Moments of a Wishart matrix

Grant Hillier and Raymond Kan
University of Southampton and University of Toronto

May 2021

Abstract

The paper discusses the moments of Wishart matrices, in both the central and noncentral cases. The first part of the paper shows that the expectation map has certain homogeneity and equivariance properties which impose considerable structure on the moments, hitherto unrecognised. The second part of the paper explains how the moments may be computed efficiently. The two parts of the paper are completely independent, but the computations produce precisely the algebraic structure predicted in the first part, as well as reproducing all previously known formulae. A number of examples are given for the more manageable cases.

JEL Classification: C01, C08

Preface

It is a great pleasure to contribute to this special issue in honour of Professor A. L. Nagar. Nagar's contributions to finite sample econometrics, in particular to inference in simultaneous equations models, were extensive, and important. Many of the procedures he worked on involved functions of either a single noncentral Wishart matrix (e.g., TSLS, OLS estimators), or a noncentral Wishart and an independent central Wishart matrix (e.g., LIML and k-class estimators). So it is quite appropriate that our contribution should focus on that particular family of distributions - albeit on a rather different problem than those usually met in econometrics.

1 Introduction

Let x_i, $i = 1, \ldots, n$ be independent $N_m(\mu_i, \Sigma)$ vectors, with possibly different means but common covariance matrix Σ. Let $X = [x_1, \ldots, x_n]'$. If the μ_i are not all zero, then the matrix $W = X'X$ has a non-central Wishart distribution with n degrees of freedom, covariance matrix Σ and matrix of noncentrality parameters $\Omega = \Sigma^{-1}M'M$, where $M = E[X]$ (Muirhead (1982, Section 10.3)).[1] When $M = 0$, then W has a central Wishart distribution. In this paper, we investigate the matrix-valued moments of W

$$\Psi_k(\Sigma, \Lambda) = \mathbb{E}[W^k], \quad k = 1, 2 \ldots, \tag{1}$$

where $\Lambda = M'M$.

It is easy to see that the expectation of $\mathrm{tr}(W^k)$ exists, so the $\Psi_k(\Sigma, \Lambda)$ certainly exist for all positive semi-definite Σ and Λ. Obviously, too, $\Psi_k(\Sigma, \Lambda)$ must be symmetric and positive semi-definite (because W^k is). However, there are deeper properties of the expectation map that provide considerable information about the structure of $\Psi_k(\Sigma, \Lambda)$, and it is these properties that we explain first. Both the central and noncentral Wishart distributions have many applications, in statistics (Bayesian and frequentist), in econometrics (in both finite-sample and asymptotic distribution theory), and in many other disciplines. Properties of the distributions, including of course its moments, are therefore of wide interest. This is the motivation for the present paper, as it has been for a considerable literature on the problem. The extensive bibliography to the paper by Di Nardo (2014) provides excellent coverage of the literature up to 2014.

For $k = 1$ the result is well-known, and straightforward to derive from the definition of W since $\mathbb{E}[x_i x_i'] = \Sigma + \mu_i \mu_i'$, $i = 1, \ldots, n$:

$$\Psi_1(\Sigma, \Lambda) = n\Sigma + \Lambda. \tag{2}$$

For the central case results were given for $k = 2, 3, 4$ in Gupta and Nagar (1999), and Letac and Massam (2004) provide methods that can be used for any k. For the noncentral case the result is known for $k = 2$ (see Magnus and Neudecker (1979) and Neudecker and Wansbeek (1987)). However, for the noncentral case, general results are not known, and are far from straightforward.

This paper is in two parts. In the first part we point out that the expectation operator has certain important homogeneity and invariance properties. These provide, by themselves, considerable insight into the structure of the moments as functions of Σ in the central case, and of (Σ, Λ) in the noncentral case. We start with the central case, and then develop the more complex analogous argument for the noncentral case. These are qualitative results, so in the second part of the paper we explain how the terms in the expansions can be efficiently computed. It must be emphasised that the computational method in no way *imposes* the algebraic structure discussed in the first part of the paper on the formulae obtained, but it does produce results that confirm its validity completely. The expressions for the moments of course become rapidly unmanageable as k increases, but details for a number of the manageable low-order moments are given, here for the first time. We denote by $\mathcal{O}(m)$ the group of real $m \times m$

[1] We do not require $n \geq m$, and W can be singular.

orthogonal matrices, and denote that a non-increasing sequence $\kappa = (\kappa_1, \kappa_2, \ldots)$ of non-negative integers is a partition of k, i.e., $\sum_{i=1}^{m} \kappa_i = k$, by $\kappa \vdash k$.

2 General structure: central case

In general, if P_m is the space of positive semi-definite $m \times m$ symmetric matrices, it is clear that for each $k \geq 1$ the expectation operator $\mathbb{E}[W^k]$ maps P_m into itself: $\Sigma :\to \Psi_k(\Sigma)$. Two properties of the map $\Psi_k(\Sigma)$ are immediate from the density:

Proposition 1 *(i) The function $\Psi_k(\Sigma)$ is equivariant under the conjugate action of $\mathcal{O}(m)$:*

$$\Psi_k(H\Sigma H') = H\Psi_k(\Sigma)H', \quad H \in \mathcal{O}(m). \tag{3}$$

And, (ii) for any $t > 0$,

$$\Psi_k(t\Sigma) = t^k \Psi_k(\Sigma), \tag{4}$$

so $\Psi_k(\Sigma)$ is homogeneous of degree k in Σ.

Proof. (i) We let $\tilde{X} = XH$. When $M = 0$, $W = X'X$ has a central Wishart distribution with covariance matrix Σ and $\tilde{W} = \tilde{X}'\tilde{X} = H'WH$ has a central Wishart distribution with a covariance matrix $H\Sigma H'$, and $\Psi_k(H\Sigma H') = \mathbb{E}[\tilde{W}^k] = \mathbb{E}[HW^kH'] = H\Psi_k(\Sigma)H'$. (ii) In this case, we set $\tilde{X} = \sqrt{t}X$, then $\tilde{W} = \tilde{X}'\tilde{X}$ is a central Wishart with a covariance matrix $t\Sigma$, and $\Psi_k(t\Sigma) = \mathbb{E}[\tilde{W}^k] = \mathbb{E}[t^k W^k] = t^k \Psi_k(\Sigma)$. ■

Now, the subset of P_m consisting of those matrices that are equivariant under $\mathcal{O}(m)$ and homogeneous of degree k consists of those matrices of the form

$$\Psi_k(\Sigma) = \sum_{r=1}^{k} c_{k-r}(\Sigma)\Sigma^r, \tag{5}$$

where the scalar coefficients $c_{k-r}(\Sigma)$ are homogeneous of degree $k - r$, and invariant under $\Sigma \to H\Sigma H'$, $H \in \mathcal{O}(m)$. See, for instance, Procesi (1976) (where the equivariants under the group action are called "concomitants"), or the argument in Letac and Massam (2004).[2]

For $r = k$, c_0 is of degree 0, i.e., a constant independent of Σ. For $r < k$ the coefficients may, because they are homogeneous symmetric polynomials, be written as linear combinations of the power-sum symmetric functions indexed by partitions of $k - r$:

$$p_\lambda(\Sigma) = \prod_{i=1}^{\ell(\lambda)} p_{\lambda_i}(\Sigma), \quad \lambda \vdash k - r, \tag{6}$$

[2]It is easy to see that the expression given for $\Psi_k(\Sigma)$ has the required properties. Showing that any matrix in P_m having these properties can be written in this form is a more subtle argument, but is classical.

where $p_i(\Sigma) = \mathrm{tr}(\Sigma^i)$ and $\ell(\lambda)$ stands for the number of nonzero parts in λ. That is, for certain numerical coefficients c_λ, where $\lambda \vdash k-r$, $c_{k-r}(\Sigma)$ may be written as

$$c_{k-r}(\Sigma) = \sum_{\lambda \vdash k-r} c_\lambda p_\lambda(\Sigma). \tag{7}$$

Thus, we may write

$$\Psi_k(\Sigma) = c_0 \Sigma^k + \sum_{r=1}^{k-1} \Sigma^r \left[\sum_{\lambda \vdash k-r} c_\lambda p_\lambda(\Sigma) \right]. \tag{8}$$

Any other basis for the set of homogeneous invariant polynomials could also be used. These properties of $\Psi_k(\Sigma)$ are evident in the results obtained in the literature for some small values of k (e.g., those reported in Gupta and Nagar (1999), p.99, or the formulae given in Letac and Massam (2004)). The problem therefore reduces to that of finding the coefficients c_λ in this expression for the expectation: *the c_λ are the only terms in the expansion (8) that are unknown.* For $k = 2, 3, 4, 5$, we obtain, by the methods described below, the following values for the c_λ :

	$k=2$	$k=3$	$k=4$	$k=5$
c_0	$n(n+1)$	$n(n^2+3n+4)$	$n(n^3+6n^2+21n+20)$	$n(n^4+10n^3+65n^2+160n+148)$
$c_{(1)}$	n	$2n(n+1)$	$3n(n^2+3n+4)$	$4n(n^3+6n^2+21n+20)$
$c_{(2)}$	-	$n(n+1)$	$n(2n^2+5n+5)$	$3n(n^3+5n^2+14n+12)$
$c_{(11)}$	-	n	$3n(n+1)$	$6n(n^2+3n+4)$
$c_{(3)}$	-	-	$n(n^2+3n+4)$	$2n(n^3+5n^2+14n+12)$
$c_{(21)}$	-	-	$3n(n+1)$	$4n(2n^2+5n+5)$
$c_{(111)}$	-	-	n	$4n(n+1)$
$c_{(4)}$	-	-	-	$n(n^3+6n^2+21n+20)$
$c_{(31)}$	-	-	-	$4n(n^2+3n+4)$
$c_{(22)}$	-	-	-	$n(2n^2+5n+5)$
$c_{(211)}$	-	-	-	$6n(n+1)$
$c_{(1111)}$	-	-	-	n

Table 1: Coefficients in the expansion of $\mathbb{E}[W^k]$ for $k = 2, \ldots, 5$.

These coefficients agree with the expressions reported by Gupta and Nagar (1999) for $k = 2, 3$, but both their formula for the case $k = 4$, and that given in its source, de Waal and Nel (1973), are incorrect. The results reported for $k = 4$ agree with those given by Letac and Massam (2004), equation (46). Evidently the coefficients are polynomials in n, and there is clearly a great deal of structure here that we have yet to understand.

Remark 1 *The coefficients* c_λ *in the expansion do not depend on the matrix dimension* m. *Considering the case* $m = 1$, *therefore, we just have the* k*-th moment of* $w \sim \sigma^2\chi^2_n$, *namely:*

$$E[w^k] = (2\sigma^2)^k \left(\frac{n}{2}\right)_k. \tag{9}$$

Writing out equation (8) for $m = 1$, *on the other hand, we have*

$$E[w^k] = (\sigma^2)^k \sum_{r=1}^{k} \sum_{\lambda \vdash k-r} c_\lambda. \tag{10}$$

We therefore have, for any k, *and any* m, *the identity*

$$\sum_{r=1}^{k} \sum_{\lambda \vdash k-r} c_\lambda = 2^k \left(\frac{n}{2}\right)_k. \tag{11}$$

This identity is exactly true in the above table, and there are evidently other properties of the c_λ *that we have yet to unearth.*

It should be emphasised again that the computational procedure described below does not *impose* the structure in equation (8) on the result, but, as expected, reproduces it precisely. We give the more general analogous results for the noncentral case next.

3 General structure: noncentral case

In the noncentral case the expectation operator defines a map $\Psi_k(\Sigma, \Lambda)$ from $P_m \times P_m$ to the set of $m \times m$ symmetric matrices. It is obvious again that the expectation $\Psi_k(\Sigma, \Lambda)$ must be a positive semi-definite symmetric matrix, so the analogue of Proposition 1 is the following:

Proposition 2 *The map* $\Psi_k(\Sigma, \Lambda) = \mathbb{E}[W^k]$ *defined on* $P_m \times P_m$ *has the following properties: (i)* $\Psi_k(\Sigma, \Lambda)$ *is positive semi-definite symmetric, so the map is* $P_m \times P_m \longmapsto P_m$*; (ii)* $\Psi_k(\Sigma, \Lambda)$ *is equivariant under the simultaneous action of* $\mathcal{O}(m)$*:* $\Sigma \to H\Sigma H'$, $\Lambda \to H\Lambda H'$, $H \in \mathcal{O}(m)$, *i.e., for all* $H \in \mathcal{O}(m)$,

$$\Psi_k(H\Sigma H', H\Lambda H') = H\Psi_k(\Sigma, \Lambda)H', \tag{12}$$

and (iii) $\Psi_k(\Sigma, \Lambda)$ *is homogeneous of degree* k *in* (Σ, Λ), *i.e.,* $\Psi_k(t\Sigma, t\Lambda) = t^k\Psi_k(\Sigma, \Lambda)$.

The proof of (ii) and (iii) is identical to that of Proposition 1, and is left to the reader. The fact that $\Psi_k(\Sigma, \Lambda)$ must be positive semi-definite symmetric (with probability 1), i.e., $\Psi_k(\Sigma, \Lambda) \in P_m$ follows directly from the fact that W^k has that property almost surely.[3]

We are therefore concerned with the set of symmetric, homogeneous 2-matrix polynomials that are invariant under the simultaneous transformations $\Sigma \to H\Sigma H'$, $\Lambda \to H\Lambda H'$, $H \in \mathcal{O}(m)$. For each $r = 1, \dots, k$, there are 2^r monomials in the two matrices of degree r. Each such product has the required equivariance property, and there are $\Sigma_{r=1}^k 2^r = 2(2^k - 1)$ monomials in total. Let $a(r) = (a_1, \dots, a_r)$ be a sequence of length r in which each term is either 0 or 1, the order of terms being important. The monomials of degree r in (Σ, Λ) may be written as

$$Q_{a(r)}(\Sigma, \Lambda) = \prod_{i=1}^{r} \Sigma^{a_i} \Lambda^{1-a_i}. \tag{13}$$

Matrices of the form

$$\sum_{r=1}^{k} \sum_{a(r)} c_{a(r)}(\Sigma, \Lambda) Q_{a(r)}(\Sigma, \Lambda), \tag{14}$$

in which the coefficients $c_{a(r)}(\Sigma, \Lambda)$ are homogeneous of degree $k-r$, and are invariant under the simultaneous transformations $\Sigma \to H\Sigma H'$, $\Lambda \to H\Lambda H'$, $H \in \mathcal{O}(m)$, have the required invariance and homogeneity properties, and any homogeneous equivariant polynomial in (Σ, Λ) can be written in this form.

Now, this observation implies that $\mathbb{E}[W^k]$ has an expansion in which the terms are a smaller set of matrices, namely, sums of the matrices in the equivalence class defined by constancy of the trace function. That is, the coefficients $c_{a(r)}(\Sigma, \Lambda)$ in equation (14) may be assumed constant on all sequences $a(r)$ for which $\mathrm{tr}(Q_{a(r)}(\Sigma, \Lambda))$ is constant. The argument, briefly, is as follows: taking the trace of the matrix (14) produces an expansion in which the terms are the sums of all monomials with a common trace. It follows by linearity that, for every nonsingular matrix A, $\mathbb{E}[\mathrm{tr}(AW^k)]$ has an expansion of this form, and this implies that $\mathbb{E}[W^k]$ itself has such an expansion. See, for instance, the proof of Theorem 2.1 in Procesi (1976). We call this set of matrices a *basis* for the expectation $\Psi_k(\Sigma, \Lambda)$.

Example 1 *r=4. In this case there are $2^4 = 16$ monomials, but the basis consists of the following 6 matrices (defined by the value of* $\mathrm{tr}(\cdot)$*):*

$\mathrm{tr}(\Sigma^4)$:	Σ^4
$\mathrm{tr}(\Lambda^4)$:	Λ^4
$\mathrm{tr}(\Sigma^3\Lambda)$:	$\Sigma^3\Lambda + \Lambda\Sigma^3 + \Sigma^2\Lambda\Sigma + \Sigma\Lambda\Sigma^2$
$\mathrm{tr}(\Sigma^2\Lambda^2)$:	$\Sigma^2\Lambda^2 + \Sigma\Lambda^2\Sigma + \Lambda\Sigma^2\Lambda + \Lambda^2\Sigma^2$
$\mathrm{tr}(\Sigma\Lambda\Sigma\Lambda)$:	$\Sigma\Lambda\Sigma\Lambda + \Lambda\Sigma\Lambda\Sigma$
$\mathrm{tr}(\Sigma\Lambda^3)$:	$\Sigma\Lambda^3 + \Lambda^3\Sigma + \Lambda\Sigma\Lambda^2 + \Lambda^2\Sigma\Lambda$

.

Now, to examine this basis further, we make the following observations:

[3]When Σ is positive definite, then $\Psi_k(\Sigma, \Lambda)$ is also positive definite.

1. The terms in a sum with constant trace necessarily have the same number of matrices of each type. That is, $\Sigma_{i=1}^{r} a_i = s$ is constant (and $\Sigma_{i=1}^{r}(1-a_i) = r-s$) on the terms in a sum, for $s = 0, 1, \ldots, r$. We shall say that a term with s Σs and $(r-s)$ Λs is of *type* $(s\Sigma, (r-s)\Lambda)$, for $s = 0, 1, \ldots, r$. Note that for $r = 4$, there are two possible traces of type $(2\Sigma, 2\Lambda)$, but in all other cases there is one possible trace from a given type.

2. The trace function defines an equivalence relation on the set of (non-commutative) matrix-monomials $Q_{a(r)}(\Sigma, \Lambda)$. The basis elements are sums over the orbits defined by $\mathrm{tr}(\cdot)$, i.e., invariants under the group operation that leaves $\mathrm{tr}(\cdot)$ fixed. For arbitrary $m \times m$ matrices the orbit representatives for the equivalence classes defined by $\mathrm{tr}(\cdot)$ are called *necklaces;* the terms in the orbit are generated from a single term of type $(s\Sigma, (r-s)\Lambda)$ by cyclic permutation of its members. This is the only subgroup of the symmetric group S_r that leaves the trace invariant for general $m \times m$ matrices. However, our matrices Σ and Λ are both symmetric, so the trace of a product is also invariant under transposition of the product, which implies invariance under *reflections* as well. The orbit representatives for our problem are therefore two-color *bracelets,* the terms in an orbit being generated by cyclic permutations and reflections, i.e., by the action of the dihedral group D_r. The basis matrices defined by this equivalence are necessarily symmetric, because invariant under reflections.[4]

It follows from these observations that $\Psi_k(\Sigma, \Lambda)$ can be written in the form

$$\Psi_k(\Sigma, \Lambda) = \sum_{r=1}^{k} \sum_{s=0}^{r} \sum_{\pi \in \mathcal{B}(s, r-s)} c_\pi(\Sigma, \Lambda) B_\pi(\Sigma, \Lambda), \tag{15}$$

where $\mathcal{B}(s, r-s)$ is the set of two-color bracelets of type $(s\Sigma, (r-s)\Lambda)$, and $B_\pi(\Sigma, \Lambda)$ is the sum of the terms in the orbit indexed by π. That is,

$$B_\pi(\Sigma, \Lambda) = \sum_{a(r):\mathrm{tr}(Q_{a(r)}(\Sigma,\Lambda)) = \mathrm{tr}(Q_\pi(\Sigma,\Lambda))} Q_{a(r)}(\Sigma, \Lambda) \tag{16}$$

Here we denote different two-color bracelets by the subscript π, which can be regarded again as a string of zeros and ones, a particular sequence $a(r)$ identifying an orbit representative. Following convention, we define the two-color bracelet π as the lexicographically smallest element of $a(r)$ among the set that produces the same $\mathrm{tr}(Q_{a(r)}(\Sigma, \Lambda))$. In the example above for $r = 4$ there is just a single bracelet for $s = 0, 1, 3, 4$, but two bracelets for $s = 2$, a total of 6. Denoting by $B(r, 2)$ the number of two-color bracelets of length r we have:

[4]Thus, $\Psi_k(\Sigma, \Lambda)$ is a linear combination of symmetric matrices, and is therefore itself symmetric. There is no need to impose symmetry as a separate condition.

r	1	2	3	4	5	6	7	8	9	10
$B(r,2)$	2	3	4	6	8	13	18	30	46	78
$\sum_{i=1}^{r} B(i,2)$	2	5	9	15	23	36	54	84	130	208

Table 2: Numbers of two-color bracelets of length $r = 1, \ldots, 10$.

The bottom row of the table gives the cumulative sum of the numbers of 2-color bracelets. So, for instance, for $k = 8$ there are a total of 84 bracelets (i.e., terms in the sum (15)), considerably less than the number of monomials in (14), which is $2^9 - 2 = 510$.

Turning now to the coefficients of the various orbit-sums $B_\pi(\Sigma, \Lambda)$, for $r = k$ the coefficients of all $B_\pi(\Sigma, \Lambda)$ are of degree zero, i.e., constants independent of (Σ, Λ). Thus, we may write

$$\Psi_k(\Sigma, \Lambda) = \sum_{s=0}^{k} \sum_{\pi \in \mathcal{B}(s,k-s)} c_\pi B_\pi(\Sigma, \Lambda) + \sum_{r=1}^{k-1} \sum_{s=0}^{r} \sum_{\pi \in \mathcal{B}(s,r-s)} c_\pi(\Sigma, \Lambda) B_\pi(\Sigma, \Lambda), \tag{17}$$

where the c_π in the first term are numerical constants. For $r < k$ the coefficients $c_\pi(\Sigma, \Lambda)$ in the second term are homogeneous polynomials in (Σ, Λ) of degree $k - r$, invariant under the simultaneous transformations $\Sigma \to H\Sigma H'$, $\Lambda \to H\Lambda H'$, $H \in \mathcal{O}(m)$. We can therefore write

$$c_\pi(\Sigma, \Lambda) = \sum_{l_1 + l_2 = k - r} c_\pi^{l_1, l_2}(\Sigma, \Lambda), \tag{18}$$

the notation here indicating that the coefficient is of degree l_1 in Σ, and l_2 in Λ.

Now, we can express the coefficient $c_\pi^{l_1,l_2}(\Sigma, \Lambda)$ in terms of any basis for the space of polynomials of degree l_1 in Σ, l_2 in Λ, with $l_1 + l_2 = k - r$, invariant under the simultaneous transformations $\Sigma \to H\Sigma H'$, $\Lambda \to H\Lambda H'$, $H \in \mathcal{O}(m)$. Davis (1980) discusses two bases for the space of polynomials with these properties — his invariant polynomials with two symmetric matrix arguments, $C_\phi^{\alpha,\lambda}(\Sigma, \Lambda)$, where $\alpha \vdash l_1$, $\lambda \vdash l_2$, and $\phi \vdash l_1 + l_2$, and a set of trace-polynomials — generalizations of the power-sum basis in the case of a single matrix. These are products of terms like

$$\left[\mathrm{tr}\left(\Sigma^{a_1} \Lambda^{b_1} \Sigma^{c_1} \cdots\right)\right]^{r_1} \left[\mathrm{tr}\left(\Sigma^{a_2} \Lambda^{b_2} \Sigma^{c_2} \cdots\right)\right]^{r_2} \cdots \tag{19}$$

of total degrees l_1, l_2 in Σ, Λ respectively (only distinct terms appearing in the basis). For instance, for $(l_1, l_2) = (2, 1)$ the distinct elements are given by the vector

$$P^{2,1}(\Sigma, \Lambda) = [\mathrm{tr}(\Sigma^2 \Lambda),\ \mathrm{tr}(\Sigma^2)\mathrm{tr}(\Lambda),\ \mathrm{tr}(\Sigma\Lambda)\mathrm{tr}(\Sigma),\ \mathrm{tr}(\Sigma)^2 \mathrm{tr}(\Lambda)]', \tag{20}$$

where we use $P^{l_1,l_2}(\Sigma,\Lambda)$ to denote the vector of distinct products of traces of degrees l_1 in Σ and l_2 in Λ.[5] We shall express our results in terms of the latter basis, because it is a natural generalization of the power-sum basis in the central case. Since the coefficients $c_\pi^{l_1,l_2}(\Sigma,\Lambda)$ can be written as linear combinations of elements in the basis vector $P^{l_1,l_2}(\Sigma,\Lambda)$, we can write them in the form

$$c_\pi^{l_1,l_2}(\Sigma,\Lambda) = \sum_{i=1}^{|P^{l_1,l_2}|} c_{\pi,i}^{l_1,l_2} P_i^{l_1,l_2}(\Sigma,\Lambda), \tag{21}$$

where $|P^{l_1,l_2}|$ denotes the number of elements in $P^{l_1,l_2}(\Sigma,\Lambda)$. That is,

$$c_\pi(\Sigma,\Lambda) = \sum_{l_1+l_2=k-r} \sum_{i=1}^{|P^{l_1,l_2}|} c_{\pi,i}^{l_1,l_2} P_i^{l_1,l_2}(\Sigma,\Lambda). \tag{22}$$

As in the central case, the only unknowns in the problem are the coefficients $c_{\pi,i}^{l_1,l_2}$ and these turn out again to be just polynomials in n. In the examples below we report not the numerical coefficients $c_{\pi,i}^{l_1,l_2}$ (because that would require a complicated diversion into the question of indexing the trace-polynomials in the vectors $P^{l_1,l_2}(\Sigma,\Lambda)$), but just the coefficients $c_\pi(\Sigma,\Lambda)$ in equation (17).

We end this section with examples of these expansions for $k = 1,..,4$. We have already noted the result for $\mathbb{E}[W]$, and the following tables give the explicit expressions for $\mathbb{E}[W^k]$ for $k = 2,3,4$. These expansions were obtained by the algorithm to be discussed in the next section. In the tables, we provide the terms in the expansion of $\mathbb{E}[W^k]$ and use the notation (X) for $\mathrm{tr}(X)$.

r	$B_\pi(\Sigma,\Lambda)$	$c_\pi(\Sigma,\Lambda)$
1	Σ	$n(\Sigma)+(\Lambda)$
	Λ	(Σ)
2	Σ^2	$n(n+1)$
	$\Sigma\Lambda+\Lambda\Sigma$	$n+1$
	Λ^2	1

Table 3a: Expansion of $\mathbb{E}[W^2]$. There are 6 monomials, but 5 bracelets in the expansion.

[5] We have developed a computer program to generate the vector $P^{l_1,l_2}(\Sigma,\Lambda)$ for arbitrary degrees (l_1,l_2).

r	$B_\pi(\Sigma,\Lambda)$	$c_\pi(\Sigma,\Lambda)$
1	Σ	$n(\Sigma)^2+(n+1)[n(\Sigma^2)+2(\Sigma\Lambda)]+2(\Sigma)(\Lambda)+(\Lambda^2)$
	Λ	$(\Sigma)^2+(n+1)(\Sigma^2)+(\Sigma\Lambda)$
2	Σ^2	$2(n+1)[n(\Sigma)+(\Lambda)]$
	$\Sigma\Lambda+\Lambda\Sigma$	$2(n+1)(\Sigma)+(\Lambda)$
	Λ^2	$2(\Sigma)$
3	Σ^3	$n(n^2+3n+4)$
	$\Sigma^2\Lambda+\Sigma\Lambda\Sigma+\Lambda\Sigma^2$	n^2+3n+4
	$\Sigma\Lambda^2+\Lambda\Sigma\Lambda+\Lambda^2\Sigma$	$n+2$
	Λ^3	1

Table 3b: Expansion of $\mathbb{E}[W^3]$. There are 14 monomials, but 9 bracelets in the expansion.

r	$B_\pi(\Sigma,\Lambda)$	$c_\pi(\Sigma,\Lambda)$
1	Σ	$n(\Sigma)^3+3n(n+1)(\Sigma)(\Sigma^2)+6(n+1)(\Sigma)(\Sigma\Lambda)$ $+(n^2+3n+4)[n(\Sigma^3)+3(\Sigma^2\Lambda)]$ $+3(n+1)(\Lambda)(\Sigma^2)+3(\Sigma)^2(\Lambda)+3(n+2)(\Sigma\Lambda^2)$ $+3(\Lambda)(\Sigma\Lambda)+3(\Sigma)(\Lambda^2)+(\Lambda^3)$
	Λ	$(\Sigma)^3+3(n+1)(\Sigma)(\Sigma^2)+(n^2+3n+4)(\Sigma^3)$ $+3(\Sigma)(\Sigma\Lambda)+2(n+2)(\Sigma^2\Lambda)+(\Lambda)(\Sigma^2)+(\Sigma\Lambda^2)$
2	Σ^2	$3n(n+1)(\Sigma)^2+(2n^2+5n+5)[n(\Sigma^2)+2(\Sigma\Lambda)]$ $+6(n+1)(\Sigma)(\Lambda)+(2n+3)(\Lambda^2)+(\Lambda)^2$
	$\Sigma\Lambda+\Lambda\Sigma$	$3(n+1)(\Sigma)^2+(2n^2+5n+5)(\Sigma^2)$ $+(3n+5)(\Sigma\Lambda)+3(\Sigma)(\Lambda)+(\Lambda^2)$
	Λ^2	$3(\Sigma)^2+(2n+3)(\Sigma^2)+2(\Sigma\Lambda)$
3	Σ^3	$3(n^2+3n+4)[n(\Sigma)+(\Lambda)]$
	$\Sigma^2\Lambda+\Sigma\Lambda\Sigma+\Lambda\Sigma^2$	$3(n^2+3n+4)(\Sigma)+2(n+2)(\Lambda)$
	$\Sigma\Lambda^2+\Lambda\Sigma\Lambda+\Lambda^2\Sigma$	$3(n+2)(\Sigma)+(\Lambda)$
	Λ^3	$3(\Sigma)$
4	Σ^4	$n(n^3+6n^2+21n+20)$
	$\Sigma^3\Lambda+\Sigma^2\Lambda\Sigma+\Sigma\Lambda\Sigma^2+\Lambda\Sigma^3$	$n^3+6n^2+21n+20$
	$\Sigma^2\Lambda^2+\Sigma\Lambda^2\Sigma+\Lambda\Sigma^2\Lambda+\Lambda^2\Sigma^2$	$n^2+5n+12$
	$\Sigma\Lambda\Sigma\Lambda+\Lambda\Sigma\Lambda\Sigma$	$n^2+5n+10$
	$\Sigma\Lambda^3+\Lambda\Sigma\Lambda^2+\Lambda^2\Sigma\Lambda+\Lambda^3\Sigma$	$n+3$
	Λ^4	1

Table 3c: Expansion of $\mathbb{E}[W^4]$. There are 30 monomials, but 15 bracelets in the expansion.

4 Computation

There remains the problem of how to compute the terms in the expansions given in equations (8) and (17). An ideal method would produce just the numerical coefficients in the expansions, since the other terms – both matrix and scalar polynomials in (Σ, Λ) – are known. Even better would be recursive relations for the coefficients, which seems a distinct possibility in view of the results given above. However, at the time of writing this is not possible, and the method we use produces the entire expansion. The required coefficients are extracted from this. It must be emphasised that this computation, although formidable, need only be done once for each k.

4.1 An expression for W^k

The starting point is to recognize that the (i, j) element of W^k can be written as a sum of products of k quadratic forms in normal random vectors. To see this, we let e_i be an n-vector of zeros except its i-th element is equal to one, and let $X = [x_1, \ldots, x_n]'$. Defining $A_{ij} = I_n \otimes (e_i e_j' + e_j e_i')/2$ and $x = \mathrm{vec}(X')$, we can write the (i, j) element of W as

$$W_{ij} = e_i' W e_j = \mathrm{tr}(e_j e_i' X' X) = \mathrm{vec}(X')' \mathrm{vec}(e_j e_i' X') = x'(I_n \otimes e_j e_i')x = x' A_{ij} x. \quad (23)$$

Note that $x \sim N_{mn}(\mu, I_n \otimes \Sigma)$, where $\mu = \mathrm{vec}(M')$ and $M = \mathbb{E}[X]$.

The (i, j) element of W^k may be written as

$$(W^k)_{ij} = \sum_{r_1=1}^{m} \sum_{r_2=1}^{m} \cdots \sum_{r_{k-1}=1}^{m} W_{i,r_1} W_{r_1,r_2} \cdots W_{r_{k-1},j}, \quad (24)$$

so is simply a sum of m^{k-1} products of k quadratic forms in x. This is the basis of our computational method. Since W is symmetric, there are only $h = m(m+1)/2$ unique elements in W. We will label the elements on and below the main diagonal of W as follows: $q_1 = W_{11},\ q_2 = W_{21},\ \ldots,\ q_m = W_{m1},\ q_{m+1} = W_{22},\ q_{m+2} = W_{32},\ \ldots,\ q_h = W_{mm}$. The terms in equation (24) are products of the type

$$q^\kappa \equiv q_1^{\kappa_1} q_2^{\kappa_2} \cdots q_h^{\kappa_h},$$

where the $\kappa = (\kappa_1, \ldots, \kappa_h)$ are sequences of non-negative integers of length h with $|\kappa| = \sum_{i=1}^h \kappa_i = k$ (i.e., weak compositions of k with h parts). Equivalently, they are products of the type $q_{i_1} q_{i_2} \cdots q_{i_k}$ with each i_j an integer from $(1, \ldots, h)$, repetitions allowed. Therefore, in order to evaluate $(W^k)_{ij}$ using (24), we first need to evaluate expectations of the type

$$\mathbb{E}[q^\kappa] \equiv \mathbb{E}[q_1^{\kappa_1} q_2^{\kappa_2} \cdots q_h^{\kappa_h}], \quad |\kappa| = k. \quad (25)$$

We will first describe a direct and efficient method for numerical evaluation of $\mathbb{E}[(W^k)_{ij}]$. However, at the time of writing this approach does not seem to be adaptable to produce the symbolic expansion discussed in previous sections. We therefore, in the following subsection, present an algorithm that produces the complete expansion for $\Psi_k(\Sigma, \Lambda)$ given in equation (15). This allows us - at least in principle - to obtain explicit expressions for the coefficients $c_\pi(\Sigma, \Lambda)$ in the expansion of $\mathbb{E}[W^k]$ for arbitrary k, as we have seen in Tables 3a–3c for $k = 2, 3, 4$.

4.2 Numerical Evaluation of $\mathbb{E}[W^k]$

There are a number of existing methods for computing the expectation of a product of k quadratic forms in normal random variables (see, for example, Kan (2008), or Bao and Ullah (2010)). Since we need to compute $\mathbb{E}[q^\kappa]$ for all $|\kappa| = k$, the most efficient method is probably due to Hillier, Kan, and Wang (2014), in which they present a super-short recursive algorithm for computing $\mathbb{E}[q^\kappa]$. However, the matrix A_{ij} in the quadratic form representation of W_{ij} is potentially a very large matrix ($mn \times mn$), so applying their algorithm (as well as other algorithms) directly would be computationally expensive. In the following Proposition, we adapt the algorithm for computing $\mathbb{E}[q^\kappa]$ in Hillier, Kan, and Wang (2014) to take advantage of the Kronecker structure of the A_{ij}. The proof of the following Proposition is given in the Appendix.

Proposition 3 *Let $q_i = x'(I_n \otimes B_i)x$, $i = 1, \ldots, h$, where $x \sim N_{mn}(\mathrm{vec}(M'), I_n \otimes \Sigma)$, and B_i, $i = 1, \ldots, h$ being $m \times m$ symmetric matrices. Define*

$$d_\kappa = \frac{\mathbb{E}[q^\kappa]}{2^k \kappa_1! \cdots \kappa_h!}. \tag{26}$$

Then d_κ can be obtained by using

$$d_\kappa = \frac{1}{2k}\left[n\mathrm{tr}(G_\kappa) + \mathrm{tr}(H_\kappa \tilde{\Lambda})\right], \tag{27}$$

where $\tilde{\Lambda} = \Sigma^{-\frac{1}{2}} M' M \Sigma^{-\frac{1}{2}}$, and G_κ and H_κ are two $m \times m$ matrices which can be obtained recursively using

$$G_\kappa = \sum_{\substack{1 \le i \le h,\\ \kappa_i > 0}} \tilde{B}_i \left[d_{\kappa_{(i)}} I_m + G_{\kappa_{(i)}}\right], \tag{28}$$

$$H_\kappa = G_\kappa + \sum_{\substack{1 \le i \le h,\\ \kappa_i > 0}} \tilde{B}_i H_{\kappa_{(i)}}, \tag{29}$$

with the boundary conditions of $d_0 = 1$, $G_0 = H_0 = 0_{m \times m}$. In the above expressions, $\kappa_{(i)} = (\kappa_1, \ldots, \kappa_{i-1}, \kappa_i - 1, \kappa_{i+1}, \ldots, \kappa_h)$ and $\tilde{B}_i = \Sigma^{\frac{1}{2}} B_i \Sigma^{\frac{1}{2}}$.

Two remarks on Proposition 3 are in order. The first is that the computational time of $\mathbb{E}[q^\kappa]$ depends on m but not n, which is a significant improvement over existing algorithms for computing $\mathbb{E}[q^\kappa]$. The second is that this efficient algorithm for computing $\mathbb{E}[q^\kappa]$ also allows us to compute the expectations of more complicated functions of W. For example, we can also compute $\mathbb{E}[WC_1WC_2\dots WC_k]$ for arbitrary matrices C_1 to C_k, or $\mathbb{E}[\prod_{i=1}^r \operatorname{tr}(W^i)^{\kappa_i}]$, where $\kappa_1 + 2\kappa_2 \dots + r\kappa_r = k$, because they can all be written as the expectation of a sum of products of elements of W. An example of the computation described in Proposition 3 for the case $k = 2$ is provided in Appendix B.

The approach based on Proposition 3 is suitable for computing $\mathbb{E}[W^k]$ numerically when m is small. However, when m is large, there will be many quadratic forms involved. In addition, we need to sum m^{k-1} terms $\mathbb{E}[q^\kappa]$, and when k is large this would also be computationally expensive. As a result, if one would like to compute $\mathbb{E}[W^k]$ numerically for large m, it is much more efficient to use the expansion for $\mathbb{E}[W^k]$. That is, the coefficients $c_\pi(\Sigma, \Lambda)$ in equation (17) are crucial, and we take up the task of computing them in the next subsection.

4.3 Analytical Expression for $\mathbb{E}[W^k]$

For $q_i = x'(I_n \otimes B_i)x$, $i = 1, \dots, h$, where $x \sim N_{mn}(\mu, I_n \otimes \Sigma)$ and $\mu = \operatorname{vec}(M')$, we can obtain an analytical (i.e, symbolic) expression for the expectation of a product of k of the q_i, say $\mathbb{E}[q_1q_2\dots q_k]$. Such an expression can be obtained, for example, by using the recursion given in Theorem 1 of Bao and Ullah (2010). The terms in this expression can be written as a sum of various products of the quantities $\tau_{i_1,\dots,i_r}$ and $\theta_{i_1,\dots,i_s}$, which are defined by

$$\begin{aligned}
\tau_{i_1,\dots,i_r} &= \operatorname{tr}((I_n \otimes B_{i_1})(I_n \otimes \Sigma) \cdots (I_n \otimes B_{i_r})(I_n \otimes \Sigma)) \\
&= n\operatorname{tr}(B_{i_1}\Sigma B_{i_2}\Sigma \cdots B_{i_r}\Sigma), && (30)\\
\theta_{i_1,\dots,i_s} &= \mu'(I_n \otimes B_{i_1})(I_n \otimes \Sigma) \cdots (I_n \otimes \Sigma)(I_n \otimes B_{i_s})\mu \\
&= \operatorname{tr}(B_{i_1}\Sigma B_{i_2}\Sigma \cdots \Sigma B_{i_s}\Lambda), && (31)
\end{aligned}$$

with $\Lambda = M'M$. As an example, the following expansions are obtained:

$$\begin{aligned}
\mathbb{E}[q_1q_2] &= 4\theta_{1,2} + 2\tau_{1,2} + \tau_1\tau_2 + \tau_1\theta_2 + \tau_2\theta_1 + \theta_1\theta_2, && (32)\\
\mathbb{E}[q_1q_2q_3] &= \tau_1\tau_2\tau_3 + 2(\tau_1\tau_{2,3} + \tau_2\tau_{1,3} + \tau_3\tau_{1,2}) + 8\tau_{1,2,3} \\
&\quad + (\tau_1\tau_2\theta_3 + \tau_1\tau_3\theta_2 + \tau_2\tau_3\theta_1) + 2(\tau_{1,2}\theta_3 + \tau_{1,3}\theta_2 + \tau_{2,3}\theta_1) \\
&\quad + 4(\tau_1\theta_{2,3} + \tau_2\theta_{1,3} + \tau_3\theta_{1,2}) + (\tau_1\theta_2\theta_3 + \tau_2\theta_1\theta_3 + \tau_3\theta_1\theta_2) \\
&\quad + \theta_1\theta_2\theta_3 + 4(\theta_1\theta_{2,3} + \theta_2\theta_{1,3} + \theta_3\theta_{1,2}) + 8(\theta_{1,2,3} + \theta_{1,3,2} + \theta_{2,1,3}). && (33)
\end{aligned}$$

The first step in our algorithm (for given k) is to generate an expansion like this for each term in equation (24). Note that in each term, each of the subscripts on the q_i (an integer from $(1, \dots, h)$) appears exactly once in the subscripts of τ and θ.

The expansion consists of all terms that can be constructed with this property. In applying this Theorem to our case we have, as described earlier, $W_{ij} = x'(I_n \otimes B_{ij})x$, where $B_{ij} = (e_i e_j' + e_j e_i')/2$. Since W_{ij} and the corresponding B_{ij} are indexed by subscript-pairs (i, j), we modify the notation for the $\theta's$ and $\tau's$ as follows:

$$\tau_{(i_1,j_1),\dots,(i_r,j_r)} = n\mathrm{tr}(B_{i_1,j_1}\Sigma B_{i_2,j_2}\Sigma \cdots B_{i_r,j_r}\Sigma), \tag{34}$$

$$\theta_{(i_1,j_1),\dots,(i_s,j_s)} = \mathrm{tr}(B_{i_1,j_1}\Sigma B_{i_2,j_2}\Sigma \cdots B_{i_s,j_s}\Lambda). \tag{35}$$

Before we describe the general algorithm, it is instructive to see how this algorithm is obtained by working through the steps for the case $k = 2$. For $k = 2$, we use (32) to obtain

$$\begin{aligned}\mathbb{E}[(W^2)_{ij}] &= \sum_{r_1=1}^{m} \mathbb{E}[W_{i,r_1} W_{r_1,j}] \\ &= \sum_{r_1=1}^{m} [4\theta_{(i,r_1),(r_1,j)} + 2\tau_{(i,r_1),(r_1,j)} + \tau_{(i,r_1)}\tau_{(r_1,j)} \\ &\quad + \tau_{(i,r_1)}\theta_{(r_1,j)} + \tau_{(r_1,j)}\theta_{(i,r_1)} + \theta_{(i,r_1)}\theta_{(r_1,j)}],\end{aligned} \tag{36}$$

Writing $\Sigma = \{\sigma_{ij}\}$ and $\Lambda = \{\lambda_{ij}\}$, the terms that appear in these expressions can be written in terms of the σ_{ij} 's and λ_{ij}'s. For the example we have:

$$\begin{aligned}\theta_{(i,r_1),(r_1,j)} &= \mathrm{tr}\left(\left(\frac{e_i e_{r_1}' + e_{r_1} e_i'}{2}\right)\Sigma\left(\frac{e_{r_1} e_j' + e_j e_{r_1}'}{2}\right)\Lambda\right) \\ &= \frac{1}{4}(\sigma_{r_1,r_1}\lambda_{i,j} + \sigma_{r_1,j}\lambda_{r_1,i} + \sigma_{i,r_1}\lambda_{j,r_1} + \sigma_{i,j}\lambda_{r_1,r_1}),\end{aligned} \tag{37}$$

$$\begin{aligned}\tau_{(i,r_1),(r_1,j)} &= n\mathrm{tr}\left(\left(\frac{e_i e_{r_1}' + e_{r_1} e_i'}{2}\right)\Sigma\left(\frac{e_{r_1} e_j' + e_j e_{r_1}'}{2}\right)\Sigma\right) \\ &= \frac{n}{4}(\sigma_{r_1,r_1}\sigma_{i,j} + \sigma_{r_1,j}\sigma_{r_1,i} + \sigma_{i,r_1}\sigma_{j,r_1} + \sigma_{i,j}\sigma_{r_1,r_1}) \\ &= \frac{n}{2}(\sigma_{r_1,r_1}\sigma_{i,j} + \sigma_{i,r_1}\sigma_{r_1,j}),\end{aligned} \tag{38}$$

$$\tau_{(i,r_1)} = n\mathrm{tr}\left(\left(\frac{e_i e_{r_1}' + e_{r_1} e_i'}{2}\right)\Sigma\right) = n\sigma_{i,r_1}, \tag{39}$$

$$\tau_{(r_1,j)} = n\mathrm{tr}\left(\left(\frac{e_{r_1} e_j' + e_j e_{r_1}'}{2}\right)\Sigma\right) = n\sigma_{j,r_1},$$

$$\theta_{(i,r_1)} = \mathrm{tr}\left(\left(\frac{e_i e_{r_1}' + e_{r_1} e_i'}{2}\right)\Lambda\right) = \lambda_{i,r_1}, \tag{40}$$

$$\theta_{(r_1,j)} = \mathrm{tr}\left(\left(\frac{e_{r_1} e_j' + e_j e_{r_1}'}{2}\right)\Lambda\right) = \lambda_{j,r_1}. \tag{41}$$

Putting these terms together, we obtain

$$\mathbb{E}[(W^2)_{ij}] = \sum_{r_1=1}^{m} (\sigma_{r_1,r_1}\lambda_{ij} + \sigma_{r_1,j}\lambda_{r_1,i} + \sigma_{i,r_1}\lambda_{j,r_1} + \sigma_{ij}\lambda_{r_1,r_1})$$

$$
\begin{aligned}
&+ n\sum_{r_1=1}^{m}(\sigma_{r_1,r_1}\sigma_{ij} + \sigma_{i,r_1}\sigma_{r_1,j}) \\
&+ n^2\sum_{r_1=1}^{m}\sigma_{i,r_1}\sigma_{r_1,j} + n\sum_{r_1=1}^{m}(\sigma_{i,r_1}\lambda_{r_1,j} + \lambda_{i,r_1}\sigma_{r_1,j}) + \sum_{r_1=1}^{m}\lambda_{i,r_1}\lambda_{r_1,j} \\
= \; &\mathrm{tr}(\Sigma)\lambda_{ij} + e_i'[\Lambda\Sigma + \Sigma\Lambda]e_j + \mathrm{tr}(\Lambda)\sigma_{ij} + n\mathrm{tr}(\Sigma)\sigma_{ij} + ne_i'\Sigma^2 e_j \\
&+ n^2 e_i'\Sigma^2 e_j + ne_i'[\Sigma\Lambda + \Lambda\Sigma]e_j + e_i'\Lambda^2 e_j. \qquad (42)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{E}[W^2] &= \mathrm{tr}(\Sigma)\Lambda + (\Lambda\Sigma + \Sigma\Lambda) + \mathrm{tr}(\Lambda)\Sigma + n\mathrm{tr}(\Sigma)\Sigma + n\Sigma^2 \\
&\quad + n^2\Sigma^2 + n(\Sigma\Lambda + \Lambda\Sigma) + \Lambda^2 \\
&= \mathrm{tr}(\Sigma)\Lambda + (n+1)(\Sigma\Lambda + \Lambda\Sigma) + [n\mathrm{tr}(\Sigma) + \mathrm{tr}(\Lambda)]\Sigma + n(n+1)\Sigma^2 + \Lambda^2. \qquad (43)
\end{aligned}
$$

Picking out the coefficients in the above expression, we obtain the $c_\pi(\Sigma, \Lambda)$ as reported in Table 3a.

For general k, we need an algorithm to automate the above process. We first need an explicit expression for the expectation of a product of k quadratic forms $q_{i_1}q_{i_2}\cdots q_{i_k}$, and have developed a computer program for generating the terms in this expression for arbitrary k. This program is a modified version of the recursion given in Theorem 1 of Bao and Ullah (2010). Each term is a product of various τ's and θ's, with each integer in the set $\{i_1, i_2, \cdots, i_k\}$ appearing exactly once in the subscripts.

We next need to express these τ's and θ's in terms of the elements of Σ and Λ, and then sum over the $k-1$ r_i's. To do so we proceed as follows: for the term indexed by $(i, r_1, r_2, \ldots r_{k-1}, j)$ in equation (24), we represent this by a $2k$-tuple $(0, 1, 1, 2, 2, \ldots, k-1, k-1, k)$, where 0 stands for i, 1 stands for $r_1, \ldots, k-1$ stands for r_{k-1}, and k stands for j. We then break the $2k$-tuple into disjoint subsets, where the number of subsets is determined by the number of elements (τ's and θ's) in that term, and membership of each subset is based on the subscripts of the τ and θ terms. We illustrate this with an example.

Example 2 *When $k = 4$, one of the terms in the expansion of $\mathbb{E}[q_1q_2q_3q_4]$ is $8\tau_{14}\theta_{23}$. Since there are two terms in this product, we will have two subsets, one corresponding to τ_{14} and the other corresponding to θ_{23}. The first subset corresponds to the first and fourth pairs in the sequence (01122334) and it is (0134), the second subset corresponds to the second and third pairs in the sequence and it is (1223). Therefore, for the term $8\tau_{14}\theta_{23}$, we break (01122334) into $(0134)(1223)$. Then, since the corresponding τ and θ are the traces of products, we can cycle the first element to the last in each subset, which then gives $(1340)(2231)$. For a term that involves τs there will be a factor n attached for each occurrence of τ, and for the terms that involve θ, we adopt the convention that the last pair corresponds to λ. Based on this convention, the term $(1340)(2231)$ will produce*

$$8n\sigma_{r_1,r_3}\sigma_{j,i}\sigma_{r_2,r_2}\lambda_{r_3,r_1} = 8n\sigma_{i,j}\sigma_{r_1,r_3}\lambda_{r_3,r_1}\sigma_{r_2,r_2}.$$

After summing each r_i from 1 to m, $i = 1, 2, 3,$ a term of this type will contribute

$$8ne_i'\Sigma e_j \mathrm{tr}(\Sigma\Lambda)\mathrm{tr}(\Sigma)$$

to the overall expectation.

For the general case, the result of summing over each r_i, $i = 1, .., k-1$, can be obtained by constructing cycles from the k pairs of numbers in the final sequence. We start with the pair that has an element of 0. If the other element in that pair is k, then we are done with constructing the first cycle. If not, then we use the value of the other element in the first pair to find the other pair that has the same element. There is exactly one, since, except for 0 and k, all other elements appear twice in the sequence. If the pair has k in it, then we are done. If not, we continue the process by using the other element in the pair to locate the next pair. We keep this process going on until we come to a pair that has k in it. From this first cycle, we identify the (i, j) element of a matrix which is a product of Σ and Λ, which corresponds to the pairs in the cycle (each pair is either an element of Σ or Λ). After the first cycle is identified, we search the smallest index in the remaining pairs and continue to find another cycle that ends with the same index. The second and subsequent cycles give us the trace terms.

In the above example the first cycle comes from the second pair, (40), and contributes an element of Σ. The second cycle starts with 1 and ends with 1, which consists of the first and fourth pairs in the sequence. As the first pair is an element of Σ and the fourth pair is an element of Λ, we obtain $\mathrm{tr}(\Sigma\Lambda)$ from the second cycle. Finally, the last cycle starts with 2 and ends with 2, which is obtained by using the third pair in the sequence and this cycle gives us $\mathrm{tr}(\Sigma)$.

By going through the above process, we create a typical term in $\mathbb{E}[(W^k)_{ij}]$. However, the matrices in the product of k quadratic forms involve terms like $(e_i e_{r_1}' + e_{r_1} e_i')/2$, $(e_{r_1} e_{r_2}' + e_{r_2} e_{r_1}')/2$ and so on. Therefore, to generate all terms in the expectation, we need to include not just the sequence $(0, 1, 1, 2, 2, \ldots, k-1, k-1, k)$, but also all sequences that can be obtained from it by interchanging the elements in each of the k pairs, resulting in altogether 2^k sequences. We repeat the same exercise as above and divide the sum of all the terms by 2^k. In the example for $k = 4$, we need to consider also (01122343), (01123234) all the way to (10213243), resulting in 16 sequences. With each sequence, we repeat the same process and sum the results across the 16 sequences, then divide the sum by 16.

We can therefore summarize our algorithm for obtaining the symbolic expression expansion of $\mathbb{E}[W^k]$ as follows:

1. Using the program mentioned above, generate an expression for the expected value of each product of k quadratic forms that occurs in equation (24) in terms of τ's and θ's.

2. For each term in the explicit expression, break the $2k$-tuple $(0, 1, 1, 2, 2, \ldots, k-1, k-1, k)$ into a number of subsequences, depending on how many elements (τ's and θ's) are in each term. The length of each subsequence is twice the number of subscripts in τ or θ, and the pairs in the subsequence are indicated by the subscripts of τ's and θ's.

3. For each subsequence, move the first element to the last, group them into pairs. Each pair represents an element in Σ with the exception that the last pair for each θ actually represents an element in Λ. Multiply the product by n for each occurrence of τ.

4. Take this product of elements of Σ and Λ and construct cycles from their indices. The cycle that begins with 0 and ends with k represents a matrix. The other cycles represent traces.

5. Take the original $2k$-tuple, consider altogether 2^k possible combinations of reshuffling the elements in each pair, and repeat the exercise. Add all the terms up and divide the final answer by 2^k.

To make the above steps a little easier to follow, we work out the first four steps for the case of $k = 2$ in the following table, which corresponds to the $2k$-tuple of $(0, 1, 1, 2)$.

1	Coefficient	4	2	1	1	1	1
	Term	$\theta_{1,2}$	$\tau_{1,2}$	$\tau_1\tau_2$	$\tau_1\theta_2$	$\tau_2\theta_1$	$\theta_1\theta_2$
2	Coefficient	4	2	1	1	1	1
	Term	(0,1,1,2)	(0,1,1,2)	(0,1)(1,2)	(0,1)(1,2)	(1,2)(0,1)	(0,1)(1,2)
3	Coefficient	4	$2n$	n^2	n	n	1
	Term	(1,1,2,0)	(1,1,2,0)	(1,0)(2,1)	(1,0)(2,1)	(2,1)(1,0)	(1,0)(2,1)
4	Coefficient	4	$2n$	n^2	n	n	1
	Term	$(\Sigma)\Lambda$	$(\Sigma)\Sigma$	Σ^2	$\Sigma\Lambda$	$\Lambda\Sigma$	Λ^2

Table 4: Steps to obtain $\mathbb{E}[W^2]$.

It follows that for the first $2k$-tuple of $(0, 1, 1, 2)$, the sum of the six terms is

$$4(\Sigma)\Lambda + 2n(\Sigma)\Sigma + n^2\Sigma^2 + n\Sigma\Lambda + n\Lambda\Sigma + \Lambda^2.$$

For the other three $2k$-tuples, the corresponding expression is

$$\begin{aligned}
(1, 0, 1, 2) &: \ 4\Sigma\Lambda + 2n\Sigma^2 + n^2\Sigma^2 + n\Sigma\Lambda + n\Lambda\Sigma + \Lambda^2 \\
(0, 1, 2, 1) &: \ 4\Lambda\Sigma + 2n\Sigma^2 + n^2\Sigma^2 + n\Sigma\Lambda + n\Lambda\Sigma + \Lambda^2 \\
(1, 0, 2, 1) &: \ 4(\Lambda)\Sigma + 2n(\Sigma)\Sigma + n^2\Sigma^2 + n\Sigma\Lambda + n\Lambda\Sigma + \Lambda^2
\end{aligned}$$

Adding these four expressions together and dividing the sum by four, we obtain the results in Table 3a:

$$\begin{aligned}\mathbb{E}[W^2] &= (\Sigma)\Lambda + (\Lambda)\Sigma + \Sigma\Lambda + \Lambda\Sigma + n(\Sigma)\Sigma + n\Sigma^2 + n^2\Sigma^2 + n\Sigma\Lambda + n\Lambda\Sigma + \Lambda^2 \\ &= (\Sigma)\Lambda + [(\Lambda) + n(\Sigma)]\Sigma + n(n+1)\Sigma^2 + (n+1)(\Sigma\Lambda + \Lambda\Sigma) + \Lambda^2, \end{aligned} \tag{44}$$

We have implemented this algorithm and computed the $c_\pi(\Sigma, \Lambda)$ for $\mathbb{E}[W^k]$ for $k \leq 10$. The coefficients as well as the Matlab programs are available upon request. For $k > 10$, the program can still handle it but will take a very long time to run because the expression for the expectation of a product of k quadratic forms already has 63,673,506 distinct terms when $k = 10$, and it goes up to 835,724,952 distinct terms when $k = 11$. Therefore, for $k > 10$, the current method is not practical and a method for determining $c_\pi(\Sigma, \Lambda)$ in a recursive way would be preferable.[6]

5 Concluding Comments

By exploiting the homogeneity and group-theoretic properties of the mapping defining the moments of both a central and a noncentral Wishart matrix, and invoking some classical results on matrix invariants, we have shown that the expressions for the moments of a Wishart matrix have a well-defined structure. This structure is reasonably simple to describe, even in the noncentral case, but it is challenging to compute because the dimension of the problem increases rapidly with the degree, k, of the moment of interest. Nevertheless, we have been able to construct a program to compute the full expansions that, at least in principle, is valid for any k.

The structure results do, however, highlight the fact that the only unknowns in the expressions for the moments are certain purely numerical coefficients – polynomials in the degrees of freedom, n. It would be highly desirable that the computations could be focused purely on those, perhaps being based on an at-present unknown recursive scheme. At the time of writing our understanding of these coefficients is not sufficient to permit this, and work continues on that challenge.

[6] Kuriki and Numata (2010, Theorem 1) presents an explicit expression for the expectation of a product of elements from a noncentral Wishart matrix. Their results can also be used to obtain $\mathbb{E}[W^k]$. However, their expression requires adding up a lot of terms even when k is modestly large. For example, when $k = 10$, their expression has 23,758,664,096 terms and it is less efficient to use than our method.

APPENDIX A: Proof of Proposition 3

Let $Z = X\Sigma^{-\frac{1}{2}}$, and $z = \mathrm{vec}(Z') = (I_n \otimes \Sigma^{-\frac{1}{2}})x$. We can then write

$$q_i = x'(I_n \otimes B_i)x = z'(I_n \otimes \tilde{B}_i)z, \tag{45}$$

where $\tilde{B}_i = \Sigma^{\frac{1}{2}} B_i \Sigma^{\frac{1}{2}}$. It is easy to see that $z \sim N_{mn}(\tilde{\mu}, I_{mn})$, where $\tilde{\mu} = \mathrm{vec}(\Sigma^{-\frac{1}{2}} M')$. Hillier, Kan, and Wang (2014) provided a recursive algorithm for computing

$$d_\kappa = \frac{\mathbb{E}[q^\kappa]}{2^k \kappa_1! \cdots \kappa_h!}. \tag{46}$$

Specifically, they show that

$$d_\kappa = \frac{1}{2k}\left[\mathrm{tr}(\tilde{G}_\kappa) + \tilde{\mu}' h_\kappa\right], \tag{47}$$

where

$$\tilde{G}_\kappa = \sum_{\substack{1 \le i \le h, \\ \kappa_i > 0}} (I_n \otimes \tilde{B}_i)\left[d_{\kappa_{(i)}} I_{nm} + \tilde{G}_{\kappa_{(i)}}\right], \tag{48}$$

$$h_\kappa = \tilde{G}_\kappa \tilde{\mu} + \sum_{\substack{1 \le i \le h, \\ \kappa_i > 0}} (I_n \otimes \tilde{B}_i) h_{\kappa_{(i)}}, \tag{49}$$

with the boundary conditions of $d_0 = 1$, $\tilde{G}_0 = 0_{nm \times nm}$ and $h_0 = 0_{nm}$. Given the Kronecker structure of the matrices in the quadratic forms in z, we can write

$$\tilde{G}_\kappa = I_n \otimes G_\kappa. \tag{50}$$

and G_κ can be obtained recursively using

$$G_\kappa = \sum_{\substack{1 \le i \le h, \\ \kappa_i > 0}} \tilde{B}_i \left[d_{\kappa_{(i)}} I_m + G_{\kappa_{(i)}}\right]. \tag{51}$$

Using the fact that

$$\begin{aligned}
\mathrm{tr}(H_\kappa \tilde{\Lambda}) &= \mathrm{tr}(H_\kappa \Sigma^{-\frac{1}{2}} M' M \Sigma^{-\frac{1}{2}}) \\
&= \mathrm{vec}(\Sigma^{-\frac{1}{2}} M')'(I_n \otimes H_\kappa)\mathrm{vec}(\Sigma^{-\frac{1}{2}} M') \\
&= \tilde{\mu}'(I_n \otimes H_\kappa)\tilde{\mu},
\end{aligned} \tag{52}$$

we can define $h_\kappa = (I_n \otimes H_\kappa)\tilde{\mu}$ and obtain a recurrence relation for H_κ using

$$H_\kappa = G_\kappa + \sum_{\substack{1 \le i \le h, \\ \kappa_i > 0}} \tilde{B}_i H_{\kappa_{(i)}}. \tag{53}$$

Finally, using G_κ and H_κ, d_κ can be obtained by using

$$d_\kappa = \frac{1}{2k}\left[n\mathrm{tr}(G_\kappa) + \mathrm{tr}(H_\kappa \tilde{\Lambda})\right]. \tag{54}$$

This completes the proof.

APPENDIX B: Example for Proposition 3

To illustrate how Proposition 3 works, we consider an example of obtaining $\mathbb{E}[W^2]$ for $m = 2$. For this case, W has 3 unique elements: $q_1 = W_{11}$, $q_2 = W_{21}$ and $q_3 = W_{22}$. The unique elements of W^2 are

$$e_1' W^2 e_1 = \sum_{r_1=1}^{2} W_{1,r_1} W_{r_1,1} = q_1^2 + q_2^2, \tag{55}$$

$$e_1' W^2 e_2 = \sum_{r_1=1}^{2} W_{1,r_1} W_{r_1,2} = q_1 q_2 + q_2 q_3, \tag{56}$$

$$e_2' W^2 e_2 = \sum_{r_1=1}^{2} W_{2,r_1} W_{r_1,2} = q_2^2 + q_3^2. \tag{57}$$

Proposition 3 can then be applied to obtain $\mathbb{E}[q_1^2]$, $\mathbb{E}[q_2^2]$, $\mathbb{E}[q_3^2]$, $\mathbb{E}[q_1 q_2]$, and $\mathbb{E}[q_2 q_3]$. To avoid repetition, we only work out $\mathbb{E}[q_1^2]$ and $\mathbb{E}[q_1 q_2]$ because the calculations of the other terms are similar, and they can also be obtained by symmetry.

In order to compute $\mathbb{E}[q_1^2]$ and $\mathbb{E}[q_1 q_2]$, we use Proposition 3 to first obtain $d_{(2,0)} = \mathbb{E}[q_1^2]/8$ and $d_{(1,1)} = \mathbb{E}[q_1 q_2]/4$, where

$$d_{(2,0)} = \frac{n\mathrm{tr}(G_{(2,0)}) + \mathrm{tr}(H_{(2,0)}\tilde{\Lambda})}{4}, \tag{58}$$

$$d_{(1,1)} = \frac{n\mathrm{tr}(G_{(1,1)}) + \mathrm{tr}(H_{(1,1)}\tilde{\Lambda})}{4}. \tag{59}$$

Using the fact that $B_1 = e_1 e_1'$ and $B_2 = (e_1 e_2' + e_2 e_1')/2$, we can now use Proposition 3 to recursively obtain $G_{(2,0)}$, $G_{(1,1)}$, $H_{(2,0)}$, and $H_{(1,1)}$:

$$
\begin{aligned}
G_{(1,0)} &= \tilde{B}_1 d_{(0,0)} I_2 = \Sigma^{\frac{1}{2}} e_1 e_1' \Sigma^{\frac{1}{2}}, && (60)\\
H_{(1,0)} &= G_{(1,0)}, && (61)\\
d_{(1,0)} &= \frac{1}{2}[n\mathrm{tr}(G_{(1,0)}) + \mathrm{tr}(H_{(1,0)}\tilde{\Lambda})] = \frac{n e_1' \Sigma e_1 + e_1' \Lambda e_1}{2}, && (62)\\
G_{(0,1)} &= \tilde{B}_2 d_{(0,0)} I_2 = \frac{\Sigma^{\frac{1}{2}}(e_1 e_2' + e_2 e_1')\Sigma^{\frac{1}{2}}}{2}, && (63)\\
H_{(0,1)} &= G_{(0,1)}, && (64)\\
d_{(0,1)} &= \frac{1}{2}[n\mathrm{tr}(G_{(0,1)}) + \mathrm{tr}(H_{(0,1)}\tilde{\Lambda})] \ = \frac{n e_1' \Sigma e_2 + e_1' \Lambda e_2}{2}, && (65)\\
G_{(2,0)} &= \tilde{B}_1[d_{(1,0)} I_2 + G_{(1,0)}] \\
&= (d_{(1,0)} + e_1' \Sigma e_1)\Sigma^{\frac{1}{2}} e_1 e_1' \Sigma^{\frac{1}{2}}, && (66)\\
H_{(2,0)} &= G_{(2,0)} + \tilde{B}_1 H_{(1,0)} = G_{(2,0)} + \Sigma^{\frac{1}{2}} e_1 e_1' \Sigma^{\frac{1}{2}} G_{(1,0)} && (67)
\end{aligned}
$$

$$
\begin{aligned}
G_{(1,1)} &= \tilde{B}_1[d_{(0,1)}I_2 + G_{(0,1)}] + \tilde{B}_2[d_{(1,0)}I_2 + G_{(1,0)}] \\
&= d_{(0,1)}\Sigma^{\frac{1}{2}}e_1e_1'\Sigma^{\frac{1}{2}} + \frac{e_1'\Sigma e_1\Sigma^{\frac{1}{2}}e_1e_2'\Sigma^{\frac{1}{2}} + e_1'\Sigma e_2\Sigma^{\frac{1}{2}}e_1e_1'\Sigma^{\frac{1}{2}}}{2} \\
&\quad + d_{(1,0)}\Sigma^{\frac{1}{2}}\frac{(e_1e_2' + e_2e_1')}{2}\Sigma^{\frac{1}{2}} + \frac{e_1'\Sigma e_2\Sigma^{\frac{1}{2}}e_1e_1'\Sigma^{\frac{1}{2}} + e_1'\Sigma e_1\Sigma^{\frac{1}{2}}e_2e_1'\Sigma^{\frac{1}{2}}}{2}, \qquad (68) \\
H_{(1,1)} &= G_{(1,1)} + \tilde{B}_1H_{(0,1)} + \tilde{B}_2H_{(1,0)} = G_{(1,1)} + \tilde{B}_1G_{(0,1)} + \tilde{B}_2G_{(1,0)}. \qquad (69)
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{E}[q_1^2] &= 2n\mathrm{tr}(G_{(2,0)}) + 2\mathrm{tr}(H_{(2,0)}\tilde{\Lambda}) \\
&= (ne_1'\Sigma e_1 + e_1'\Lambda e_1)^2 + 2e_1'\Sigma e_1(ne_1'\Sigma e_1 + 2e_1'\Lambda e_1), \qquad (70) \\
\mathbb{E}[q_1q_2] &= n\mathrm{tr}(G_{(1,1)}) + \mathrm{tr}(H_{(1,1)}\tilde{\Lambda}) \\
&= (ne_1'\Sigma e_1 + e_1'\Lambda e_1)(ne_1'\Sigma e_2 + e_1'\Lambda e_2) \\
&\quad + 2(ne_1'\Sigma e_1e_1'\Sigma e_2 + e_1'\Lambda e_1e_1'\Sigma e_2 + e_1'\Lambda e_2e_1'\Sigma e_1). \qquad (71)
\end{aligned}
$$

REFERENCES

Bao, Yong, and Ullah, Aman (2010) Expectation of quadratic forms in normal and nonnormal variables with applications, *Journal of Statistical Planning and Inference* 140, 1193–1205.

Davis, A. W. (1980) Invariant polynomials with two matrix arguments, extending the zonal polynomials, in *Multivariate Analysis–V* (ed. P. R. Krishnaiah), 287–299.

de Waal, D. J., and Nel, D. G. (1973) On some expectations with respect to Wishart matrices, *South African Statistical Journal* 7, 61–67.

Di Nardo, Elvira (2014) On a symbolic representation of non-central Wishart random matrices with applications, *Journal of Multivariate Analysis* 135, 121–135.

Hillier, Grant, Kan, Raymond, and Wang, Xiaolu (2014) Generating functions and short recursions, with application to the moments of quadratic forms in noncentral normal vectors, *Econometric Theory* 30, 436–473.

Kan, Raymond (2008) From moments of sum to moments of product, *Journal of Multivariate Analysis* 99, 542–554.

Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.

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

Gupta, A. K., and Nagar, N. K. (1999), *Matrix Variate Distributions*, Chapman & Hall, New York.

Kuriki, Satoshi, and Numata, Yasuhide (2010) Graph presentations for moments of noncentral Wishart and their applications, *Annals of the Institute of Statistical Mathematics* 62, 645–672.

Letac, Gérard, and Massam, Hélène (2004) All invariant moments of the Wishart distribution, *Scandinavian Journal of Statistics* 31, 295–318.

Letac, Gérard, and Massam, Hélène (2008) The noncentral Wishart as an exponential family, and its moments, *Journal of Multivariate Analysis* 99, 1393–1417.

Magnus, Jan R., and Neudecker, R. (1979) The commutation matrix: some properties and applications, *Annals of Statistics* 2, 381–394.

Neudecker, Heinz, and Wansbeek, Tom (1987) Fourth-order properties of normally distributed random matrices, *Linear Algebra and its Applications* 97, 13–21.