. *Journal of Theoretical Probability* *24*(4), 1044–1062.

Svensson L. (2004). A useful identity for complex Wishart forms. *Technical Report*. Department of Signals and Systems, Chalmers University of Technology.

Haagerup U., & Thosbjørnsen S. (2003) Random matrices with complex Gaussian entries. *Expositiones Mathematicae*, *21*, 293–337.

Watamori Y. (1990). On the moments of traces of Wishart and inverted Wishart matrices. *South African Statistical Journal*, *24*, 153–176.

Corresponding Author: Raymond Kan, Rotman School of Management, University of Toronto, 105 St. George Street, Ontario, Canada, M5S 3E6. E-mail: Raymond.Kan@rotman.utoronto.ca

APPENDIX A

Proof of Theorem 1: Let V be an $m \times q$ matrix, which is a function of W and Σ. Haff (1981) shows that under mild regularity conditions and when the moment exists, we have

$$E[\Sigma^{-1}V] = 2E[DV] + (n-m-1)E[W^{-1}V], \tag{92}$$

where $D = (d_{i,j})$ is an $m \times m$ matrix of differentiation operator for a symmetric matrix, with

$$d_{i,j} = \left(\frac{1+\delta_{i,j}}{2}\right)\frac{\partial}{\partial w_{i,j}}, \tag{93}$$

where $w_{i,j}$ is the (i,j)-th element of W, and $\delta_{i,j} = 1$ if $i = j$ and zero otherwise.

We first present two basic properties for this differential operator.

Lemma 1. *Let $w^{i,j}$ be the (i,j)-th element of W^{-1}. We have*

$$d_{s,t}w_{i,j} = \frac{\delta_{i,s}\delta_{t,j} + \delta_{i,t}\delta_{s,j}}{2}, \tag{94}$$

$$d_{s,t}w^{i,j} = -\frac{w^{i,s}w^{t,j} + w^{i,t}w^{s,j}}{2}. \tag{95}$$

Proof. For (94), we first consider the case that $s = t$. For this case, we must have $i = j = s$ for $d_{s,s}w_{i,j} \neq 0$, and

$$d_{s,s}w_{s,s} = \frac{\partial}{\partial w_{s,s}} w_{s,s} = 1. \tag{96}$$

If $s \neq t$, then there are two cases that $d_{s,t}w_{i,j}$ can be potentially nonzero: (1) $i = s$ and $j = t$, and (2) $i = t$ and $j = s$. For the first case, we have

$$d_{s,t}w_{s,t} = \frac{1}{2}\frac{\partial}{\partial w_{s,t}} w_{s,t} = \frac{1}{2}. \tag{97}$$

For the second case, we have $w_{t,s} = w_{s,t}$ because W is a symmetric matrix. It follows that

$$d_{s,t}w_{t,s} = \frac{1}{2}\frac{\partial}{\partial w_{s,t}} w_{t,s} = \frac{1}{2}. \tag{98}$$

Combining these three cases, we obtain (94).

For (95), we consider the (i,j)-th element of $W^{-1} = W^{-1}WW^{-1}$, which gives us

$$\begin{aligned} d_{s,t}w^{i,j} &= \sum_{r_1=1}^{m}\sum_{r_2=1}^{m} d_{s,t}(w^{i,r_1}w_{r_1,r_2}w^{r_2,j}) \\ &= \sum_{r_1=1}^{m}\sum_{r_2=1}^{m}(d_{s,t}w^{i,r_1})w_{r_1,r_2}w^{r_2,j} + \sum_{r_1=1}^{m}\sum_{r_2=1}^{m} w^{i,r_1}(d_{s,t}w_{r_1,r_2})w^{r_2,j} + \sum_{r_1=1}^{m}\sum_{r_2=1}^{m} w^{i,r_1}w_{r_1,r_2}(d_{s,t}w^{r_2,j}). \end{aligned} \tag{99}$$

For the first term on the right hand side, we have

$$\sum_{r_1=1}^{m}\sum_{r_2=1}^{m}(d_{s,t}w^{i,r_1})w_{r_1,r_2}w^{r_2,j} = \sum_{r_1=1}^{m}(d_{s,t}w^{i,r_1})\delta_{r_1,j} = d_{s,t}w^{i,j}, \tag{100}$$

where the first equality follows because $\sum_{r_2=1}^{m} w_{r_1,r_2} w^{r_2,j}$ is the (r_1,j)-th element of $WW^{-1} = I_m$. Similarly, the third term is also equal to $d_{s,t}w^{i,j}$. Therefore, we have by using (94)

$$\begin{aligned} d_{s,t}w^{i,j} &= -\sum_{r_1=1}^{m}\sum_{r_2=1}^{m} w^{i,r_1}(d_{s,t}w_{r_1,r_2})w^{r_2,j} \\ &= -\sum_{r_1=1}^{m}\sum_{r_2=1}^{m} w^{i,r_1}\left(\frac{\delta_{r_1,s}\delta_{r_2,t}+\delta_{r_1,t}\delta_{r_2,s}}{2}\right)w^{r_2,j} \\ &= -\frac{w^{i,s}w^{t,j}+w^{i,t}w^{s,j}}{2}. \end{aligned} \tag{101}$$

This completes the proof. □

We choose $V = W^{r+1}p_\lambda(W)$ for (28) and $V = W^{-r}p_\lambda(W^{-1})$ for (29). Using $V = W^{r+1}p_\lambda(W)$ in (92), we obtain

$$\Sigma^{-1}E[W^{r+1}p_\lambda(W)] = 2E[DW^{r+1}p_\lambda(W)] + (n-m-1)E[W^r p_\lambda(W)]. \tag{102}$$

It remains to obtain $E[DW^{r+1}p_\lambda(W)]$. In the following, we use Lemma 1 to establish that

$$DW^{r+1}p_\lambda(W) = \frac{1}{2}\sum_{j=0}^{r} W^{r-j}p_j(W)p_\lambda(W) + \left(\frac{r+1}{2}\right)W^r p_\lambda(W) + \sum_{i=1}^{\ell(\lambda)} \lambda_i W^{r+\lambda_i} p_{\lambda_{(i)}}(W). \tag{103}$$

Substituting this in (102), we obtain

$$\begin{aligned} \Sigma^{-1}E[W^{r+1}p_\lambda(W)] &= \sum_{j=0}^{r} E[W^{r-j}p_j(W)p_\lambda(W)] + (r+1)E[W^r p_\lambda(W)] + 2\sum_{i=1}^{\ell(\lambda)} \lambda_i E[W^{r+\lambda_i}p_{\lambda_{(i)}}(W)] \\ &\quad + (n-m-1)E[W^r p_\lambda(W)] \\ &= (n+r)E[W^r p_\lambda(W)] + \sum_{j=1}^{r} E[W^{r-j}p_j(W)p_\lambda(W)] + 2\sum_{i=1}^{\ell(\lambda)} \lambda_i E[W^{r+\lambda_i}p_{\lambda_{(i)}}(W)], \end{aligned} \tag{104}$$

where the last equality is obtained by taking out the term for $j = 0$ in the first summation and using the fact that $p_0(W) = \mathrm{tr}(W^0) = \mathrm{tr}(I_m) = m$. Pre-multiplying by Σ on both sides of the equation gives us (28).

For (103), it can be obtained by using the chain rule and the following two derivatives:

$$DW^{r+1} = \frac{1}{2}\sum_{j=0}^{r} W^{r-j}p_j(W) + \left(\frac{r+1}{2}\right)W^r \quad \text{for } r \geq 0, \tag{105}$$

$$D\mathrm{tr}(W^j) = jW^{j-1} \quad \text{for } j \geq 1. \tag{106}$$

Both (105) and (106) can be proved by using (94). For (105), we use the chain rule to write the (s,t)-th element of DW^{r+1} as

$$\begin{aligned} \sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m} d_{s,l}(w_{l,i_1}w_{i_1,i_2}\cdots w_{i_r,t}) &= \sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m}(d_{s,l}w_{l,i_1})w_{i_1,i_2}\cdots w_{i_r,t} \\ &\quad + \sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m} w_{l,i_1}(d_{s,l}w_{i_1,i_2})\cdots w_{i_r,t} + \cdots \\ &\quad + \sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m} w_{l,i_1}w_{i_1,i_2}\cdots(d_{s,l}w_{i_r,t}). \end{aligned} \tag{107}$$

For the first term, we use (94) to obtain

$$\sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m}(d_{s,l}w_{l,i_1})w_{i_1,i_2}\cdots w_{i_r,t} = \frac{1}{2}\sum_{l=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_r=1}^{m} w_{s,i_2}\cdots w_{i_r,t} + \frac{1}{2}\sum_{i_2=1}^{m}\cdots\sum_{i_r=1}^{m} w_{s,i_2}\cdots w_{i_r,t}, \tag{108}$$

which is the (s,t)-th element of $mW^r/2 + W^r/2 = p_0(W)W^r/2 + W^r/2$. For the second term, we have

$$\sum_{l=1}^{m}\sum_{i_1=1}^{m}\cdots\sum_{i_r=1}^{m} w_{l,i_1}(d_{s,l}w_{i_1,i_2})w_{i_2,i_3}\cdots w_{i_r,t} = \frac{1}{2}\sum_{l=1}^{m}\sum_{i_3=1}^{m}\cdots\sum_{i_r=1}^{m} w_{l,l}w_{s,i_3}\cdots w_{i_r,t} + \frac{1}{2}\sum_{l=1}^{m}\sum_{i_3=1}^{m}\cdots\sum_{i_r=1}^{m} w_{s,l}w_{l,i_3}\cdots w_{i_r,t}, \tag{109}$$

which is the (s,t)-th element of $p_1(W)W^{r-1}/2 + W^r/2$. Repeating this exercise for the other terms, we obtain for $r \geq 0$,

$$DW^{r+1} = \frac{1}{2}\sum_{j=0}^{r} W^{r-j}p_j(W) + \left(\frac{r+1}{2}\right)W^r. \tag{110}$$

For (106), we have

$$\begin{aligned} d_{s,t}\mathrm{tr}(W^j) &= \sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_j=1}^{m} d_{s,t}(w_{i_1,i_2}w_{i_2,i_3}\cdots w_{i_j,i_1}) \\ &= \sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_j=1}^{m}(d_{s,t}w_{i_1,i_2})w_{i_2,i_3}\cdots w_{i_j,i_1} + \sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_j=1}^{m} w_{i_1,i_2}(d_{s,t}w_{i_2,i_3})\cdots w_{i_j,i_1} + \cdots \\ &\quad + \sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_j=1}^{m} w_{i_1,i_2}w_{i_2,i_3}\cdots(d_{s,t}w_{i_j,i_1}). \end{aligned} \tag{111}$$

Using (94), the first term can be written as

$$\begin{aligned} \sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\cdots\sum_{i_j=1}^{m}(d_{s,t}w_{i_1,i_2})w_{i_2,i_3}\cdots w_{i_j,i_1} &= \frac{1}{2}\sum_{i_3=1}^{m}\sum_{i_4=1}^{m}\cdots\sum_{i_j=1}^{m} w_{t,i_3}\cdots w_{i_j,s} + \frac{1}{2}\sum_{i_3=1}^{m}\sum_{i_4=1}^{m}\cdots\sum_{i_j=1}^{m} w_{s,i_3}\cdots w_{i_j,t} \\ &= \frac{1}{2}e_t^{\mathrm{T}}W^{j-1}e_s + \frac{1}{2}e_t^{\mathrm{T}}W^{j-1}e_s = e_s^{\mathrm{T}}W^{j-1}e_t, \end{aligned} \tag{112}$$

which is the (s,t)-th element of W^{j-1}. Repeating this for the other terms, we obtain $D\mathrm{tr}(W^j) = jW^{j-1}$ for $j \geq 1$.

For $V = W^{-r}p_\lambda(W^{-1})$, we first use (95) to show that (the proof is omitted as it is similar to the one before)

$$DW^{-r} = -\frac{1}{2}\sum_{j=1}^{r} W^{-r-1+j}p_j(W^{-1}) - \left(\frac{r}{2}\right)W^{-r-1} \quad \text{for } r \geq 1, \tag{113}$$

$$D\mathrm{tr}(W^{-j}) = -jW^{-j-1} \quad \text{for } j \geq 1, \tag{114}$$

which then gives us

$$DW^{-r}p_\lambda(W^{-1}) = -\frac{1}{2}\sum_{j=1}^{r} W^{-r-1+j}p_j(W^{-1})p_\lambda(W) - \left(\frac{r}{2}\right)W^{-r-1}p_\lambda(W) - \sum_{i=1}^{\ell(\lambda)}\lambda_i W^{-r-1-\lambda_i}p_{\lambda_{(i)}}(W^{-1}). \tag{115}$$

Substituting this in (92) and after simplification, we obtain (29). This completes the proof.

Proof of Theorem 4: For the proof of (86) and (87), we need a complex version of Haff's identity. Such an identity was first developed by Svensson (2004). The version that we use is based on a matrix version of Lemma 3.1 of Konno (2010), which suggests that under mild regularity conditions and when the moments exist, we have

$$E[\Sigma^{-1}V] = E[\tilde{D}V] + (n-m)E[W^{-1}V], \tag{116}$$

where V is an $m \times q$ matrix, which is a function of $W \sim \mathcal{W}_m^c(n,\Sigma)$ and Σ, and $\tilde{D} = (\tilde{d}_{i,j})$ is an $m \times m$ matrix of differentiation operators for a Hermitian matrix, with

$$\tilde{d}_{i,j} = \left(\frac{1+\delta_{i,j}}{2}\right)\left[\frac{\partial}{\partial\Re(w_{i,j})} + \mathrm{i}(1-\delta_{i,j})\frac{\partial}{\partial\Im(w_{i,j})}\right], \tag{117}$$

where $\mathrm{i} = \sqrt{-1}$, and $\Re(w_{i,j})$ and $\Im(w_{i,j})$ stand for the real and imaginary parts of the (i,j)-th element of W, respectively.

The following Lemma summarizes the basic properties of $\tilde{D}$.

Lemma 2. *Let $w^{i,j}$ be the (i,j)-th element of W^{-1}. We have*

$$\tilde{d}_{s,t} w_{i,j} = \delta_{i,t}\delta_{s,j}, \tag{118}$$
$$\tilde{d}_{s,t} w^{i,j} = -w^{i,t} w^{s,j}. \tag{119}$$

We skip the proof because it is similar to that of Lemma 1. Using Lemma 2, we can follow the same proof as in Theorem 1 to establish

$$\tilde{D} W^{r+1} p_\lambda(W) = \sum_{j=0}^{r} W^{r-j} p_j(W) p_\lambda(W) + \sum_{i=1}^{\ell(\lambda)} \lambda_i W^{r+\lambda_i} p_{\lambda_{(i)}}(W), \tag{120}$$
$$\tilde{D} W^{-r} p_\lambda(W^{-1}) = -\sum_{j=1}^{r} W^{-r-1-j} p_j(W^{-1}) p_\lambda(W^{-1}) - \sum_{i=1}^{\ell(\lambda)} \lambda_i W^{-r-1-\lambda_i} p_{\lambda_{(i)}}(W^{-1}). \tag{121}$$

Then substituting these two equations in (116) and after simplification, we obtain (86) and (87). This completes the proof.

APPENDIX B

This Appendix discusses the structure of the matrix D_k defined in Proposition 3. The key to the structure of D_k is the recurrence equation (28),

$$E[W^{r+1} p_\lambda(W)] = (n+r)\Sigma E[W^r p_\lambda(W)] + \sum_{j=1}^{r} \Sigma E[W^{r-j} p_j(W) p_\lambda(W)] + 2\sum_{i=1}^{\ell(\lambda)} \lambda_i \Sigma E\left[W^{r+\lambda_i} p_{\lambda_{(i)}}(W)\right], \tag{122}$$

and the defining equation

$$\mathcal{Q}_{k+1} = (D_{k,a} \otimes \Sigma)\mathcal{Q}_k + D_{k,b} q_k \otimes \Sigma. \tag{123}$$

The question then becomes: which elements of $\mathcal{Q}_k$ and q_k on the right are present in the recursion for a given term on the left, and what are their coefficients? These are the elements of D_k.

To answer this, we partition the $d_{k+1} \times d_{k+1}$ matrix D_k into four sub-matrices as follows:

$$D_k = \begin{bmatrix} D_{k,11} & D_{k,12} \\ D_{k,21} & D_{k,22} \end{bmatrix}, \tag{124}$$

where $D_{k,11}$ is $d_k \times d_k$ and $D_{k,12}$ is $d_k \times \pi(k)$, etc. Note that the first d_k rows of D_k are determined by the recursion for the sub-matrices $E[p^{(k+1-r)}(W) \otimes W^r]$ with $1 < r \le k+1$, and the last $\pi(k)$ rows of D_k are determined by the recursion for $E[p^{(k)}(W)W]$.

Consider first $D_{k,11}$, which identifies the elements of $\mathcal{Q}_k$ that enter the recursion for $E[p^{(k+1-r)}(W) \otimes W^r]$ when $1 < r \le k+1$. From the recurrence relation, we know (1) in the block relating to $p^{(k+1-r)}(W) \otimes W^r$, the diagonal element is $n+r$ (from the first term of the recursion), (2) there are $r-1$ elements to the right of the diagonal, all equal to one (coming from the second term), and (3) to the left of the diagonal, the number of nonzero elements is equal to the number of distinct elements in λ (because of the third term).

Next, the $d_k \times \pi(k)$ matrix $D_{k,12}$, which identifies the elements of q_k that enter the recursion for $E[p^{(k+1-r)}(W) \otimes W^r]$ when $1 < r \le k+1$. Each row has a single element of unity, coming from the term $j = r$ in the sum in the second term. Specifically, the non-zero term in column κ of $D_{k,12}$ is that for which $p_r p_\lambda = p_\kappa$. This matrix encodes useful information: for each row, it tells us how to map $(r,\ \lambda \vdash k-r)$ to $\kappa \vdash k$. On the other hand, each column contains non-zero terms for just those pairs (r, λ) that map to the same $\kappa \vdash k$. It can be seen, in fact, that the number of nonzero elements in column κ is the number of distinct integers in κ.

The lower blocks of D_k tell us the mapping from $\mathcal{Q}_k$ and q_k to each $E[Wp_\kappa(W)]$, with $|\kappa| = k$. The recurrence relation for these terms is

$$E[Wp_\kappa(W)] = nE[p_\kappa(W)]\Sigma + 2\Sigma\sum_{i=1}^{\ell(\kappa)} \kappa_i E[W^{\kappa_i} p_{\kappa_{(i)}}(W)]. \tag{125}$$

From this recurrence relation, we see that there is only one element, equal to n, in the lower block of D_k that comes from q_k, and those are the diagonal elements of $D_{k,22}$. That is, $D_{k,22} = nI_{\pi(k)}$.

It follows at once from this result, and the fact that each row-sum of D_k is $n + 2k$, that the row sums of $D_{k,21}$ are all equal to $2k$. Moreover, the location of the nonzero elements in each row of $D_{k,21}$ is exactly the same as the location of the nonzero elements in the corresponding column in $D_{k,12}$, because the nonzero elements correspond to all the combinations of $(\kappa_i, \kappa_{(i)})$ or (r, λ) such that $\kappa = (r, \lambda)$. Therefore, multiplying a given row of $D_{k,21}$ by the corresponding column of $D_{k,12}$ must give $2k$. Finally, the row of $D_{k,21}$ labelled κ is obviously orthogonal to all columns of $D_{k,12}$ other than that corresponding to κ. These facts imply the important result:

$$D_{k,21}D_{k,12} = 2kI_{\pi(k)}. \tag{126}$$

This result is used in the proof of Proposition 5. These various properties of D_k make its construction for each k quite straightforward, and quick.

TABLE 1 Speed comparison of different methods for computing $\mathcal{H}_k$ and $\mathcal{C}_k$

	Computation Time (in seconds)			
	$\mathcal{H}_k$		$\mathcal{C}_k$	
k	$\mathcal{D}_k\mathcal{A}_k\mathcal{D}_k^{-1}$	Our Method	Hillier and Kan (2021)	Our Method
5	0.235	0.0006	0.243	0.0006
6	0.418	0.0011	0.459	0.0009
7	0.739	0.0021	2.414	0.0014
8	1.396	0.0033	19.182	0.0025
9	2.597	0.0046	351.13	0.0038
10	5.187	0.0068	9598.2	0.0061
11	9.744	0.0105	–	0.0094
12	20.287	0.0192	–	0.0159
13	41.105	0.0401	–	0.0310
14	90.988	0.0843	–	0.0573
15	216.51	0.1703	–	0.1304
16	537.77	0.3152	–	0.2740
17	1207.8	0.6237	–	0.5699
18	2820.7	1.2419	–	1.1509
19	6857.6	2.4334	–	2.2165
20	23329	4.7775	–	4.3033