Scand J Statist 15: 97–109, 1988

Moments for the Inverted Wishart Distribution

DIETRICH VON ROSEN

University of Stockholm

ABSTRACT. In order to obtain moments for the inverted Wishart distribution two new approaches are brought forward. One is applied to moments of order up to three whereas another, bearing less information, is applied to moments of arbitrary order. Expressions for the moments of arbitrary order are given in a recursive and non-recursive manner. As an application the growth curve model is considered. Results are given using the third order moments for the inverted Wishart distribution.

Key words: inverted Wishart distribution, inverse matrix moments, growth curve model

1. Introduction

The inverted Wishart distribution (IW) has various applications in statistics. For instance, the distribution can be used as a natural conjugate prior when discussing covariances under sampling from the normal distribution (Anderson, 1984, section 7.7). Hence, moments for the IW are of importance. Moreover, moments for the IW have also been utilized in discriminant analysis (Das Gupta, 1968; Siskind, 1972; Haff, 1982), when obtaining moments of the maximum likelihood estimators in the growth curve model (von Rosen, 1985, and section 5 of this paper) as well as in several other fields of statistics.

Useful results for the IW can be found in Press (1982). Several different techniques have been presented when deriving the first and second order moments for the IW. Kaufman (1967) derived the moments by aid of a factorization theorem. Das Gupta (1968) utilized some invariance arguments. Both these authors based their calculations on the inverse moments of chi-squared variables. In a series of papers Haff (1977, 1979a, b, 1980, 1981, 1982) presented moment identities which among others are useful when obtaining moments for the IW. The identities were established by applying Stokes' theorem. Independently of Haff, von Rosen (1985) discovered that moments for the IW where obtainable with the help of the fundamental theorem in calculus. Hence von Rosen's approach is almost identical to Haff's. For moments of higher order than the second, however, Haff's approach needs further explication. To some extent this was done by von Rosen (1985) who considered several algebraic aspects. In this paper we intend to present, clarify, and extend the results of von Rosen and point out serious problems when discussing moments of general order.

2. Preparation

This section is devoted to some basic tools needed in the sequel. The cummutation matrix $\mathbf{K}_{p,q}$ given by

$$\mathbf{K}_{p,q} = \sum_{I} \mathbf{d}_i \mathbf{e}_j' \otimes \mathbf{e}_j \mathbf{d}_i' \tag{2.1}$$

is of utmost importance where in the definition $\otimes$ is the right Kronecker product, defined by $\mathbf{A}\otimes\mathbf{B} = (a_{ij}B)$ for $\mathbf{A} = (a_{ij})$, $I = \{i,j;\ 1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant q\}$ and $\mathbf{d}_i: p \times 1$, $\mathbf{e}_j: q \times 1$ stand for vectors with 1 in the ith and jth positions, respectively, and zero elsewhere. We assume that the reader is familiar with the properties of $\mathbf{K}_{p,q}$ and for useful results is referred to Magnus & Neudecker (1979). Throughout the report lower cases with subscripts will denote vectors like those in (2.1).

In multivariate analysis matrix derivatives, in a sense of a collection of partial derivatives, play a key role. Here we use the following definition:

Let $\mathbf{Y}: q \times r$ be a function of $\mathbf{X}: p \times n$. If $\mathbf{X}$ is symmetric but otherwise consists of functionally independent elements then

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}} = \sum_I \frac{\partial y_{ij}}{\partial x_{kl}} (\mathbf{d}_i \mathbf{e}_j') \otimes \varepsilon_{kl} (\mathbf{f}_k \mathbf{g}_l') \qquad \varepsilon_{kl} = \begin{cases} 1 & k = l \\ \frac{1}{2} & k \neq l \end{cases} \tag{2.2}$$

where $I = \{i, j, k, l;\ 1 \leqslant i \leqslant q,\ 1 \leqslant j \leqslant r,\ 1 \leqslant k \leqslant p,\ 1 \leqslant l \leqslant n\}$, $\mathbf{d}_i: q \times 1$, $\mathbf{e}_j: r \times 1$, $\mathbf{f}_k: p \times 1$ and $\mathbf{g}_l: n \times 1$. Haff (1981, p. 217) used a transposition of $\partial \mathbf{Y}/\partial \mathbf{X}$, namely $\mathbf{K}_{p,q}(\partial \mathbf{Y}/\partial \mathbf{X})\mathbf{K}_{r,n}$. In the next lemma some elementary rules for our derivative are given. Vec is the function defined by

$$\text{vec}: R^{n \times m} \to R^{nm},\ \mathbf{x}\mathbf{y}' \to \mathbf{y} \otimes \mathbf{x} \text{ where } \mathbf{x}: n \times 1 \text{ and } \mathbf{y}: m \times 1.$$

Lemma 2.1

Let $\mathbf{X}$: $p \times p$ be a symmetric matrix whose elements are otherwise functionally independent.

(a) *Let $\mathbf{Y}: q \times r$, $\mathbf{Z}: s \times q$ and y a scalar be functions of $\mathbf{X}$ and $\mathbf{A}: s \times q$, $\mathbf{B}: r \times t$ matrices of constants. Then*

(i) $$\frac{\partial \overleftarrow{\mathbf{ZY}}}{\partial \mathbf{X}} = \frac{\partial \mathbf{Z}}{\partial \mathbf{X}} (\mathbf{Y} \otimes \mathbf{I}_p) + (\mathbf{Z} \otimes \mathbf{I}_p) \frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$$

(ii) $$\frac{\partial \mathbf{Z} y}{\partial \mathbf{X}} = \frac{\partial \mathbf{Z}}{\partial \mathbf{X}} y + \mathbf{Z} \otimes \frac{\partial y}{\partial \mathbf{X}}$$

(iii) $$\frac{\partial \mathbf{AYB}}{\partial \mathbf{X}} = (\mathbf{A} \otimes \mathbf{I}_p) \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} (\mathbf{B} \otimes \boldsymbol{I}_p).$$

(b) *Let $\mathbf{Y}: q \times r$ and $\mathbf{Z}: s \times t$ be functions of $\mathbf{X}$. Then*

$$\frac{\partial \mathbf{Y} \otimes \mathbf{Z}}{\partial \mathbf{X}} = \mathbf{Y} \otimes \frac{\partial \mathbf{Z}}{\partial \mathbf{X}} + (\mathbf{K}_{q,s} \otimes \mathbf{I}_p) \left(\mathbf{Z} \otimes \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} \right) (\mathbf{K}_{t,r} \otimes \mathbf{I}_p).$$

(c) *Let $\mathbf{Y}: q \times q$, y a scalar function with scalar argument and z a scalar, all functions of $\mathbf{X}$. Then*

(i) $$\frac{\partial y\{z(\mathbf{X})\}}{\partial \mathbf{X}} = \frac{\partial y}{\partial z} \frac{\partial z}{\partial \mathbf{X}}$$

(ii) $$\frac{\partial\, tr\,(\mathbf{Y})}{\partial \mathbf{X}} = \sum_{m=1}^{q} (\mathbf{e}_m' \otimes \mathbf{I}_p) \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} (\mathbf{e}_m \otimes \mathbf{I}_p) \qquad (\mathbf{e}_m: q \times 1)$$

(iii) $$\frac{\partial \mathbf{Y}^{-1}}{\partial \mathbf{X}} = -(\mathbf{Y}^{-1} \otimes \boldsymbol{I}_p) \frac{\partial \mathbf{Y}}{\partial \mathbf{X}} (\mathbf{Y}^{-1} \otimes \boldsymbol{I}_p) \qquad (\mathbf{Y} \text{ non-singular}).$$

(d) *Let $\mathbf{A}: p \times p$ and $\mathbf{B}: p \times p$ be constants.*

(i) $$\frac{\partial \mathbf{X}}{\partial \mathbf{X}} = \tfrac{1}{2}\{vec\,(\mathbf{I}_p)\, vec\,(\mathbf{I}_p)' + \mathbf{K}_{p,p}\}$$

(ii) $$\frac{\partial\, tr\,(\mathbf{AX})}{\partial \mathbf{X}} = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}')$$

(iii) $$\frac{\partial\, tr\,(\mathbf{XAXB})}{\partial \mathbf{X}} = \tfrac{1}{2}(\mathbf{BXA} + \mathbf{B}'\mathbf{XA}' + \mathbf{A}'\mathbf{XB}' + \mathbf{AXB}).$$

(e) *For every real r*

$$\frac{\partial |\mathbf{X}|^r}{\partial \mathbf{X}} = r\mathbf{X}^{-1} |\mathbf{X}|^r \qquad (\mathbf{X} \textit{ non-singular}).$$

Proof. Other than $\{\partial\, \mathrm{tr}(\mathbf{Y})\}/\partial \mathbf{X}$ all statements can be found in Rogers (1980). Since $\sum \mathbf{e}_m \mathbf{e}'_m = \mathbf{I}_q$, $\mathrm{tr}(\mathbf{Y}) = \mathrm{tr}(\sum \mathbf{e}_m \mathbf{e}'_m \mathbf{Y}) = \sum \mathbf{e}'_m \mathbf{Y} \mathbf{e}_m$ and therefore a(iii) implies c(ii). Furthermore, we are going to show (e) in such a manner that it is apparent that the derivative of the determinant is linked to the derivative of the trace function which is not so well-known. Using the density of the multivariate normal distribution and d(iii) yield $[\mathbf{y} \sim N_p\{0, (\mathbf{XX})^{-1}\}]$

$$\frac{\partial |\mathbf{X}|^{-1}}{\partial \mathbf{X}} = \frac{\partial |\mathbf{XX}|^{-1/2}}{\partial \mathbf{X}} = -\tfrac{1}{2}|\mathbf{XX}|^{-1/2} E\left[\frac{\partial\, \mathrm{tr}(\mathbf{XXyy}')}{\partial \mathbf{X}}\right]$$

$$= -\tfrac{1}{2}|\mathbf{XX}|^{-1/2} E[\mathbf{Xyy}' + \mathbf{yy}'\mathbf{X}] \qquad (E[\cdot] \text{ means expectation})$$

which is identical to (e) for $r = -1$. The general case follows from c(i). □

Lemma 2.2

Let $\mathbf{X}: p \times p$ be a symmetric matrix whose elements are otherwise functionally independent:

(i) $$\frac{\partial \mathbf{X}^{-1}}{\partial \mathbf{X}} = -\tfrac{1}{2}\{vec(\mathbf{X}^{-1})\, vec(\mathbf{X}^{-1})' + \mathbf{K}_{p,p}(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1})\}$$

(ii) $$\frac{\partial \overset{r}{\otimes} \mathbf{X}^{-1}}{\partial \mathbf{X}} = \sum_{i=0}^{r-1} (\mathbf{I}_{p^i} \otimes \mathbf{K}_{p,p^{r-1-i}} \otimes \mathbf{I}_p)\left(\overset{r-1}{\otimes} \mathbf{X} \otimes \frac{\partial \mathbf{X}^{-1}}{\partial \mathbf{X}}\right)(\mathbf{I}_{p^i} \otimes \mathbf{K}_{p^{r-1-i},p} \otimes \mathbf{I}_p) \qquad r = 1, 2, \ldots$$

$$\left(\overset{r}{\otimes} \mathbf{X}^{-1} = \underbrace{\mathbf{X}^{-1} \otimes \mathbf{X}^{-1} \otimes \ldots \otimes \mathbf{X}^{-1}}_{r}, \overset{1}{\otimes} \mathbf{X}^{-1} = \mathbf{X}^{-1}, \overset{0}{\otimes} \mathbf{X}^{-1} = 1\right)$$

Proof. (i) is obtained by aid of lemma 2.1 c(iii) and d(i). From lemma 2.1 (b) follows that (ii) is true for $r = 1, 2$. Further, lemma 2.1 (b) implies that

$$\frac{\partial \overset{r}{\otimes} \mathbf{X}^{-1}}{\partial \mathbf{X}} = \mathbf{X}^{-1} \otimes \frac{\partial \overset{r-1}{\otimes} \mathbf{X}^{-1}}{\partial \mathbf{X}} + (\mathbf{K}_{p,p^{r-1}} \otimes \mathbf{I}_p)\left(\overset{r-1}{\otimes} \mathbf{X}^{-1} \otimes \frac{\partial \mathbf{X}^{-1}}{\partial \mathbf{X}}\right)(\mathbf{K}_{p^{r-1},p} \otimes \mathbf{I}_p). \tag{2.3}$$

Suppose that (ii) is valid for $r-1$ and inserting $(\partial \otimes^{r-1} \mathbf{X}^{-1})/\partial \mathbf{X}$ in (2.3) together with some calculations establishes (ii). □

The next lemma provides some important auxiliary results. The proofs are omitted since they are rather straightforward. For notational convenience we write in (i) of the lemma $\mathbf{A}, \mathbf{B}, \ldots$ instead of $\mathbf{A}_i, \mathbf{B}_i, \ldots$ and in (ii) $\mathbf{A}_1, \mathbf{B}_1, \ldots$ instead of $\mathbf{A}_{1i}, \mathbf{B}_{1i}, \ldots$.

Lemma 2.3

(i) $$\sum_i \{\mathbf{A} \otimes \mathbf{B} + vec(\mathbf{D}')\, vec(\mathbf{C})' + \mathbf{K}_{p,r}(\mathbf{F}' \otimes \mathbf{E})\} = 0$$

is equivalent to

$$\sum_i \{vec(\mathbf{A})\, vec(\mathbf{B})' + \mathbf{C} \otimes \mathbf{D} + \mathbf{K}_{q,p}(\mathbf{E} \otimes \mathbf{F})\} = 0$$

where $\mathbf{A}: p \times q$, $\mathbf{B}: r \times s$, $\mathbf{C}: q \times s$, $\mathbf{D}: p \times r$, $\mathbf{E}: p \times s$ and $\mathbf{F}: q \times r$.

(ii) *Let $\mathbf{P}_1 = \mathbf{K}_{p,p} \otimes \mathbf{I}_p$ and $\mathbf{P}_2 = \mathbf{I}_p \otimes \mathbf{K}_{p,p}$.*

$$\sum_i \Big\{\mathbf{A}_1 \otimes \mathbf{B}_1 \otimes \mathbf{C}_1 + vec(\mathbf{A}_2)\, vec(\mathbf{B}_2)' \otimes \mathbf{C}_2 + \mathbf{P}_1(\mathbf{A}_3 \otimes \mathbf{B}_3 \otimes \mathbf{C}_3)$$

$$+ \mathbf{P}_2\{vec(\mathbf{A}_4)\, vec(\mathbf{B}_4)' \otimes \mathbf{C}_4\}\mathbf{P}_2 + \mathbf{P}_2\mathbf{P}_1(\mathbf{A}_5 \otimes \mathbf{B}_5 \otimes \mathbf{C}_5)\mathbf{P}_2$$

$$+ \mathbf{P}_1\mathbf{P}_2\{vec(\mathbf{A}_6)\, vec(\mathbf{B}_6)' \otimes \mathbf{C}_6\}\mathbf{P}_2\mathbf{P}_1 + \mathbf{P}_1\mathbf{P}_2\mathbf{P}_1(\mathbf{A}_7 \otimes \mathbf{B}_7 \otimes \mathbf{C}_7)\mathbf{P}_2\mathbf{P}_1\Big\} = 0$$

is equivalent to

$$\sum_i \{vec(\mathbf{B}_1')\,vec(\mathbf{A}_1')'\otimes\mathbf{C}_1+\mathbf{A}_2\otimes\mathbf{B}_2\otimes\mathbf{C}_2+\mathbf{P}_1(\mathbf{B}_3'\otimes\mathbf{A}_3\otimes\mathbf{C}_3)$$

$$+\{vec(\mathbf{C}_4')\,vec(\mathbf{B}_4)'\otimes\mathbf{A}_4\}\mathbf{P}_2\mathbf{P}_1+\{vec(\mathbf{C}_5')\,vec(\mathbf{B}_5')'\otimes\mathbf{A}_5\}\mathbf{P}_2$$

$$+\mathbf{P}_2\{vec(\mathbf{A}_6)\,vec(\mathbf{C}_6')'\otimes\mathbf{B}_6'\}+\mathbf{P}_1\mathbf{P}_2\{vec(\mathbf{A}_7)\,vec(\mathbf{C}_7')'\otimes\mathbf{B}_7\}\}=0$$

where the **A***'s,* **B***'s and* **C***'s are square matrices with dimensions* $p\times p$. □

We are going to use the density function of the Wishart distribution, $f(\mathbf{W})$, given by

$$f(\mathbf{W})=c|\mathbf{W}|^{1/2(n-p-1)}|\boldsymbol{\Sigma}|^{-1/2n}\exp\{-\tfrac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\mathbf{W})\} \qquad n\geqslant p \tag{2.4}$$

where c is the normalizing constant. The next theorem can for example be verified using Stokes' theorem (Haff, 1981, section 2) or applying the fundamental theorem in calculus (von Rosen, 1985, p. 37). Furthermore, $d\mathbf{W}$ denotes the Lebesque measure $\Pi_{k\leqslant l}dw_{kl}$ in $R^{1/2p(p+1)}\{\mathbf{W}=(w_{kl})\}$ and $\int_{\mathbf{W}_{p.d.}}$ means an ordinary multiple integral with integration performed over the subset of $R^{1/2p(p+1)}$ where $\mathbf{W}$ is positive definite.

Theorem 2.1

Let $\mathbf{W}\sim\mathbf{W}_p(\boldsymbol{\Sigma},n)$ *and* $f(\mathbf{W})$ *its density given by (2.4). Then*

$$\int_{\mathbf{W}_{p.d.}}\frac{\partial}{\partial\mathbf{W}}\left(\overset{r}{\otimes}\mathbf{W}^{-1}f(\mathbf{W})\right)d\mathbf{W}=0 \qquad \text{if} \quad n-p-1-2r>0 \qquad r=0,1,2,\ldots$$ □

Corollary 2.1

Let $\mathbf{W}\sim\mathbf{W}_p(\boldsymbol{\Sigma},n)$.

$$2E\left[\frac{\partial}{\partial\mathbf{W}}\overset{r}{\otimes}\mathbf{W}^{-1}\right]+(n-p-1)E\left[\overset{r+1}{\otimes}\mathbf{W}^{-1}\right]=E\left[\overset{r}{\otimes}\mathbf{W}^{-1}\right]\otimes\boldsymbol{\Sigma}^{-1} \qquad \text{if} \quad n-p-1-2r>0.$$

Proof. Applying lemma 2.1 c(i), d(ii) and (e) yield

$$\frac{\partial f(\mathbf{W})}{\partial\mathbf{W}}=\tfrac{1}{2}(n-p-1)\mathbf{W}^{-1}f(\mathbf{W})-\tfrac{1}{2}\boldsymbol{\Sigma}^{-1}f(\mathbf{W}).$$

Thus by utilizing lemma 2.1 a(ii) the theorem follows. □

Remark. Using another type of matrix derivative Haff (1981, p. 217) presented a similar identity.

3. Moments of order less than four

All results presented in the following will depend on corollary 2.1. Our intention is to present moment relations for $\otimes^r\mathbf{W}^{-1}$: and from now on it is a matter of using

$$E\left[\frac{\partial}{\partial\mathbf{W}}\overset{r}{\otimes}\mathbf{W}^{-1}\right]+(n-p-1)E\left[\overset{r+1}{\otimes}\mathbf{W}^{-1}\right]$$

in order to obtain $E[\otimes^{r+1}\mathbf{W}^{-1}]$. This is an algebraic problem. On one the problem is straightforward whereas on the other, being somewhat more ambitious about the expressions for the moments, the difficulties pile up. For $r=3$, already, it will be seen that it is far from evident how one should proceed; and for $r\geqslant 4$ we leave the problem unsolved.

In the next two theorems we derive $E[\mathbf{W}^{-1}]$, $E[\otimes^2\mathbf{W}^{-1}]$ and $E[\otimes^3\mathbf{W}^{-1}]$. The first and second order moments are well known whereas the expression for the third is new. The method of obtaining the second order moments can be viewed as a multivariate extension of Siskind's (1972) approach. Moreover, in the proofs of the theorems the following relation appears to be useful;

$$\begin{pmatrix} -a & -a & b \\ -a & b & -a \\ b & -a & -a \end{pmatrix}^{-1} = c \begin{pmatrix} a & a & b-a \\ a & b-a & a \\ b-a & a & a \end{pmatrix} \tag{3.1}$$

where $c^{-1}=(b-2a)(b+a)$.

Theorem 3.1
Let $\mathbf{W} \sim W_p(\mathbf{\Sigma}, n)$.

(i) $E[\mathbf{W}^{-1}] = \dfrac{1}{n-p-1}\mathbf{\Sigma}^{-1}$ *if* $n-p-1>0$

(ii) $E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right] = c_1 \overset{2}{\otimes}\mathbf{\Sigma}^{-1} + c_2\, vec(\mathbf{\Sigma}^{-1})\, vec(\mathbf{\Sigma}^{-1})' + c_2 \mathbf{K}_{p,p} \overset{2}{\otimes}\mathbf{\Sigma}^{-1}$

where $c_2^{-1} = (n-p)(n-p-1)(n-p-3)$ *and* $c_1 = (n-p-2)c_2$.

Proof. Putting $r=0$ in corollary 2.1 yields (i). Putting $r=1$ in the corollary and then together with lemma 2.2 (i) it is established that

$$-E[\mathrm{vec}(\mathbf{W}^{-1})\,\mathrm{vec}(\mathbf{W}^{-1})'] - \mathbf{K}_{p,p}E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right] + (n-p-1)E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right]$$
$$= \frac{1}{n-p-1}\overset{2}{\otimes}\mathbf{\Sigma}^{-1}. \tag{3.2}$$

Let

$$\mathbf{T}' = \left\{E[\mathrm{vec}(\mathbf{W}^{-1})\,\mathrm{vec}(\mathbf{W}^{-1})'], \mathbf{K}_{p,p}E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right], E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right]\right\}$$

$$\mathbf{M}' = \left\{\overset{2}{\otimes}\mathbf{\Sigma}^{-1}, \mathbf{K}_{p,p}\overset{2}{\otimes}\mathbf{\Sigma}^{-1}, \mathrm{vec}(\mathbf{\Sigma}^{-1})\,\mathrm{vec}(\mathbf{\Sigma}^{-1})'\right\}.$$

Now, pre-multiplying (3.2) with the non-singular matrix $\mathbf{K}_{p,p}$ and additionally applying lemma 2.3 (i) to (3.2) yields the following equation in $\mathbf{T}$:

$$\begin{pmatrix} -1 & -1 & n-p-1 \\ -1 & n-p-1 & -1 \\ n-p-1 & -1 & -1 \end{pmatrix}\mathbf{T} = \frac{1}{n-p-1}\mathbf{M}. \tag{3.3}$$

Thus, (3.1) leads to the relation in (ii). □

Remark. In (ii) we were concerned only with $E[\otimes^2\mathbf{W}^{-1}]$ but an advantage of solving (3.3) is that we immediately obtain $E[\mathrm{vec}(\mathbf{W}^{-1})\,\mathrm{vec}(\mathbf{W}^{-1})']$ which is useful if one is interested in the dispersion matrix for $\mathbf{W}^{-1}$. This is of course somewhat trivial since this result is obtainable from lemma 2.3 (i) but the idea of solving (3.3) is that for higher order moments we get expressions which are not readily obtainable from lemma 2.3.

Theorem 3.2
Let $\mathbf{W} \sim \mathbf{W}_p(\boldsymbol{\Sigma}, n)$, c_1 and c_2 *are defined in theorem 3.1. Set* $\mathbf{P}_1 = (\mathbf{K}_{p,p} \otimes \mathbf{I}_p)$, $\mathbf{P}_2 = (\mathbf{I}_p \otimes \mathbf{K}_{p,p})$ *and*

$$c_3 = \frac{n-p-3}{(n-p-5)(n-p+1)}, \quad c_4 = \frac{2}{(n-p-5)(n+p+1)}.$$

$$E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right] = c_3 c_1 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1} + (c_4 c_1 + c_3 c_2)\, vec(\boldsymbol{\Sigma}^{-1})\, vec(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}$$

$$+ (c_3 c_2 + c_4 c_2)\mathbf{P}_1 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1} + c_3 c_2 \mathbf{P}_2\, vec(\boldsymbol{\Sigma}^{-1})\, vec(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\mathbf{P}_2$$

$$+ c_3 c_2 \mathbf{P}_2 \mathbf{P}_1 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1}\mathbf{P}_2 + (c_4 c_1 - c_3 c_2)\mathbf{P}_2 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1} + c_4 c_2 \mathbf{P}_2 \mathbf{P}_1 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1}$$

$$+ 2c_4 c_2\, vec(\boldsymbol{\Sigma}^{-1})\, vec(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\mathbf{P}_2 + c_4 c_2 \mathbf{P}_1 \mathbf{P}_2 \overset{3}{\otimes}\boldsymbol{\Sigma}^{-1}$$

$$+ c_4 c_2\, vec(\boldsymbol{\Sigma}^{-1})\, vec(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}\mathbf{P}_2\mathbf{P}_1 - (c_3 c_2 + c_4 c_2)\boldsymbol{\Sigma}^{-1} \otimes vec(\boldsymbol{\Sigma}^{-1}) vec(\boldsymbol{\Sigma}^{-1})'$$

$$- c_4 c_2 \mathbf{P}_1 \mathbf{P}_2\, vec(\boldsymbol{\Sigma}^{-1})\, vec(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}.$$

Proof. Putting $r=2$ in corollary 2.1 implies that

$$2E\left[\mathbf{W}^{-1} \otimes \frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}}\right] + 2\mathbf{P}_1 E\left[\mathbf{W}^{-1} \otimes \frac{2\mathbf{W}^{-1}}{\partial \mathbf{W}}\right]\mathbf{P}_1 + (n-p-1)E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right]$$

$$= E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right] \otimes \boldsymbol{\Sigma}^{-1}. \tag{3.4}$$

Set

$$\mathbf{M}_1 = E\left[\overset{2}{\otimes}\mathbf{W}^{-1}\right] \otimes \boldsymbol{\Sigma}^{-1}, \quad \mathbf{M}_2 = \mathbf{P}_2 \mathbf{M}_1 \mathbf{P}_2, \quad \mathbf{M}_3 = \mathbf{P}_1 \mathbf{M}_2 \mathbf{P}_1.$$

Now, pre- and post-multiplying (3.4) with $\mathbf{P}_2$ which then in turn is pre- and post-multiplied with $\mathbf{P}_1$ we obtain, in addition, the following two equations:

$$2E\left[\mathbf{W}^{-1} \otimes \frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}}\right] + 2E\left[\frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}} \otimes \mathbf{W}^{-1}\right] + (n-p-1)E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right] = \mathbf{M}_2 \tag{3.5}$$

$$2\mathbf{P}_1 E\left[\mathbf{W}^{-1} \otimes \frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}}\right]\mathbf{P}_1 + 2E\left[\frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}} \otimes \mathbf{W}^{-1}\right] + (n-p-1)E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right] = \mathbf{M}_3 \tag{3.6}$$

Adding (3.4) and (3.5) and then subtracting (3.6) leads to the consideration of

$$4E\left[\mathbf{W}^{-1} \otimes \frac{\partial \mathbf{W}^{-1}}{\partial \mathbf{W}}\right] + (n-p-1)E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right] = \mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3. \tag{3.7}$$

Note that (3.7) is similar to (3.2) and the rest of the proof is more or less a copy of the proof of theorem 3.1 (ii). Let

$$\mathbf{T}' = \left(E[\mathbf{W}^{-1} \otimes \mathrm{vec}(\mathbf{W}^{-1})\, \mathrm{vec}(\mathbf{W}^{-1})'],\ \ \mathbf{P}_2 E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right],\ \ E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right]\right)$$

$$\mathbf{M}' = (\mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3,\ \ (\mathbf{M}_1 + \mathbf{M}_2 - \mathbf{M}_3)\mathbf{P}_2,\ \ \tilde{\mathbf{M}}_1' + \tilde{\mathbf{M}}_2' - \tilde{\mathbf{M}}_3')$$

where

$$\tilde{\mathbf{M}}_1 = c_1 \operatorname{vec}(\boldsymbol{\Sigma}^{-1})\operatorname{vec}(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1} + c_2(\mathbf{I} + \mathbf{P}_1) \overset{3}{\otimes} \boldsymbol{\Sigma}^{-1}$$

$$\tilde{\mathbf{M}}_2 = c_1 \operatorname{vec}(\boldsymbol{\Sigma}^{-1})\operatorname{vec}(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1} + c_2 \operatorname{vec}(\boldsymbol{\Sigma}^{-1})\operatorname{vec}(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}(\mathbf{P}_2\mathbf{P}_1 + \mathbf{P}_2)$$

$$\tilde{\mathbf{M}}_3 = c_1 \operatorname{vec}(\boldsymbol{\Sigma}^{-1})\operatorname{vec}(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1} + c_2(\mathbf{P}_2 + \mathbf{P}_1\mathbf{P}_2)\operatorname{vec}(\boldsymbol{\Sigma}^{-1})\operatorname{vec}(\boldsymbol{\Sigma}^{-1})' \otimes \boldsymbol{\Sigma}^{-1}.$$

Pre-multiplying (3.7) with $\mathbf{P}_2$ and together with lemma 2.3 (ii) we obtain

$$\begin{pmatrix} -2 & -2 & n-p-1 \\ -2 & n-p-1 & -2 \\ n-p-1 & -2 & -2 \end{pmatrix} \mathbf{T} = \mathbf{M}$$

Thus the theorem is established by aid of (3.1) and some calculations. □

In the next corollary some consequences of theorems 3.1 and 3.2 are given, especially functions involving both $\mathbf{W}^{-1}$ and $\mathbf{W}$. These are useful when obtaining moments for the estimators in the growth curve model.

Corollary 3.1

Let $\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n)$ and $c_1 - c_4$ be the same constants as in theorems 3.1 and 3.2.

(i) $E[\mathbf{W}^{-1}\mathbf{W}^{-1}] = (c_1 + c_2)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} + c_1\boldsymbol{\Sigma}^{-1}\, tr(\boldsymbol{\Sigma}^{-1})$ *if* $n-p-3>0$

(ii) $$\begin{aligned} E[\mathbf{W}^{-1}\mathbf{W}^{-1}\mathbf{W}^{-1}] &= (c_3c_1 + c_3c_2 + c_4c_1 + 5c_4c_2)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} \\ &\quad + (2c_3c_2 + c_4c_1 + c_4c_2)\, tr(\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} \\ &\quad - (c_3c_2 + c_4c_2)\, tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1} \\ &\quad - c_4c_2\{tr(\boldsymbol{\Sigma}^{-1})\}^2\boldsymbol{\Sigma}^{-1} \end{aligned}$$ *if* $n-p-5>0$

(iii) $$\begin{aligned} E[tr(\mathbf{W}^{-1}\mathbf{W}^{-1})\mathbf{W}^{-1}] &= (c_4c_1 + c_4c_2)\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} + (c_4c_1 + c_3c_2)\{tr(\boldsymbol{\Sigma}^{-1})\}^2\boldsymbol{\Sigma}^{-1} \\ &\quad + (c_3c_1 + c_3c_2 + c_4c_2)\, tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1} + 2c_4c_2\, tr(\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1} \end{aligned}$$ *if* $n-p-5>0$

(iv) $$\begin{aligned} E[\{tr(\mathbf{W}^{-1})\}^3] &= c_3c_1\{tr(\boldsymbol{\Sigma}^{-1})\}^3 + 4c_4c_2\, tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}) \\ &\quad + 2(c_4c_1 + c_3c_2)\, tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1})\, tr(\boldsymbol{\Sigma}^{-1}) \end{aligned}$$ *if* $n-p-5>0$

(v) $E[tr(\mathbf{W}^{-1})\mathbf{W}^{-1}] = c_1\, tr(\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma}^{-1} + 2c_2\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}$ *if* $n-p-3>0$

(vi) $E[\mathbf{W}^{-1} \otimes \mathbf{W}] = \dfrac{n}{n-p-1}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma} - \dfrac{1}{n-p-1}\{vec(\mathbf{I}_p)\, vec(\mathbf{I}_p)' + \mathbf{K}_{p,p}\}$ *if* $n-p-1>0$

(vii) $E[tr(\mathbf{W}^{-1})\mathbf{W}] = \dfrac{n}{n-p-1}\, tr(\boldsymbol{\Sigma}^{-1})\boldsymbol{\Sigma} - \dfrac{2}{n-p-1}\mathbf{I}_p$ *if* $n-p-1>0$

Proof. Post-multiplying theorem 3.1 (ii) with $\operatorname{vec}(\mathbf{I}_p)$ establishes (i). (ii) follows from theorem 3.2 by noting that

$$E\left[\overset{3}{\otimes}\mathbf{W}^{-1}\right]\operatorname{vec}(\mathbf{I}_p) \otimes \mathbf{I}_p = E[\operatorname{vec}(\mathbf{W}^{-1}\mathbf{W}^{-1}) \otimes \mathbf{W}^{-1}],$$

$$(\mathbf{K}_{p,p} \otimes \mathbf{I}_{p^2})\operatorname{vec}(E[\operatorname{vec}(\mathbf{W}^{-1}\mathbf{W}^{-1}) \otimes \mathbf{W}^{-1}]) = \operatorname{vec}(E[\mathbf{W}^{-1}\mathbf{W}^{-1} \otimes \mathbf{W}^{-1}])$$

and thus $E[\mathbf{W}^{-1}\mathbf{W}^{-1}\otimes\mathbf{W}^{-1}]$ is obtained which then is post-multiplied with $\mathrm{vec}(\mathbf{I}_p)$. (iii) is verified by using the fact that

$$E[\mathrm{tr}(\mathbf{W}^{-1}\mathbf{W}^{-1})\mathbf{W}^{-1}]=\sum_{m=1}^{p}(\mathbf{e}_m'\otimes\mathbf{I}_p)E[\mathbf{W}^{-1}\mathbf{W}^{-1}\otimes\mathbf{W}^{-1}](\mathbf{e}_m\otimes\mathbf{I}_p).$$

Differentiating both sides in theorem 3.1 (i) with respect to $\mathbf{\Sigma}^{-1}$ shows (vi). (iv) is obtained by aid of theorem 3.2 since $\mathrm{tr}(\mathbf{W}^{-1}\otimes\mathbf{W}^{-1}\otimes\mathbf{W}^{-1})=\{\mathrm{tr}(\mathbf{W}^{-1})\}^3$. (v) and (vii) are determined in the same way as (iii). □

4. Moments of arbitrary order

Moments of arbitrary order are available for the multivariate normal distribution as well as for the Wishart distribution (see Homlquist, 1985; von Rosen, 1987). It is interesting to note that IW moments of arbitrary order are also obtainable.

The starting point of the derivation of these moments is the relation

$$2\sum_{i=0}^{r-1}(\mathbf{I}_{p^i}\otimes\mathbf{K}_{p,p^{r-1-i}}\otimes\mathbf{I}_p)E\left[\bigotimes^{r-1}\mathbf{W}^{-1}\otimes\frac{\partial\mathbf{W}^{-1}}{\partial\mathbf{W}}\right](\mathbf{I}_{p^i}\otimes\mathbf{K}_{p^{r-1-i},p}\otimes\mathbf{I}_p)$$
$$+(n-p-1)E\left[\bigotimes^{r+1}\mathbf{W}^{-1}\right]=E\left[\bigotimes^{r}\mathbf{W}^{-1}\right]\otimes\mathbf{\Sigma}^{-1}\qquad \textit{if}\quad n-p-1-2r>0 \tag{4.1}$$

obtained by combining lemma 2.2 (ii) and corollary 2.1. In order to simplify notations we define two permutation matrices $\mathbf{P}_1(i)$ and $\mathbf{P}_2(i)$ as follows:

$$\mathbf{P}_1(i)=\left\{\bigotimes^{2}(\mathbf{I}_{p^i}\otimes\mathbf{K}_{p,p^{r-1-i}}\otimes\mathbf{I}_p)\right\}(\mathbf{I}_{p^r}\otimes\mathbf{K}_{p,p^r}\otimes\mathbf{I}_p)(\mathbf{I}_{p^{r-1}}\otimes\mathbf{K}_{p^{r+2},p})$$

$$\mathbf{P}_2(i)=\left\{\bigotimes^{2}(\mathbf{I}_{p^i}\otimes\mathbf{K}_{p,p^{r-1-i}}\otimes\mathbf{I}_p)\right\}(\mathbf{I}_{p^{2r}}\otimes\mathbf{K}_{p,p}).$$

Applying lemma 2.2 (i) and then some calculations give that (4.1) is equivalent to

$$-\sum_{i=0}^{r-1}\{\mathbf{P}_1(i)+\mathbf{P}_2(i)\}\,\mathrm{vec}\left(E\left[\bigotimes^{r+1}\mathbf{W}^{-1}\right]\right)+(n-p-1)\,\mathrm{vec}\left(E\left[\bigotimes^{r+1}\mathbf{W}^{-1}\right]\right)$$
$$=\mathrm{vec}\left(E\left[\bigotimes^{r}\mathbf{W}^{-1}\right]\otimes\mathbf{\Sigma}^{-1}\right).$$

Therefore

$$\mathrm{vec}\left(E\left[\bigotimes^{r+1}\mathbf{W}^{-1}\right]\right)=\left[(n-p-1)\mathbf{I}-\sum_{i=0}^{r-1}\{\mathbf{P}_1(i)+\mathbf{P}_2(i)\}\right]^{-1}\mathrm{vec}\left(E\left[\bigotimes^{r}\mathbf{W}^{-1}\right]\otimes\mathbf{\Sigma}^{-1}\right). \tag{4.2}$$

In (4.2) is supposed that the inverse of

$$\mathbf{V}(r)=(n-p-1)\mathbf{I}-\sum_{i=0}^{r-1}\{\mathbf{P}_1(i)+\mathbf{P}_2(i)\} \tag{4.3}$$

exists and we now show that this will always be the case.

Since $\mathbf{P}_1(i)$ and $\mathbf{P}_2(i)$ are permutation matrices the sum of the rows as well as the sum of the columns of these matrices equal unity. Moreover, using that $n-p-1-2r>0$, the diagonal elements of $\mathbf{V}(r)$ are greater than the corresponding row or column sums of the absolute values of the off-diagonal elements. Thus $\mathbf{V}(r)$ is a diagonally dominant matrix and from, for instance, theorem 7.13 in Basilevsky (1983) follows that the inverse exists.

Using (4.2) the moments can be expressed in a non-recursive manner. Note that

$$\mathrm{vec}\left(E\left[\overset{r}{\otimes}\mathbf{W}^{-1}\right]\otimes\mathbf{\Sigma}^{-1}\right)=(\mathbf{I}_{p^r}\otimes\mathbf{K}_{p,p^r}\otimes\mathbf{I}_p)\mathbf{K}_{p^{2r},p^2}\left\{\mathrm{vec}\,(\mathbf{\Sigma}^{-1})\otimes\mathrm{vec}\left(E\left[\overset{r}{\otimes}\mathbf{W}^{-1}\right]\right)\right\}$$

and let

$$\mathbf{A}(r)=\mathbf{V}(r)^{-1}(\mathbf{I}_{p^r}\otimes\mathbf{K}_{p,p^r}\otimes\mathbf{I}_p)\mathbf{K}_{p^{2r},p^2}. \tag{4.4}$$

Then (4.2) can be written

$$\mathrm{vec}\left(E\left[\overset{r+1}{\otimes}\mathbf{W}^{-1}\right]\right)=\mathbf{A}(r)\left\{\mathrm{vec}\,(\mathbf{\Sigma}^{-1})\otimes\mathrm{vec}\left(E\left[\overset{r}{\otimes}\mathbf{W}^{-1}\right]\right)\right\}$$

and thus (remember theorem 3.1 (i))

$$\mathrm{vec}\left(E\left[\overset{r+1}{\otimes}\mathbf{W}^{-1}\right]\right)=\frac{1}{n-p-1}\prod_{i=0}^{r-1}\{\mathbf{I}_{p^{2i}}\otimes\mathbf{A}(r-i)\}\overset{r+1}{\otimes}\mathrm{vec}\,(\mathbf{\Sigma}^{-1})\qquad r>0. \tag{4.5}$$

Theorem 4.1

Let $\mathbf{W}\sim W_p(\mathbf{\Sigma},n)$ and suppose $n-p-1-2r>0$. Then $E[\otimes^{r+1}\mathbf{W}^{-1}]$, $r>0$, is given by (4.2) in a recursive manner and by (4.5) in a non-recursive manner. □

We conclude this section by showing that the approach of this section and the approach of the previous section differ. (4.5) will be applied when $p=2$ and $r=1$. Let

$$\mathbf{C}=c_1(\mathbf{I}_2\otimes\mathbf{K}_{2,2}\otimes\mathbf{I}_2)+c_2\mathbf{I}_8+c_2(\mathbf{I}_2\otimes\mathbf{K}_{2,2}\otimes\mathbf{I}_2)(\mathbf{K}_{2^2,2}\otimes\mathbf{I}_2)$$

where c_1 and c_2 are defined in theorem 3.1. From theorem 3.1 (ii) and some calculations it follows that $E[\otimes^2\mathbf{W}^{-1}]=\mathbf{C}\otimes^2\mathrm{vec}\,(\mathbf{\Sigma}^{-1})$ and therefore is reasonable to believe that $\mathbf{C}=1/(n-3)\mathbf{A}(1)$. However, some tedious calculations indicate that this is not the case since it appears that $\mathbf{C}$ consists of fewer elements different from zero than $\mathbf{A}(1)$.

It is worth knowing that the reason both approaches give correct expressions is that many of the elements in $\otimes^2\mathrm{vec}\,(\mathbf{\Sigma}^{-1})$ are identical. This indicates, in fact, that in order to extend the results of section 3 one ought to take into account the relationships between the elements in $\otimes^r\mathbf{W}^{-1}$ in a more detailed manner than hitherto. For instance, design special matrix derivatives, define suitable operators and then also, of course, refine the algebraic treatment.

5. Application

In this section we show an application of corollary 3.1.

Let $\mathbf{X}:p\times n$ have columns which are independent p-variate normally distributed with an unknown covariance matrix Σ, and $E[X]=\mathbf{ABC}$ $\mathbf{A}:p\times q$, $q\leqslant p$, and $\mathbf{C}:k\times n$, $k+p\leqslant n$, are known design matrices and $\mathbf{B}$ is unknown. This is known as the growth curve model (see Potthoff & Roy, 1964; Srivastava & Khatri, 1979). The maximum likelihood estimators are given by

$$\begin{aligned}\hat{\mathbf{B}}&=(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\mathbf{X}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\\ n\hat{\mathbf{\Sigma}}&=(\mathbf{X}-\mathbf{A}\hat{\mathbf{B}}\mathbf{C})(\mathbf{X}-\mathbf{A}\hat{\mathbf{B}}\mathbf{C})'\end{aligned} \tag{5.1}$$

where

$$\mathbf{S}=\mathbf{X}(\mathbf{I}-\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C})\mathbf{X}'$$

and the inverses are supposed to exist.

The distributions of $\hat{\mathbf{B}}$ and $\hat{\Sigma}$ are complicated (see Gleser & Olkin, 1970). However, moments are obtainable. These can be used when studying properties of $\hat{\mathbf{B}}$ and $\hat{\Sigma}$, as well as when approximating the distributions. In order to obtain these moments we need certain functions of the moments for the IW. To show an application of the third order moments for the IW we consider $E[\otimes^6(\hat{\mathbf{B}}-\mathbf{B})]$. Results could also have been presented for the third order moments for $\hat{\Sigma}$ but $E[\otimes^6(\hat{\mathbf{B}}-\mathbf{B})]$ is somewhat easier to deal with. In the next we will be very sparse with details since the purpose is to show an application and not to discuss the growth curve model. For details the reader is referred to von Rosen (1985).

We are going to apply a result concerning moments for matrix normally distributed variables. A derivation of the next relation is given by von Rosen (1987).

Let $\mathbf{Y}:p\times n$ have columns which are independent normally distributed with mean zero and covariance Σ. Then

$$E\left[\overset{2r}{\otimes}\mathbf{Y}\right]=\sum_{\substack{2\leqslant i_j\leqslant 2r-2j\\ j=0,1,\ldots,r-1}}\mathbf{H}(p,r,i_0,i_1,\ldots,i_{r-1})\overset{r}{\otimes}\operatorname{vec}(\Sigma)\operatorname{vec}(\mathbf{I})'\mathbf{H}(n,r,i_0,\ldots,i_{r-1})' \qquad r=1,2,3 \tag{5.2}$$

where

$$\mathbf{H}(a,r,i_0,\ldots,i_{r-1})=\prod_{k=0}^{r-1}\{\mathbf{I}_{a^{2k}}\otimes\mathbf{P}(a,i_k,2r-2k)\}$$

and

$$\mathbf{P}(a,b,c)=(\mathbf{K}_{a,a^{b-2}}\otimes\mathbf{I}_a)\mathbf{K}_{a^{b-2},a^2}\otimes\mathbf{I}_{a^{c-b}}=\mathbf{I}_a\otimes\mathbf{K}_{a^{b-2},a}\otimes\mathbf{I}_{a^{c-b}}.$$

Although $\hat{\mathbf{B}}$ given by (5.1) is a complicated stochastic expression we have the nice property that $\mathbf{S}$ and $\mathbf{XC}'$ are independently distributed. Hence, $E[\hat{\mathbf{B}}]=\mathbf{B}$ and using (5.2)

$$\begin{aligned}
E\left[\overset{6}{\otimes}(\hat{\mathbf{B}}-\mathbf{B})\right]&=E\left[\overset{6}{\otimes}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}(\mathbf{X}-\mathbf{ABC})\mathbf{C}'(\mathbf{CC}')^{-1}\right]\\
&=E\left[\overset{6}{\otimes}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\right]E\left[\overset{6}{\otimes}(\mathbf{X}-\mathbf{ABC})\right]\overset{6}{\otimes}\mathbf{C}'(\mathbf{CC}')^{-1}\\
&=\sum_{\substack{2\leqslant i_j\leqslant 6-2j\\ j=0,1,2}}\mathbf{H}(q,3,i_0,i_1,i_2)E\left[\overset{6}{\otimes}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\right]\\
&\quad\overset{3}{\otimes}\{\operatorname{vec}(\Sigma)\operatorname{vec}(\mathbf{I})'\}\overset{6}{\otimes}\mathbf{C}'(\mathbf{CC}')^{-1}\mathbf{H}(k,3,i_0,i_1,i_2)'=\sum_{\substack{2\leqslant i_j\leqslant 6-2j\\ j=0,1,2}}\\
&\quad\mathbf{H}(q,3,i_0,i_1,i_2)E\left[\overset{3}{\otimes}\operatorname{vec}\{(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\Sigma\mathbf{S}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\}\right]\\
&\quad\overset{3}{\otimes}\operatorname{vec}\{(\mathbf{CC}')^{-1}\}'\mathbf{H}(k,3,i_0,i_1,i_2)'.
\end{aligned}\tag{5.3}$$

It is always possible to find matrices $\mathbf{M}$, $\mathbf{T}$ and Γ where $\mathbf{M}$ and $\mathbf{T}$ are non-singular and Γ orthogonal such that

$$\Sigma=\mathbf{MM}' \quad\text{and}\quad \mathbf{A}'=\mathbf{T}(\mathbf{I}_q:\mathbf{O})\Gamma\mathbf{M}'. \tag{5.4}$$

Since $\mathbf{S} \sim W_p(\Sigma, n-k)$ we obtain

$$\overset{3}{\otimes}\operatorname{vec}\{(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\mathbf{A}'\mathbf{S}^{-1}\Sigma\mathbf{S}^{-1}\mathbf{A}(\mathbf{A}'\mathbf{S}^{-1}\mathbf{A})^{-1}\}$$

$$=\overset{3}{\otimes}\operatorname{vec}\{(\mathbf{T}')^{-1}\{\mathbf{I}+\mathbf{Y}_1\mathbf{Y}_2'(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\mathbf{Y}_2\mathbf{Y}_1'\}\mathbf{T}^{-1}\} \tag{5.5}$$

where the elements of $\mathbf{Y}'=(\mathbf{Y}_1':\mathbf{Y}_2'):(n-k\times q:n-k\times p-q)$ are independently univariate normally distributed with mean zero and variance one $\{N(0,1)\}$. Moreover, let $\mathbf{Z}=\mathbf{Y}_2'(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\mathbf{Y}_2$ and it follows that the right-hand side of (5.5) equals

$$\overset{6}{\otimes}(\mathbf{T}')^{-1}\overset{3}{\otimes}\operatorname{vec}(\mathbf{I}+\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')=\overset{6}{\otimes}(\mathbf{T}')^{-1}\left\{\overset{3}{\otimes}\operatorname{vec}(\mathbf{I})\right.$$

$$+\{\mathbf{I}+(\mathbf{K}_{q^2,q^2}\otimes\mathbf{I}_{q^2})+\mathbf{K}_{q^4,q^2}\}\left[\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')\otimes\left\{\overset{2}{\otimes}\operatorname{vec}(\mathbf{I})\right\}\right.$$

$$\left.\left.+\operatorname{vec}(\mathbf{I})\otimes\left\{\overset{2}{\otimes}\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')\right\}\right]+\overset{3}{\otimes}\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')\right\}. \tag{5.6}$$

Observe that

$$E\left[\overset{3}{\otimes}\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')\right]=E\left[\overset{6}{\otimes}\mathbf{Y}_1\right]E[\operatorname{vec}(\mathbf{Z})]$$

$$=\sum_{\substack{2\leq i_j\leq 6-2j\\ j=0,1}}\mathbf{H}(q,3,i_0,i_1,2)\overset{3}{\otimes}\operatorname{vec}(\mathbf{I})s(i_0,i_1) \tag{5.7}$$

where

$$s(i_0,i_1)=\overset{3}{\otimes}\operatorname{vec}(\mathbf{I})'\mathbf{H}(n-k,3,i_0,i_1,2)'E\left[\overset{3}{\otimes}\operatorname{vec}(\mathbf{Z})\right].$$

$s(i_0,i_1)$ is a scalar and since $\mathbf{H}(n-k,3,i_0,i_1,2)$ is a permutation operator it is easy to see that $s(i_0,i_1)$ equals either $E[\{\operatorname{tr}(\mathbf{Z})\}^3]$, $E[\operatorname{tr}(\mathbf{Z})\operatorname{tr}(\mathbf{ZZ})]$ or $E[\operatorname{tr}(\mathbf{ZZZ})]$. In fact, some computations show that

$$s(2,2)=E[\{\operatorname{tr}(\mathbf{Z})\}^3]$$

$$s(2,3)=s(2,4)=s(3,2)=s(4,2)=s(5,4)=s(6,4)=E[\operatorname{tr}(\mathbf{ZZ})\operatorname{tr}(\mathbf{Z})]$$

$$s(3,3)=s(3,4)=s(4,3)=s(4,4)=s(5,2)=s(5,3)=s(6,2)=s(6,3)=E[\operatorname{tr}(\mathbf{ZZZ})] \tag{5.8}$$

and since $\mathbf{Y}_2\mathbf{Y}_2'\sim W_{p-q}(I,n-k)$, $\operatorname{tr}(\mathbf{Z})=\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}$, $\operatorname{tr}(\mathbf{ZZ})=\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}$ and $\operatorname{tr}(\mathbf{ZZZ})=\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}$ (5.7) is determined by aid of corollary 3.1.

From (5.6) follows that $E[\otimes^2\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')]$ and $E[\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')]$ are needed. In the same manner as (5.7) was found we can obtain expressions for these moments;

$$E\left[\overset{2}{\otimes}\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')\right]=\sum_{2\leq i_0\leq 4}\mathbf{H}(q,2,i_0,2)\overset{2}{\otimes}\operatorname{vec}(\mathbf{I})s(i_0)$$

where

$$s(2)=E[\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}^2]$$

$$s(3)=s(4)=E[\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}] \tag{5.9}$$

and

$$E[\operatorname{vec}(\mathbf{Y}_1\mathbf{Z}\mathbf{Y}_1')] = s\operatorname{vec}(\mathbf{I}) \tag{5.10}$$

where $s = E[\operatorname{tr}\{(\mathbf{Y}_2\mathbf{Y}_2')^{-1}\}]$.

Finally we note that from (5.4) follows that $\otimes^6(\mathbf{T}')^{-1}\otimes^3\operatorname{vec}(\mathbf{I}) = \otimes^3\operatorname{vec}\{(\mathbf{T}')^{-1}\mathbf{T}^{-1}\} = \otimes^3\operatorname{vec}\{(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\}$ and using (5.3), (5.5)–(5.10) we are able to write down $E[\otimes^6(\hat{\mathbf{B}}-\mathbf{B})]$. Let $\mathbf{V} = \otimes^3\operatorname{vec}\{(\mathbf{A}'\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\}\operatorname{vec}(\mathbf{CC}')^{-1}\}$ and then

$$\begin{aligned} E\left[\overset{6}{\otimes}(\hat{\mathbf{B}}-\mathbf{B})\right] = \sum_{\substack{2\leqslant i_j\leqslant 6-2j\\ j=0,1}} \mathbf{H}(q,3,i_0,i_1,2)\Bigg\{ & \mathbf{V} + \{\mathbf{I}+(\mathbf{K}_{q^2,q^2}\otimes\mathbf{I}_{q^2})+\mathbf{K}_{q^4,q^2}\} \\ & \times\left[s\mathbf{V} + \sum_{2\leqslant \underline{i}_0\leqslant 4}\{\mathbf{I}_{q^2}\otimes\mathbf{H}(q,2,\underline{i}_0,2)\}s(\underline{i}_0)\mathbf{V}\right] \\ & + \sum_{\substack{2\leqslant \underline{\underline{i}}_j\leqslant 6-2j\\ j=0,1}} \mathbf{H}(q,3,\underline{\underline{i}}_0,\underline{\underline{i}}_1,2)s(\underline{\underline{i}}_0,\underline{\underline{i}}_1)\mathbf{V}\Bigg\}\mathbf{H}(k,3,i_0,\mathrm{i}_1,2)'. \end{aligned}$$

Acknowledgements

The comments made by a referee and the secretarial help of Mrs Margit Holmberg are acknowledged.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*, 2nd edn. Wiley, New York.

Basilevsky, A. (1983). *Applied matrix algebra in the statistical sciences.* North-Holland, New York.

Das Gupta, S. (1968). Some aspects of discrimination function coefficients. *Sankyā Ser. A* **30**, 387–400.

Gleser, L. J. & Olkin, I. (1970). Linear models in multivariate analysis. *Essays in probability and statistics* (ed. R. C. Bose, I. M. Chakravarti, P. C. Mahalanobis, C. R. Rao & K. J. C. Smith), 267–292. University of North Carolina Press, Chapel Hill.

Haff, L. R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7**, 374–385.

Haff, L. R. (1979a). Estimation of the inverse covariance matrix: random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7**, 1264–1276.

Haff, L. R. (1979b). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9**, 531–544.

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

Haff, L. R. (1981). Further identities for the Wishart distribution with applications in regression. *Canad. J. Statist.* **9**, 215–224.

Haff, L. R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. *Sankyā Ser. B*, **44**, 245–258.

Holmquist, B. (1985). Moments and cumulants from generating functions of Hilbert space-valued random variables and an application to the Wishart distribution. *TFMS*-3037. Dept. of Math. Statist., University of Lund, Sweden.

Kaufman, G. M. (1967). Some Bayesian moment formulae. Report No. 6710. Centre for Operations Research and Econometrics, Catholic University of Lovain, Heverlee, Belgium.

Magnus, J. R. & Neudecker, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* **7**, 381–394.

Potthoff, R. F. & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

Press, S. J. (1982). *Applied multivariate analysis: Using Bayesian and frequentist methods of inference.* Robert & Krieger, Malabar, Florida.

Rogers, G. S. (1980). *Matrix derivatives.* Marcel Dekker, New York.

von Rosen, D. (1985).Moments and some asymptotic results for maximum likelihood estimates in multivariate linear normal models with special references to the growth curve model. Research Report No. 141, Inst. of Actuar. Mathematics and Math. Statist., University of Stockholm, Sweden.

von Rosen, D. (1987). Moments for matrix normal variables. *Statistics* (accepted for publication).

Siskind, V. (1972). Second moments of inverse Wishart-matrix elements. *Biometrika* **59,** 690–691.

Srivastava, M. S. & Khatri, C. G. (1979). *An introduction to multivariate statistics.* North-Holland, New York.

Received January 1987, in final form January 1988

Dietrich von Rosen, Department of Mathematical Statistics, University of Stockholm, Box 6701, S-11385, Stockholm, Sweden