Mathematical Statistics
Stockholm University

Closed-Form Estimator for the Matrix-Variate Gamma Distribution

Gustav Alfelt

Research Report 2019:16

ISSN 1650-0377

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2019:16**,
http://www.math.su.se

Closed-Form Estimator for the Matrix-Variate Gamma Distribution

Gustav Alfelt*

Department of Mathematics, Stockholm University, Roslagsvägen 101, SE-10691 Stockholm, Sweden

November 2019

Abstract

In this paper we present a novel closed-form estimator for the parameters of the matrix-variate gamma distribution. The estimator relies on the moments of a transformation of the observed matrices, and is compared to the maximum likelihood estimator (MLE) through a simulation study. The study reveals that the suggested estimator outperforms the MLE, in terms of estimation error, when the underlying scale matrix parameter is ill-conditioned or when the shape parameter is close to its lower bound. In addition, since the suggested estimator is closed-form, it does not require numerical optimization as the MLE does, thus needing shorter computation time and is furthermore not subject to start value sensitivity or convergence issues. Finally, using the proposed estimator as start value in the optimization procedure of the MLE is shown to substantially reduce computation time, in comparison to using arbitrary start values.

1 Introduction

The matrix-variate gamma distribution is a generalization of the univariate gamma distribution to the set of positive-definite and symmetric matrices. It is also a more general form of the classical Wishart distribution and a popular approach to model e.g. the stochastic properties of covariance matrices of financial asset returns, which in turn has numerous important applications. For an overview of the matrix-variate gamma distribution and some of its properties, see Gupta and Nagar (2000).

We denote a symmetric and positive definite $p \times p$ matrix $\mathbf{A}$ that follows a matrix-variate gamma distribution with shape parameter α and symmetric scale matrix $\mathbf{\Sigma}$ as $\mathbf{A} \sim MG_p(\alpha, \mathbf{\Sigma})$, where $\alpha > (p-1)/2$, and $\mathbf{\Sigma} > 0$. Let $\mathbf{A}_1, \ldots, \mathbf{A}_n$ be a sample of i.i.d. observations of the matrix-variate gamma distribution, and let $\overline{\mathbf{A}}$ be its sample mean. In general, the maximum likelihood method is the most efficient way of estimating the parameters α and $\mathbf{\Sigma}$, given such a sample. The maximum likelihood estimates of these parameters is obtained by solving the following system of equations for α and $\mathbf{\Sigma}$:

$$\begin{aligned} \psi_p(\alpha) &= \frac{\sum_{k=1}^{n} \ln\left(|\mathbf{A}_k|\right)}{n} - \ln\left(|\mathbf{\Sigma})|\right. \qquad (1) \\ \mathbf{\Sigma} &= \frac{\overline{\mathbf{A}}}{\alpha}, \qquad (2) \end{aligned}$$

*E-mail: gustava@math.su.se

where $\psi_p(\cdot)$ is the multivariate digamma function[1] and $|\cdot|$ denotes the determinant operator. However, as there exists no closed-form inverse of $\psi_p(\cdot)$, the maximum likelihood estimate needs to be computed through numerical optimization. These procedures often have several drawbacks: it can be computationally demanding, increasingly so with matrix dimension p and sample size n; the numerical procedures most often requires a start value, which can influence what parameter value the procedure converges to; there is a risk that the optimization procedure does not converge at all. Apart from the drawbacks associated with the numerical procedure, the maximum likelihood estimator (MLE) for α and $\mathbf{\Sigma}$ tend to be very imprecise when the true parameter value of the scale matrix $\mathbf{\Sigma}$ is ill-conditioned and close to singular, or when α is closer to its lower bound $(p-1)/2$.[2] Such distributions can arise in various situations, for example when estimating covariance matrices based on small sample sizes, or in the presence of multicollinearity.

Several approaches provide alternative estimators for the univariate and multivariate gamma distributions, see for example Ye and Chen (2017) and Vani Lakshmi and Vaidyanathan (2015). Regarding the matrix-variate gamma distribution, e.g. Alfelt (2018) considers parameter estimation under Stein's loss function, when the parameter α is known. In this paper, we suggest a new, closed-form, estimator for the parameters of the matrix-variate gamma distribution. It relies on estimating α based on the relationship between the moments of the diagonal elements of $\mathbf{B}_1, \ldots, \mathbf{B}_n$, where $\mathbf{B}_k$ is a transformation of $\mathbf{A}_k$ for $k = 1, \ldots, n$. Since this estimator is of closed-form, it is not affected by any of the numerical issues discussed above. Further, its precision remains also for ill-conditioned $\mathbf{\Sigma}$'s and α close to $(p-1)/2$, unlike the precision of the MLE. To investigate the relative performance of the different estimators, we conduct a simulation study where the estimation errors of the suggested estimator and the MLE are compared for various α and $\mathbf{\Sigma}$. The benefits of using the proposed estimator as start value in the MLE optimization procedure is also considered.

The rest of this paper is organized as follows: Section 2 presents the new closed-form estimator and derives its asymptotic distribution; Section 3 presents the simulation study; Section 4 provides a general discussion. Finally, proofs of the presented results can be found in the Appendix.

2 New estimator

In this section, we will derive a new, closed-form, estimator of the parameters of the matrix-variate gamma distribution and compute its asymptotic distribution.

From here on, we will consider a sample of n i.i.d. $p \times p$ matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$, where $\mathbf{A}_k \sim MG_p(\alpha, \mathbf{\Sigma})$ for $k = 1, \ldots, n$, where $\alpha > (p-1)/2$, and $\mathbf{\Sigma} > 0$. That is, each matrix $\mathbf{A}_k$ follows a matrix-variate gamma distribution with common shape parameter α and positive-definite scale matrix $\mathbf{\Sigma}$. One way to produce an estimator in closed form is to consider the raw moments of the diagonal elements of the sample matrices $\mathbf{A}_1, \ldots, \mathbf{A}_n$. Denoting the element on row i and column j of $\mathbf{\Sigma}$ as σ_{ij}, we have that $\mathbb{E}[a_{ii}] = \alpha\sigma_{ii}$ and $\mathbb{E}[a_{ii}^2] = \alpha\sigma_{ii}^2 + \alpha^2\sigma_{ii}^2$ for $i = 1, \ldots, p$, since the marginal distribution of a_{ii} is univariate gamma with scale α and shape σ_{ii}. As such,

$$\frac{\mathbb{E}[a_{ii}]^2}{\mathbb{E}[a_{ii}^2] - \mathbb{E}[a_{ii}]^2} = \alpha.$$

By replacing the expected values in the above equation with their corresponding sample means, for each $i = 1, \ldots, p$, and then averaging over each such expression, we can obtain

[1]This is the function $\frac{\partial \ln(\Gamma_p(\alpha))}{\partial \alpha}$, where $\Gamma_p(\cdot)$ is the multivariate gamma function.
[2]The detailed reasons for the imprecision of the MLE in this case remains to be studied closer.

an estimator for α as

$$\hat{\alpha}_a = \frac{1}{p}\sum_{i=1}^{p} \frac{\left(\frac{1}{n}\sum_{k=1}^{n} a_{ii,k}\right)^2}{\left(\frac{1}{n}\sum_{k=1}^{n} a_{ii,k}^2\right) - \left(\frac{1}{n}\sum_{k=1}^{n} a_{ii,k}\right)^2}, \tag{3}$$

where $a_{ii,k}$ is the i:th diagonal element of $\mathbf{A}_k$. Subsequently, we can obtain the estimate of $\mathbf{\Sigma}$ by inserting the above estimator of α into the maximum likelihood equation for $\mathbf{\Sigma}$ presented in Equation (2). As such, we obtain

$$\hat{\mathbf{\Sigma}}_a = \frac{\overline{\mathbf{A}}}{\hat{\alpha}_a}.$$

However, when $\mathbf{\Sigma}$ is a non-diagonal matrix there will be a dependency between the diagonal elements $a_{ii}, i = 1, \ldots, p$, resulting in a dependency between the p terms in $\hat{\alpha}_a$. Hence, with the purpose of decreasing estimator variance, we will consider a transformation of $\mathbf{A}_1, \ldots, \mathbf{A}_n$ that ensures independence of each matrix' diagonal values. To this end, we derive the following theorem, which is an adaptation of Theorem 3.2.10 in Muirhead (1982) to the matrix-variate gamma distribution. The proof of this result can be found in the Appendix.

Theorem 1. *Let* $\mathbf{A} \sim MG_p(\alpha, \mathbf{\Sigma})$, *where* $\alpha > (p-1)/2$ *and* $\mathbf{\Sigma} > 0$ *and assume the following partitions:*

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$ *and* $\mathbf{\Sigma}_{11}$ *are* $q \times q$ *with* $q < p$. *Further, define* $\mathbf{A}_{11\cdot 2} = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$ *and* $\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}$. *Then the following holds*

i) $\mathbf{A}_{11\cdot 2} \sim MG_q(\alpha - (p-q)/2, \mathbf{\Sigma}_{11\cdot 2})$ *and independent of* $\mathbf{A}_{12}$ *and* $\mathbf{A}_{22}$.

ii) The conditional distribution of $\mathbf{A}_{12} \mid \mathbf{A}_{22}$ *is* $N_{q\times(p-q)}(\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{A}_{22}, \frac{1}{2}\mathbf{\Sigma}_{11\cdot 2} \otimes \mathbf{A}_{22})$.

iii) $\mathbf{A}_{22} \sim MG_{p-q}(\alpha, \mathbf{\Sigma}_{22})$.

Here $N_{q\times(p-q)}(\mathbf{M}, \mathbf{V})$ *denotes the matrix-variate normal distribution with mean matrix* $\mathbf{M}$ *and covariance matrix* $\mathbf{V}$.

For an observation $\mathbf{A}$, we will now construct a matrix $\mathbf{B}$ and, with the aid of Theorem 1, show that its diagonal elements are independent. To this end, first define $\mathbf{B}^{(p)} = \mathbf{A}$ and $c = p$, then apply the following algorithm:

Algorithm 1

1) Make the partition

$$\mathbf{B}^{(c)} = \begin{bmatrix} \mathbf{B}_{11}^{(c)} & \mathbf{b}_{12}^{(c)} \\ \mathbf{b}_{21}^{(c)} & b_{cc}^{(c)} \end{bmatrix},$$

where $\mathbf{B}_{11}^{(c)}$ is a $(c-1)\times(c-1)$ matrix, $\mathbf{b}_{12}^{(c)}$ is a $(c-1)\times 1$ vector, $\mathbf{b}_{21}^{(c)} = (\mathbf{b}_{12}^{(c)})'$, and $b_{cc}^{(c)}$ is a scalar.[a]

2) Define $\mathbf{B}^{(c-1)} = \mathbf{B}_{11}^{(c)} - \mathbf{b}_{12}^{(c)}\mathbf{b}_{21}^{(c)}/b_{cc}^{(c)}$.[b]

3) Decrease c by one.

4) Repeat steps 1) to 3) above until $c = 1$.

5) Finally set

$$\mathbf{B} = \begin{bmatrix} b_{11}^{(1)} & \mathbf{b}_{12}^{(2)} & \cdots & \mathbf{b}_{12}^{(p-1)} & \mathbf{b}_{12}^{(p)} \\ (\mathbf{b}_{21}^{(2)})' & b_{22}^{(2)} & & & \\ \vdots & & \ddots & & \\ (\mathbf{b}_{21}^{(p-1)})' & & & b_{(p-1)(p-1)}^{(p-1)} & \\ (\mathbf{b}_{21}^{(p)})' & & & & b_{pp}^{(p)} \end{bmatrix}.$$

[a] As such, $\mathbf{B}_{11}^{(c)}$ contains all the elements of $\mathbf{B}^{(c)}$ except the elements in the last row and the last column.

[b] As such, $\mathbf{B}^{(c-1)}$ is an $(c-1)\times(c-1)$ matrix.

Further, for ease of notation, let, for $i = 1, \ldots, p$,

$$\alpha^{(i)} = \alpha - \frac{(p-i)}{2}.$$

Also, let $\mathbf{\Sigma}^{(p)} = \mathbf{\Sigma}$ and,

$$\begin{aligned} \mathbf{\Sigma}^{(i-1)} &= \mathbf{\Sigma}_{11}^{(i)} - \mathbf{\Sigma}_{12}^{(i)} \left(\mathbf{\Sigma}_{22}^{(i)}\right)^{-1} \mathbf{\Sigma}_{21}^{(i)}, \quad \text{with} \\ \mathbf{\Sigma}^{(i)} &= \begin{bmatrix} \mathbf{\Sigma}_{11}^{(i)} & \mathbf{\Sigma}_{12}^{(i)} \\ \mathbf{\Sigma}_{21}^{(i)} & \mathbf{\Sigma}_{22}^{(i)} \end{bmatrix} \end{aligned}$$

and where $\mathbf{\Sigma}^{(i)} = (\sigma)_{lm}^{(i)}$ for $1 \le l \le i, 1 \le m \le i$.

By Theorem 1 *i*), we have that $\mathbf{B}^{(i)}$ is independent of $\mathbf{b}_{12}^{(j)}$, $\mathbf{b}_{21}^{(j)}$ and $b_{jj}^{(j)}$, for $i = 1, \ldots, p-1$ and $j = i+1, \ldots, p$. Further, in line with *i*) in Theorem 1 we have that

$$\mathbf{B}^{(i)} \sim MG_i(\alpha^{(i)}, \mathbf{\Sigma}^{(i)}).$$

Subsequently, the marginal distribution of $b_{ii}^{(i)}$ will be univariate gamma with shape parameter $\alpha^{(i)}$ and scale parameter $\sigma_{ii}^{(i)}$. Finally, by *ii*), the conditional distribution of $\mathbf{b}_{12}^{(i)}$ given $b_{ii}^{(i)}$ will be multivariate normal, since it is a matrix of size $(i-1)\times 1$, i.e. a vector. Specifically,

$$\mathbf{b}_{12}^{(i)} \mid b_{ii}^{(i)} \sim N_{(i-1)\times 1}\left(\mathbf{\Sigma}_{12}^{(i)} \frac{b_{ii}^{(i)}}{\sigma_{ii}^{(i)}}, \mathbf{\Sigma}^{(i-1)} \frac{b_{ii}^{(i)}}{2}\right). \tag{4}$$

We will refer to the element of $\mathbf{B}$ on row l and column m as b_{lm}. In accordance with the above, note that b_{lm} is independent of all elements in $\mathbf{B}$ except the elements $b_{\max(l,m)i}$ and $b_{i\max(l,m)}$, $i = 1, \ldots, \max(l, m)$. Further, all diagonal elements b_{ii}, $i = 1 \ldots, p$, follows a univariate gamma distribution with shape parameter $\alpha^{(i)}$.

Now, given the distribution of $b_{ii}, i = 1, \ldots, p$, we obtain that

$$\frac{\mathbb{E}\left[b_{ii}\right]^2}{\mathbb{E}\left[b_{ii}^2\right] - \mathbb{E}\left[b_{ii}\right]^2} + \frac{p-i}{2} = \alpha.$$

This relationship, together with the mutual independence of $b_{ii}, i = 1, \ldots, p$, allows us to develop a new version of the estimator presented by equation (3) as

$$\begin{aligned}
\hat{\alpha}_b &= \frac{1}{p}\sum_{i=1}^{p}\left[\frac{\left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}\right)^2}{\left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}^2\right) - \left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}\right)^2} + \frac{p-i}{2}\right] \\
&= \frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{\left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}\right)^2}{\left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}^2\right) - \left(\frac{1}{n}\sum_{k=1}^{n} b_{ii,k}\right)^2}, \qquad (5)
\end{aligned}$$

where $b_{ii,k}$ is the element on row i, column i of the matrix $\mathbf{B}_k$, which in turn is a transformation of the observed matrix $\mathbf{A}_k$, $k = 1, \ldots, n$, where the transformation is computed according to Algorithm 1. Similar to Equation (2), $\boldsymbol{\Sigma}$ will be estimated by

$$\hat{\boldsymbol{\Sigma}}_b = \frac{\overline{\mathbf{A}}}{\hat{\alpha}_b}. \qquad (6)$$

We now present the asymptotic distribution for $\hat{\alpha}_b$ and $\hat{\boldsymbol{\Sigma}}_b$. In Theorem 2, let $s = p(p-1)/2$, $\mathbf{y}$ be an $(s \times 1)$ vector and $\mathbf{Z}$ be an $s \times s$ matrix. Further, let $h(u)$ be a mapping from the u:th row in the vector $\text{vech}(\mathbf{M})$ to the row and column indexation of the equivalent element in $\mathbf{M}$, $u = 1, \ldots, s$. If for example $p = 4$, then $h(6) = (3, 2)$, referring to the element m_{32} in $\mathbf{M}$. Moreover, by Lemma 1 in the Appendix, we have that

$$\begin{aligned}
\mathbb{C}\left[a_{lm}, b_{ii}\right] &= \begin{cases} \alpha^{(i)}\sigma_{il}^{(i)}\sigma_{mi}^{(i)}, & \text{if } \max(l,m) \le i \\ 0, & \text{if } \max(l,m) > i \end{cases} \\
\mathbb{C}\left[a_{lm}, b_{ii}^2\right] &= \begin{cases} 2\left(\alpha^{(i)}+1\right)\alpha^{(i)}\sigma_{ii}^{(i)}\sigma_{il}^{(i)}\sigma_{mi}^{(i)}, & \text{if } \max(l,m) \le i \\ 0, & \text{if } \max(l,m) > i \end{cases},
\end{aligned}$$

where $\mathbb{C}[X, Y]$ denotes the covariance between X and Y.

Theorem 2. *Let $\mathbf{A}_1, \ldots, \mathbf{A}_n$ be an i.i.d. sample, where $\mathbf{A}_k \sim MG_p(\alpha, \boldsymbol{\Sigma}), \alpha > (p-1)/2$ and $\boldsymbol{\Sigma} > 0$, for each $k = 1, \ldots, n$. Then $(\hat{\alpha}_b, \hat{\boldsymbol{\Sigma}}_b)$, defined by Equation (5) and (6), is a consistent estimator for $(\alpha, \boldsymbol{\Sigma})$. Further,*

$$\sqrt{n}\left(\begin{bmatrix} \hat{\alpha}_b \\ \text{vech}\left(\hat{\boldsymbol{\Sigma}}_b\right) \end{bmatrix} - \begin{bmatrix} \alpha \\ \text{vech}\left(\boldsymbol{\Sigma}\right) \end{bmatrix}\right) \xrightarrow{d} N_{(s+1)\times 1}(\mathbf{0}, \mathbf{C})$$

as $n \to \infty$. Here

$$\mathbf{C} = \begin{pmatrix} x & \mathbf{y} \\ \mathbf{y}' & \mathbf{Z} \end{pmatrix}$$

where, for $1 \leq u \leq v \leq s$,

$$\begin{aligned}
x &= \frac{2}{p^2}\sum_{i=1}^{p}\left[\left(\alpha^{(i)}\right)^2 + \alpha^{(i)}\right] \\
y_u &= \sum_{i=1}^{p}\left[\frac{2}{p\alpha\sigma_{ii}^{(i)}}(1+\alpha^{(i)})\,\mathbb{C}\left[a_{h(u)}, b_{ii}\right] - \frac{\mathbb{C}\left[a_{h(u)}, b_{ii}^2\right]}{p\alpha(\sigma_{ii}^{(i)})^2} - \frac{2\sigma_{h(u)}\alpha^{(i)}(\alpha^{(i)}+1)}{p^2\alpha}\right]. \\
z_{uv} &= \frac{\mathbb{C}[a_{h(u)}, a_{h(v)}]}{\alpha^2} + \sum_{i=1}^{p}\frac{2\sigma_{h(u)}\sigma_{h(v)}}{p^2\alpha^2}(\alpha^{(i)}+1)\alpha^{(i)} \\
&\quad - \sum_{i=1}^{p}\frac{2(\alpha^{(i)}+1)}{p\alpha^2\sigma_{ii}^{(i)}}\left(\sigma_{h(v)}\,\mathbb{C}\left[a_{h(u)}, b_{ii}\right] + \sigma_{h(u)}\,\mathbb{C}\left[a_{h(v)}, b_{ii}\right]\right) \\
&\quad + \sum_{i=1}^{p}\frac{1}{p\alpha^2\left(\sigma_{ii}^{(i)}\right)^2}\left(\sigma_{h(v)}\,\mathbb{C}\left[a_{h(u)}, b_{ii}^2\right] + \sigma_{h(u)}\,\mathbb{C}\left[a_{h(v)}, b_{ii}^2\right]\right).
\end{aligned}$$

The proof of Theorem 2 can be found in the Appendix. The finite-sample properties of the proposed estimator $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$, and how they compare to the finite-sample properties of the MLE will be studied in the next section.

3 Simulation study

As suggested in Section 1 the MLE is in general the most suitable estimator of the matrix-variate gamma parameters. However, for certain regions of the parameter space, the estimation error of the estimator proposed in Section 2, given a finite sample of observations, tend to be significantly smaller. In particular, it seems that $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ is much more accurate when α is close to its lower bound $(p-1)/2$, or when the scale matrix $\mathbf{\Sigma}$ is so called ill-conditioned. Such matrices are characterized by a large ratio between the largest and smallest eigenvalue of the matrix, and are close to singular. This will be illustrated with a simulation study, comparing the estimates $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ and $(\hat{\alpha}_{ML}, \hat{\mathbf{\Sigma}}_{ML})$ for various parameter values and sample sizes. A small comparison between $\hat{\alpha}_b$ and $\hat{\alpha}_a$ is also presented. Furthermore, even in cases where the MLE is preferable, the suggested estimator is useful. By applying the proposed estimator $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ as a start value in the numerical optimization procedure for the MLE, the computation time can be reduced substantially, in comparison to using an arbitrary start value. This will also be illustrated by simulation.

In order to compare the magnitude of estimation error between the estimators, for N samples, define for a parameter θ and an estimator of θ denoted $\hat{\theta}$,

$$\begin{aligned}
\mathrm{MSE}\left[\hat{\theta}\right] &= \frac{1}{N}\sum_{j=1}^{N}\left(\hat{\theta}_i - \theta\right)^2, \quad r_\alpha = \frac{\mathrm{MSE}[\hat{\alpha}_{ML}]}{\mathrm{MSE}[\hat{\alpha}_b]} \\
\mathrm{F}\left[\hat{\theta}\right] &= \frac{1}{N}\sum_{j=1}^{N}||\hat{\theta}_i - \theta||, \quad r_{\mathbf{\Sigma}} = \frac{\mathrm{F}\left[\hat{\mathbf{\Sigma}}_{ML}\right]}{\mathrm{F}\left[\hat{\mathbf{\Sigma}}_b\right]}
\end{aligned}$$

where $\hat{\theta}_i$ is the estimate based on sample i and where $||\mathbf{M}||$ denote the Frobenius norm[3] of the matrix $\mathbf{M}$. As such, when r_α takes on values above 1, $\hat{\alpha}_b$ has a lower mean squared error (MSE) than $\hat{\alpha}_{ML}$, and similarly when $r_{\mathbf{\Sigma}}$ is larger than 1, the estimation error of $\hat{\mathbf{\Sigma}}_b$ is smaller than that of $\hat{\mathbf{\Sigma}}_{ML}$, in terms of the Frobenius norm. Further, let r_c denote

[3]See e.g. page 71 in Golub and Loan (2013).

the computation time for the MLE divided by the computation time of the suggested estimator.

In the following simulation study, the parameter $\mathbf{\Sigma}$ is not fixed; instead the eigenvalues of $\mathbf{\Sigma}$ is fixed, and the matrices of eigenvectors is randomly generated according to the Haar distribution[4]. To this end, denote the p eigenvalues of $\mathbf{\Sigma}$ as $\lambda_1 \geq \ldots \geq \lambda_p > 0$ and the $p \times p$ matrix of eigenvectors as $\mathbf{H}$. Now, the simulation procedure is conducted as follows:

1) Generate $\mathbf{H}$ according to the Haar distribution.
2) Compute $\mathbf{\Sigma}= \mathbf{H}'\mathbf{L}\mathbf{H}$, where $\mathbf{L} = \text{diag}(\lambda_1, \ldots, \lambda_p)$.
3) Generate $\mathbf{A}_1, \ldots, \mathbf{A}_n$, where $\mathbf{A}_k \sim MG_p(\alpha, \mathbf{\Sigma}), k = 1 \ldots, n$.
4) Compute $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ and $(\hat{\alpha}_{ML}, \hat{\mathbf{\Sigma}}_{ML})$.
5) Repeat step 1) to 4) N times.

The procedure above is implemented for various values of n, p, and α. To account for various condition numbers on $\mathbf{\Sigma}$, we set $\lambda_1 = 10^d$, $\lambda_p = 10^{-d}$ and $\lambda_i = 10^d(p-i)/p + 10^{-d}i/p$, where $i = 2, \ldots, p-1$, for different values of d.

Table 1 compares the above defined measures, and mean values for estimates of α, for the MLE and the new suggested closed-form estimator, for $N = 1000$ and different values of n, p, d and α. Note that the lower bound for α is 0.5 when $p = 2$ and 2 when $p = 5$. Values of r_α, $r_{\mathbf{\Sigma}}$ and r_c above one is emphasized in bold, indicating a better performance for the closed-form estimator. Note here that the cases with $d = 0$ corresponds to $\mathbf{\Sigma}$ being the identity matrix with probability one, while the cases with $d = 7$ corresponds to $\mathbf{\Sigma}$ being nearly singular and ill-conditioned. In the cases when $d = 0$ and α is not close to its lower bound, the MLE performs better, as the values of r_α and $r_{\mathbf{\Sigma}}$ are below one. Requiring numerical optimization, however, the MLE computation time is longer. When $d = 0$ and α is close to its lower bound ($\alpha = 0.501$ for $p = 2$ and $\alpha = 2.001$ for $p = 5$), however, the suggested estimator tends to perform better for larger sample sizes, indicated by and r_α and $r_{\mathbf{\Sigma}}$ above one for sample sizes $n = 1000$ and $n = 10000$. For the cases when $d = 7$, r_α and $r_{\mathbf{\Sigma}}$ are very large, substantially favouring the closed-form estimator. In terms of estimation error, the closed-form estimator performs similar in the case of $d = 0$ and $d = 7$, suggesting that the estimation error of the MLE increases drastically when the scale matrix $\mathbf{\Sigma}$ is ill-conditioned. On a general note, r_α and $r_{\mathbf{\Sigma}}$ tend to increase with increasing α (barring when α is close to its lower bound) and sample size n, but decrease with when the matrix dimension increases. Also, for the cases when $d = 7$, the MLE tends to have a substantial upward bias. In addition, further simulations suggest that the relative performance of the closed-form estimator increases further with an increasing d.

Moreover, Table 2 compares the performance between $(\hat{\alpha}_a, \hat{\mathbf{\Sigma}}_a)$ and $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$. It suggests that they perform similarly when $d = 0$, but that $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ has a lower estimation error when $d = 7$. This is not unexpected, since $d = 0$ represents the special case when there is no dependencies between the diagonal elements of the observed matrices, and the motivation for the estimator $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ was to decrease the estimate variance in the presence of such dependencies.

Furthermore, even in cases where observations are generated with parameters that favors the MLE, the suggested closed-form estimator can be useful. The numerical procedure used to find the MLE requires some start value, and the choice of this start value will influence the computation time of the procedure. By inserting the closed-form estimate as start value in the numerical optimization procedure, it is possible to decrease computation time in comparison to using an arbitrary start values. To illustrate this, a simulation study is conducted as follows:

[4]See Definition 4.5.1 on page 161 in Andersen (2003).

1) Generate $\mathbf{\Sigma}, \mathbf{\Sigma}_{RS} \sim MG_p(p^2, \mathbf{I}_p)$ and $\alpha, \alpha_{RS} \sim U(\frac{p-1}{2}, 100\frac{p-1}{2})$.[5]
2) Generate $\mathbf{A}_1, \ldots, \mathbf{A}_n$, where $\mathbf{A}_k \sim MG_p(\alpha, \mathbf{\Sigma}), k = 1 \ldots, n$.
3) Compute the MLE for $\mathbf{A}_1, \ldots, \mathbf{A}_n$ using $(\alpha_{RS}, \mathbf{\Sigma}_{RS})$ as start value for the numerical optimization procedure[6] and record the computation time.
4) Compute $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ for $\mathbf{A}_1, \ldots, \mathbf{A}_n$ and then compute the MLE using $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ as start value and record the computation time for the two steps combined.
5) Repeat steps 1)-4) 100 times.

As such, the true parameters and random start values are drawn from the same distribution. The summed computation time for the two different approaches, in seconds, is presented in Table 3, for different values of the matrix dimension p, and where $n = 1000$. The results suggest that in each of the considered cases, it is beneficial to use the proposed estimator as start value, compared to using an arbitrary start value. This advantage increases substantially as the matrix dimension grows large.

$p = 2$		$d = 0$					$d = 7$				
	n	$\bar{\hat{\alpha}}_{ML}$	$\bar{\hat{\alpha}}_b$	r_α	$r_{\mathbf{\Sigma}}$	r_c	$\bar{\hat{\alpha}}_{ML}$	$\bar{\hat{\alpha}}_b$	r_α	$r_{\mathbf{\Sigma}}$	r_c
	100	0.54	0.54	0.54	0.96	**47**	2.4	0.54	$\mathbf{4.7 \cdot 10^4}$	**2.4**	**29**
$\alpha = 0.501$	1000	0.54	0.51	**8.3**	**1.6**	**113**	2.9	0.51	$\mathbf{9.6 \cdot 10^5}$	**7.9**	**76**
	10000	0.54	0.50	**93**	**4.5**	**376**	2.0	0.50	$\mathbf{6.4 \cdot 10^6}$	**26**	**308**
	100	102	103	0.59	0.81	**28**	105	103	**2.1**	1.0	**46**
$\alpha = 100$	1000	100	100	0.67	0.83	**107**	103	100	**22**	**2.6**	**111**
	10000	100	100	0.73	0.86	**503**	104	100	**290**	**8.6**	**504**
	100	1012	1030	0.62	0.80	**37**	1568	1022	**775**	**1.5**	**56**
$\alpha = 1000$	1000	1001	1003	0.69	0.84	**113**	1548	1000	$\mathbf{1.7 \cdot 10^4}$	**5.0**	**131**
	10000	1000	1000	0.68	0.84	**501**	1423	998	$\mathbf{5.1 \cdot 10^4}$	**13**	**505**

$p = 5$		$d = 0$					$d = 7$				
	n	$\bar{\hat{\alpha}}_{ML}$	$\bar{\hat{\alpha}}_b$	r_α	$r_{\mathbf{\Sigma}}$	r_c	$\bar{\hat{\alpha}}_{ML}$	$\bar{\hat{\alpha}}_b$	r_α	$r_{\mathbf{\Sigma}}$	r_c
	100	2.0	2.1	0.08	0.95	**138**	2.2	2.1	**2.2**	**1.1**	**47**
$\alpha = 2.001$	1000	2.0	2.0	**1.2**	1.0	**119**	2.2	2.0	**37**	**2.3**	**58**
	10000	2.0	2.0	**16**	**1.5**	**87**	2.2	2.0	**403**	**6.9**	**80**
	100	101	103	0.27	0.62	**26**	101	103	0.7	0.71	**51**
$\alpha = 100$	1000	100	100	0.37	0.68	**50**	101	100	**6.1**	**1.7**	**65**
	10000	100	100	0.35	0.65	**77**	101	100	**59**	**4.8**	**84**
	100	1011	1029	0.29	0.57	**33**	1148	1031	**61**	**1.8**	**44**
$\alpha = 1000$	1000	1001	1003	0.32	0.58	**49**	1137	1002	**3515**	**5.6**	**65**
	10000	1000	1001	0.32	0.58	**83**	1119	999	**6517**	**17.5**	**83**

Table 1: *A comparison of estimation errors between the estimators* $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$ *and* $(\hat{\alpha}_{ML}, \hat{\mathbf{\Sigma}}_{ML})$*, for various combinations of* p, d, n *and* α*. Values of* r_α*,* $r_{\mathbf{\Sigma}}$ *and* r_c *that are larger than one is emphasized in bold, indicating a superior performance of the closed-form estimator* $(\hat{\alpha}_b, \hat{\mathbf{\Sigma}}_b)$*.*

[5] Here $U(l, u)$ denotes the uniform distribution with bounds l and u.
[6] Applying the Broyden-Fletcher-Goldfarb-Shanno algorithm.

		$p=2$				$p=5$			
		$d=0$		$d=7$		$d=0$		$d=7$	
	n	$\frac{\text{MSE}(\hat\alpha_a)}{\text{MSE}(\hat\alpha_b)}$	$\frac{\text{F}(\hat{\mathbf{\Sigma}}_a)}{\text{F}(\hat{\mathbf{\Sigma}}_b)}$	$\frac{\text{MSE}(\hat\alpha_a)}{\text{MSE}(\hat\alpha_b)}$	$\frac{\text{F}(\hat{\mathbf{\Sigma}}_a)}{\text{F}(\hat{\mathbf{\Sigma}}_b)}$	$\frac{\text{MSE}(\hat\alpha_a)}{\text{MSE}(\hat\alpha_b)}$	$\frac{\text{F}(\hat{\mathbf{\Sigma}}_a)}{\text{F}(\hat{\mathbf{\Sigma}}_b)}$	$\frac{\text{MSE}(\hat\alpha_a)}{\text{MSE}(\hat\alpha_b)}$	$\frac{\text{F}(\hat{\mathbf{\Sigma}}_a)}{\text{F}(\hat{\mathbf{\Sigma}}_b)}$
	100	**1.58**	**1.04**	**2.94**	**1.18**	**2.04**	**1.05**	**2.29**	**1.07**
$\alpha \gtrapprox \frac{p-1}{2}$	1000	**1.85**	**1.06**	**3.56**	**1.26**	**2.17**	**1.04**	**2.54**	**1.07**
	10000	**2.00**	**1.07**	**3.77**	**1.27**	**2.33**	**1.04**	**2.71**	**1.08**
	100	0.94	0.99	**1.87**	**1.40**	0.90	0.97	0.98	1.00
$\alpha = 100$	1000	1.00	1.00	**2.05**	**1.46**	1.00	1.00	**1.20**	**1.09**
	10000	1.00	1.00	**2.08**	**1.43**	**1.03**	**1.01**	**1.16**	**1.07**
	100	0.94	0.99	**2.00**	**1.41**	0.89	0.96	0.97	**1.01**
$\alpha = 1000$	1000	0.99	1.00	**2.16**	**1.47**	0.98	0.99	**1.08**	**1.05**
	10000	1.00	1.00	**1.75**	**1.31**	1.00	1.00	**1.21**	**1.13**

Table 2: *A comparison of estimation errors between the estimators $(\hat\alpha_a, \hat{\mathbf{\Sigma}}_a)$ and $(\hat\alpha_b, \hat{\mathbf{\Sigma}}_b)$, for various combinations of p, d, n and α. Values that are larger than one is emphasized in bold, indicating a superior performance of the closed-form estimator $(\hat\alpha_b, \hat{\mathbf{\Sigma}}_b)$.*

	p			
Start value	2	5	10	50
$(\hat\alpha_{RS}, \hat{\mathbf{\Sigma}}_{RS})$	8	26	148	22340
$(\hat\alpha_b, \hat{\mathbf{\Sigma}}_b)$	5	6	9	368

Table 3: *A comparison of aggregated computation times, in seconds, for the MLE using random start values, $(\hat\alpha_{RS}, \hat{\mathbf{\Sigma}}_{RS})$, and using the closed-form estimator $(\hat\alpha_b, \hat{\mathbf{\Sigma}}_b)$ as start value.*

4 Discussion

In this paper, we have presented a new closed-form estimator for the parameters of the matrix-variate gamma distribution. Its performance in terms of estimation error and computation time is compared to the performance of the maximum likelihood estimator of said parameters. First off, since the estimator is of closed-form, it does not struggle with any issues the MLE is subject to due to requiring a numerical optimization procedure. Second, the MLE tends to be very imprecise when $\mathbf{\Sigma}$ is ill-conditioned or when α is close to its lower bound, an issue that our suggested estimator does not have.

The simulation study in Section 3 reveals that the presented closed-form estimator outperforms the MLE, in terms of estimation error, when parameters are such that the generated observations are near-singular. Further, in these cases, the relative performance of the closed-form estimator increases with sample size and with increasing α. Similar results occur when d increases, indicating an increased condition number for the scale matrix $\mathbf{\Sigma}$. On the other hand, its relative performance tends to decrease as matrix dimension p increases. In addition, not requiring any numerical procedures, the suggested estimator has a considerably shorter computation time in all considered cases. Further simulations presented in Section 3 reveal that the computation time for the numerical procedure of the MLE can be substantially reduced by using the proposed estimator as start value, in comparison to using arbitrary start values. This illustrates that the closed-form estimator is important even in cases where the MLE has a lower MSE.

A question that arises is in what region of parameter space, more specifically, that the presented estimator outperforms the MLE, which is something that the considered simulation studies only hint at. Another topic concerns how to discern from a sample which of the two estimation procedures that is most appropriate, using for example the sample mean condition number. It would also be of interest to specify the bias of the estimators in order to correct for it. These important question should be considered in future research.

Appendix

Lemma 1. *Let* $\mathbf{A} \sim MG_p(\alpha, \mathbf{\Sigma})$ *and let* $\mathbf{B}$ *be defined as in Algorithm 1 in Section 2. Then, for* $i = \ldots, p$,

$$\begin{aligned}
\mathbb{C}\left[a_{lm}, b_{ii}\right] &= \begin{cases} \alpha^{(i)}\sigma_{il}^{(i)}\sigma_{mi}^{(i)}, & \textit{if } \max(l,m) \leq i \\ 0, & \textit{if } \max(l,m) > i \end{cases}, \\
\mathbb{C}\left[a_{lm}, b_{ii}^2\right] &= \begin{cases} 2\left(\alpha^{(i)}+1\right)\alpha^{(i)}\sigma_{ii}^{(i)}\sigma_{il}^{(i)}\sigma_{mi}^{(i)}, & \textit{if } max(l,m) \leq i \\ 0, & \textit{if } max(l,m) > i \end{cases}.
\end{aligned}$$

Proof. First, denote the element on row l, column m of the matrix $\mathbf{B}^{(i)}$ as $b_{lm}^{(i)}$, and note that $b_{ii} = b_{ii}^{(i)}$. Then, we have

$$\begin{aligned}
\mathbb{C}\left[a_{lm}, b_{ii}\right] &= \mathbb{C}\left[a_{lm}, b_{ii}^{(i)}\right] \\
&= \mathbb{E}\left[a_{lm}b_{ii}^{(i)}\right] - \mathbb{E}\left[a_{lm}\right]\mathbb{E}\left[b_{ii}^{(i)}\right].
\end{aligned}$$

Now, in accordance with Algorithm 1, we can write, for any $q = 2, \ldots, p$,

$$\begin{aligned}
b_{lm}^{(q-1)} &= b_{lm}^{(q)} - \frac{b_{qm}^{(q)}b_{lq}^{(q)}}{b_{qq}^{(q)}}. \\
b_{lm}^{(q)} &= b_{lm}^{(q-1)} + \frac{b_{qm}^{(q)}b_{lq}^{(q)}}{b_{qq}^{(q)}}. \qquad (7)
\end{aligned}$$

Further note that $a_{lm} = b_{lm}^{(p)}$, and by repeatedly use the relationship shown in Equation (7) we obtain

$$\begin{aligned}
a_{lm} &= b_{lm}^{(p)} \\
&= b_{lm}^{(p-1)} + \frac{b_{pm}^{(p)}b_{lp}^{(p)}}{b_{pp}^{(p)}} \\
&= b_{lm}^{(l)} - \sum_{q=l+1}^{p}\frac{b_{qm}^{(q)}b_{lq}^{(q)}}{b_{qq}^{(q)}}.
\end{aligned}$$

Now, suppose $l > i$. Then

$$\begin{aligned}
\mathbb{E}\left[a_{lm}b_{ii}^{(i)}\right] &= \mathbb{E}\left[\left(b_{lm}^{(l)} - \sum_{q=l+1}^{p}\frac{b_{qm}^{(q)}b_{lq}^{(q)}}{b_{qq}^{(q)}}\right)b_{ii}^{(i)}\right] \\
&= \mathbb{E}\left[\left(b_{lm}^{(l)} - \sum_{q=l+1}^{p}\frac{b_{qm}^{(q)}b_{lq}^{(q)}}{b_{qq}^{(q)}}\right)\right]\mathbb{E}\left[b_{ii}^{(i)}\right] \\
&= \mathbb{E}\left[a_{lm}\right]\mathbb{E}\left[b_{ii}^{(i)}\right]
\end{aligned}$$

since $b_{ii}^{(i)}$ is independent of all the terms and factors within the parenthesis, in accordance with the results stated following Algorithm 1. Since $\mathbf{A}$ is symmetric, the above holds also if $m > i$, or put differently, if $\max(l, m) > i$. As such, if $\max(l, m) > i$ we have that

$$\begin{aligned}
\mathbb{C}\left[a_{lm}, b_{ii}^{(i)}\right] &= \mathbb{E}\left[a_{lm} b_{ii}^{(i)}\right] - \mathbb{E}\left[a_{lm}\right] \mathbb{E}\left[b_{ii}^{(i)}\right] \\
&= \mathbb{E}\left[a_{lm}\right] \mathbb{E}\left[b_{ii}^{(i)}\right] - \mathbb{E}\left[a_{lm}\right] \mathbb{E}\left[b_{ii}^{(i)}\right] \\
&= 0.
\end{aligned}$$

If, on the contrary, $l \leq i$, we write

$$a_{lm} = b_{lm}^{(i)} - \sum_{q=i+1}^{p} \frac{b_{qm}^{(q)} b_{lq}^{(q)}}{b_{qq}^{(q)}}.$$

Further,

$$\begin{aligned}
\mathbb{E}\left[a_{lm} b_{ii}^{(i)}\right] &= \mathbb{E}\left[b_{ii}^{(i)} b_{lm}^{(i)}\right] - \mathbb{E}\left[b_{ii}^{(i)} \sum_{q=i+1}^{p} \frac{b_{qm}^{(q)} b_{lq}^{(q)}}{b_{qq}^{(q)}}\right] \\
&= \mathbb{E}\left[b_{ii}^{(i)} b_{lm}^{(i)}\right] - \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[\sum_{q=i+1}^{p} \frac{b_{qm}^{(q)} b_{lq}^{(q)}}{b_{qq}^{(q)}}\right] \\
&= \mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right] + \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[b_{lm}^{(i)}\right] - \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[\sum_{q=i+1}^{p} \frac{b_{qm}^{(q)} b_{lq}^{(q)}}{b_{qq}^{(q)}}\right] \\
&= \mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right] + \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[b_{lm}^{(i)} - \sum_{q=i+1}^{p} \frac{b_{qm}^{(q)} b_{lq}^{(q)}}{b_{qq}^{(q)}}\right] \\
&= \mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right] + \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[a_{lm}\right] \qquad (8)
\end{aligned}$$

where the second equality holds since $b_{ii}^{(i)}$ is independent of all the elements $b_{qm}^{(q)}, b_{lq}^{(q)}$ and $b_{qq}^{(q)}, q = i+1, \ldots, p$. With the aid of Equation (8) we can obtain

$$\begin{aligned}
\mathbb{C}\left[a_{lm}, b_{ii}^{(i)}\right] &= \mathbb{E}\left[a_{lm} b_{ii}^{(i)}\right] - \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[a_{lm}\right] \\
&= \mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right] + \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[a_{lm}\right] - \mathbb{E}\left[b_{ii}^{(i)}\right] \mathbb{E}\left[a_{lm}\right] \\
&= \mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right].
\end{aligned}$$

Note again that since $\mathbf{A}$ is symmetric, this results holds whenever $\max(l, m) \leq i$. Further, since $\mathbf{B}^{(i)} \sim MG_p(\alpha^{(i)}, \mathbf{\Sigma}^{(i)})$, in accordance with the covariance of the matrix-variate gamma distribution, we have that $\mathbb{C}\left[b_{ii}^{(i)}, b_{lm}^{(i)}\right] = \alpha^{(i)} \sigma_{li}^{(i)} \sigma_{mi}^{(i)}$, completing the proof for the first expression in the Lemma.

Following the same reasoning as above, whenever $\max(l, m) > i$, we have that $\mathbb{C}\left[a_{lm}, b_{ii}^2\right] = 0$. When $\max(l, m) \leq i$, again following the above reasoning, we get

$$\mathbb{C}\left[a_{lm}, \left(b_{ii}^{(i)}\right)^2\right] = \mathbb{C}\left[b_{im}^{(i)}, \left(b_{ii}^{(i)}\right)^2\right]$$

Further, consider first the case when $l = i$. In accordance with Equation (4), we have that

$$b_{im}^{(i)} \mid b_{ii}^{(i)} \sim N\left(\frac{\sigma_{im}^{(i)} b_{ii}^{(i)}}{\sigma_{ii}^{(i)}}, \frac{\sigma_{im}^{(i-1)} b_{ii}^{(i)}}{2}\right).$$

As such, we have

$$\begin{aligned}
\mathbb{C}\left[b_{im}^{(i)}, \left(b_{ii}^{(i)}\right)^2\right] &= \mathbb{E}\left[b_{im}^{(i)}\left(b_{ii}^{(i)}\right)^2\right] - \mathbb{E}\left[b_{im}^{(i)}\right]\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right] \\
&= \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2 \mathbb{E}\left[b_{im}^{(i)} \mid b_{ii}^{(i)}\right]\right] - \mathbb{E}\left[b_{im}^{(i)}\right]\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right] \\
&= \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3 \frac{\sigma_{im}^{(i)}}{\sigma_{ii}^{(i)}}\right] - \mathbb{E}\left[b_{im}^{(i)}\right]\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right] \\
&= \left(\alpha^{(i)}+2\right)\left(\alpha^{(i)}+1\right)\alpha^{(i)}\left(\sigma_{ii}^{(i)}\right)^2\sigma_{im}^{(i)} - \left(\alpha^{(i)}+1\right)\left(\alpha^{(i)}\right)^2\left(\sigma_{ii}^{(i)}\right)^2\sigma_{im}^{(i)} \\
&= 2\left(\alpha^{(i)}+1\right)\alpha^{(i)}\left(\sigma_{ii}^{(i)}\right)^2\sigma_{im}^{(i)},
\end{aligned}$$

where the fourth equality is due to the moments of the gamma distribution. Note that the equivalent result holds if instead $m = i$, and also if $l = m = i$. If instead $l < i$, then

$$\begin{aligned}
\mathbb{E}\left[b_{lm}^{(i)}\left(b_{ii}^{(i)}\right)^2\right] &= \mathbb{E}\left[\left(b_{lm}^{(i-1)} + \frac{b_{il}^{(i)}b_{mi}^{(i)}}{b_{ii}^{(i)}}\right)\left(b_{ii}^{(i)}\right)^2\right] \\
&= \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2 b_{lm}^{(i-1)}\right] + \mathbb{E}\left[b_{ii}^{(i)}b_{il}^{(i)}b_{mi}^{(i)}\right] \\
&= \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i-1)}\right] + \mathbb{E}\left[b_{ii}^{(i)}\,\mathbb{E}\left[b_{il}^{(i)}b_{mi}^{(i)} \mid b_{ii}^{(i)}\right]\right]. \qquad (9)
\end{aligned}$$

Now, from Equation (4), we can obtain that

$$\begin{aligned}
\mathbb{C}\left[b_{il}^{(i)}, b_{mi}^{(i)} \mid b_{ii}^{(i)}\right] &= \frac{\sigma_{lm}^{(i-1)}}{2} b_{ii}^{(i)}, \\
\mathbb{E}\left[b_{il}^{(i)} \mid b_{ii}^{(i)}\right] &= \frac{\sigma_{il}^{(i)}}{\sigma_{ii}^{(i)}} b_{ii}^{(i)} \quad \text{and} \\
\mathbb{E}\left[b_{mi}^{(i)} \mid b_{ii}^{(i)}\right] &= \frac{\sigma_{mi}^{(i)}}{\sigma_{ii}^{(i)}} b_{ii}^{(i)}.
\end{aligned}$$

As such,

$$\begin{aligned}
\mathbb{E}\left[b_{il}^{(i)}b_{mi}^{(i)} \mid b_{ii}^{(i)}\right] &= \mathbb{C}\left[b_{il}^{(i)}, b_{mi}^{(i)} \mid b_{ii}^{(i)}\right] + \mathbb{E}\left[b_{il}^{(i)} \mid b_{ii}^{(i)}\right]\mathbb{E}\left[b_{mi}^{(i)} \mid b_{ii}^{(i)}\right] \\
&= \frac{\sigma_{lm}^{(i-1)}}{2} b_{ii}^{(i)} + \frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\left(b_{ii}^{(i)}\right)^2
\end{aligned}$$

Inserting this into Equation (9) yields

$$\mathbb{E}\left[b_{lm}^{(i)}\left(b_{ii}^{(i)}\right)^2\right] = \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i-1)}\right] + \frac{\sigma_{lm}^{(i-1)}}{2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right] + \frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right].$$

Further, as

$$\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i)}\right] = \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i-1)}\right] + \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[\frac{b_{il}^{(i)}b_{mi}^{(i)}}{b_{ii}^{(i)}}\right],$$

we have that

$$
\begin{aligned}
\mathbb{C}\left[b_{im}^{(i)},\left(b_{ii}^{(i)}\right)^2\right] &= \mathbb{E}\left[b_{lm}^{(i)}\left(b_{ii}^{(i)}\right)^2\right]-\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i)}\right]\\
&= \mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i-1)}\right]+\frac{\sigma_{lm}^{(i-1)}}{2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\\
&\quad -\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{lm}^{(i-1)}\right]-\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[\frac{b_{il}^{(i)}b_{mi}^{(i)}}{b_{ii}^{(i)}}\right]\\
&= \frac{\sigma_{lm}^{(i-1)}}{2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[\frac{1}{b_{ii}^{(i)}}\mathbb{E}\left[b_{il}^{(i)}b_{mi}^{(i)}\mid b_{ii}^{(i)}\right]\right]\\
&= \frac{\sigma_{lm}^{(i-1)}}{2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\\
&\quad -\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[\frac{1}{b_{ii}^{(i)}}\left(\frac{\sigma_{lm}^{(i-1)}}{2}b_{ii}^{(i)}+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\left(b_{ii}^{(i)}\right)^2\right)\right]\\
&= \frac{\sigma_{lm}^{(i-1)}}{2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\left(\frac{\sigma_{lm}^{(i-1)}}{2}+\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[b_{ii}^{(i)}\right]\right)\\
&= \frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\mathbb{E}\left[b_{ii}^{(i)}\right]\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\\
&= \frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\left(\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^3\right]-\mathbb{E}\left[\left(b_{ii}^{(i)}\right)^2\right]\mathbb{E}\left[b_{ii}^{(i)}\right]\right)\\
&= \frac{\sigma_{il}^{(i)}\sigma_{mi}^{(i)}}{\left(\sigma_{ii}^{(i)}\right)^2}\left(\left(\alpha^{(i)}+2\right)\left(\alpha^{(i)}+1\right)\alpha^{(i)}\left(\sigma_{ii}^{(i)}\right)^3-\left(\alpha^{(i)}+1\right)\left(\alpha^{(i)}\right)^2\left(\sigma_{ii}^{(i)}\right)^3\right)\\
&= 2\left(\alpha^{(i)}+1\right)\alpha^{(i)}\sigma_{ii}^{(i)}\sigma_{il}^{(i)}\sigma_{mi}^{(i)}.
\end{aligned}
$$

The equivalent result holds when $m < i$ and as such whenever $\max(l,m) < i$. This confirms the second part of the Lemma, which completes the proof. □

Proof of Theorem 1:

Proof. Observing that the density of $\mathbf{A}$ is

$$f(\mathbf{A})=\frac{|\mathbf{\Sigma}|^{-\alpha}}{\Gamma_p(\alpha)}|\mathbf{A}|^{\alpha-(p+1)/2}\exp\left(\mathrm{tr}\left(-\mathbf{\Sigma}^{-1}\mathbf{A}\right)\right),$$

the proof of Theorem 1 follows directly from the proof of Theorem 3.2.10 in Muirhead (1982). □

Proof of Theorem 2:

Proof. Let $\mathbf{m}(\mathbf{v})$ denote the mean for n samples of the random vector $\mathbf{v}$, $\mathbf{a}=\mathrm{vech}(\mathbf{A})$, $\mathbf{b}=(b_{11},\ldots,b_{pp})$, and $\mathbf{b}^2=(b_{11}^2,\ldots,b_{pp}^2)$. In accordance with the Central limit theorem, we have

$$\sqrt{n}\left(\mathbf{m}\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix}-\mathbb{E}\left[\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix}\right]\right)\xrightarrow{\mathrm{d}}N_{(2p+s)\times 1}(\mathbf{0},\mathbf{K}_1),$$

where

$$\mathbf{K}_1 = \begin{pmatrix} \mathbf{C} & \mathbf{D} & \mathbb{C}[\mathbf{a},\mathbf{b}] \\ \mathbf{D} & \mathbf{E} & \mathbb{C}[\mathbf{a},\mathbf{b}^2] \\ \mathbb{C}[\mathbf{a},\mathbf{b}] & \mathbb{C}[\mathbf{a},\mathbf{b}^2] & \mathbb{C}[\mathbf{a},\mathbf{a}] \end{pmatrix}. \tag{10}$$

Due to the independence of the elements in $\mathbf{b}$ and $\mathbf{b}^2$, the $p \times p$ matrices $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{E}$ are all diagonal, and in accordance with the moments of the gamma distribution,

$$\begin{aligned} c_{ii} &= \alpha^{(i)} \left(\sigma_{ii}^{(i)}\right)^2 \\ d_{ii} &= 2\left(\alpha^{(i)}+1\right)\alpha^{(i)}\left(\sigma_{ii}^{(i)}\right)^3 \\ e_{ii} &= 2\left(\alpha^{(i)}+1\right)\left(2\alpha^{(i)}+3\right)\alpha^{(i)}\left(\sigma_{ii}^{(i)}\right)^4, \end{aligned}$$

for $i = 1, \ldots, p$. The elements of the $s \times p$ covariance matrices $\mathbb{C}[\mathbf{a},\mathbf{b}]$ are given by Lemma 1. Further, $\mathbb{C}[\mathbf{a},\mathbf{a}]$ is the covariance matrix for the elements in the matrix-variate gamma distribution. For the elements a_{lm} and a_{uv} in $\mathbf{A}$,

$$\mathbb{C}[a_{lm}, a_{uv}] = \frac{\alpha}{2}(\sigma_{lu}\sigma_{mv} + \sigma_{lv}\sigma_{mu}).$$

To show the result stated in the theorem, the Delta method will be applied. To this end, define the $(2p+s)\times 1 \to (s+1)\times 1$ vector function $\mathbf{g}((x_1,\ldots,x_p,y_1,\ldots,y_p,z_1,\ldots,z_s)')$ as

$$\begin{aligned} g_1((x_1,\ldots,x_p,y_1,\ldots,y_p,z_1,\ldots,z_s)') &= \frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{x_i^2}{y_i - x_i^2}, \\ g_{u+1}((x_1,\ldots,x_p,y_1,\ldots,y_p,z_1,\ldots,z_s)') &= z_u\left(\frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{x_i^2}{y_i - x_i^2}\right)^{-1}, \end{aligned}$$

for $u = 1, \ldots, s$. Note that

$$\begin{aligned} \mathbf{g}\left(\mathbf{m}\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right) &= \begin{pmatrix} \hat{\alpha}_b \\ \mathrm{vech}(\hat{\mathbf{\Sigma}}_b) \end{pmatrix}, \quad \text{and} \\ \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right]\right) &= \begin{pmatrix} \alpha \\ \mathrm{vech}(\mathbf{\Sigma}) \end{pmatrix}. \end{aligned}$$

From the Delta method, we now have that

$$\sqrt{n}\left(\mathbf{g}\left(\mathbf{m}\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right) - \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right]\right)\right) \xrightarrow{\mathrm{d}} N_{(s+1)\times 1}(\mathbf{0}, \mathbf{K}),$$

where

$$\mathbf{K} = \nabla\mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right]\right)\mathbf{K}_1 \nabla\mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix} \mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a} \end{pmatrix}\right]\right)',$$

where $\nabla\mathbf{g}(\mathbb{E}[(\mathbf{b},\mathbf{b}^2,\mathbf{a})'])$ is the $(s+1)\times(2p+s)$ gradient of $\mathbf{g}$ at the point $\mathbb{E}[(\mathbf{b},\mathbf{b}^2,\mathbf{a})']$. For the elements of $\nabla\mathbf{g}(\mathbb{E}[(\mathbf{b},\mathbf{b}^2,\mathbf{a})'])$, we have, for $i = 1\ldots,p$, and $v = 1,\ldots,s$, with the argument dropped for easier notation,

$$\nabla \mathbf{g}_{1i} = \frac{\partial g_1}{\partial x_i} = \frac{2}{p}\left(\frac{x_i}{y_i - x_i^2} + \frac{x_i^3}{\left(y_i - x_i^2\right)^2}\right) = \frac{2}{p\sigma_{ii}^{(i)}}\left(1+\alpha^{(i)}\right) := h_i, \tag{11}$$

$$\nabla \mathbf{g}_{1(p+i)} = \frac{\partial g_1}{\partial y_i} = -\frac{x_i^2}{p\left(y_i - x_i^2\right)^2} = -\frac{1}{p\left(\sigma_{ii}^{(i)}\right)^2} := l_i \tag{12}$$

$$\nabla \mathbf{g}_{1(2p+v)} = \frac{\partial g_1}{\partial z_v} = 0. \tag{13}$$

Further, for $u = 1, \ldots, s$,

$$\begin{aligned}\nabla \mathbf{g}_{(u+1)i} &= \frac{\partial g_{u+1}}{\partial x_i} = -z_u\left(\frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{x_i^2}{y_i - x_i^2}\right)^{-2}\frac{\partial g_1}{\partial x_i} = \\ &= -\frac{2z_u}{p\alpha^2\sigma_{ii}^{(i)}}\left(\alpha^{(i)}+1\right) := m_{ui}\end{aligned} \tag{14}$$

$$\begin{aligned}\nabla \mathbf{g}_{(u+1)(p+i)} &= \frac{\partial g_{u+1}}{\partial y_i} = -z_u\left(\frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{x_i^2}{y_i - x_i^2}\right)^{-2}\frac{\partial g_1}{\partial y_i} \\ &= \frac{z_u}{p\alpha^2\left(\sigma_{ii}^{(i)}\right)^2} := o_{ui}\end{aligned} \tag{15}$$

$$\nabla \mathbf{g}_{(u+1)(2p+v)} = \frac{\partial g_{u+1}}{\partial z_v} = \begin{cases}\left(\frac{p-1}{4} + \frac{1}{p}\sum_{i=1}^{p}\frac{x_i^2}{y_i - x_i^2}\right)^{-1} = \frac{1}{\alpha}, & \text{if } v = u \\ 0, & \text{if } v \neq u\end{cases} := f_{uv}. \tag{16}$$

Now, let $\mathbf{h}$ and $\mathbf{l}$ be $p \times 1$ vectors consisting of the elements h_i and l_i, $i = 1, \ldots, p$, respectively. Further, let $\mathbf{M}$ be an $s \times p$ matrix consisting of the elements m_{ui}, let $\mathbf{O}$ be an $s \times p$ matrix consisting of the elements o_{ui}, and $\mathbf{F}$ be an $s \times s$ matrix consisting of the elements f_{uv}, for $i = 1, \ldots, p$ and $u, v = 1 \ldots, s$. Note here that $\mathbf{F} = \frac{1}{\alpha}\mathbf{I}_s$. We now have

$$\nabla \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix}\mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a}\end{pmatrix}\right]\right) = \begin{pmatrix}\mathbf{h}' & \mathbf{l}' & \mathbf{0}_s' \\ \mathbf{M} & \mathbf{O} & \mathbf{F}\end{pmatrix},$$

where $\mathbf{0}_s$ is an $s \times 1$ vector of zeros. We can now compute the asymptotic covariance matrix $\mathbf{K}$ as

$$\begin{aligned}\mathbf{K} &= \nabla \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix}\mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a}\end{pmatrix}\right]\right)\mathbf{K}_1 \nabla \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix}\mathbf{b} \\ \mathbf{b}^2 \\ \mathbf{a}\end{pmatrix}\right]\right)' \\ &= \begin{pmatrix}\mathbf{h}' & \mathbf{l}' & \mathbf{0}_s' \\ \mathbf{M} & \mathbf{O} & \mathbf{F}\end{pmatrix}\begin{pmatrix}\mathbf{C} & \mathbf{D} & \mathbb{C}[\mathbf{a},\mathbf{b}] \\ \mathbf{D} & \mathbf{E} & \mathbb{C}[\mathbf{a},\mathbf{b}^2] \\ \mathbb{C}[\mathbf{a},\mathbf{b}] & \mathbb{C}[\mathbf{a},\mathbf{b}^2] & \mathbb{C}[\mathbf{a},\mathbf{a}]\end{pmatrix}\begin{pmatrix}\mathbf{h}' & \mathbf{l}' & \mathbf{0}_s' \\ \mathbf{M} & \mathbf{O} & \mathbf{F}\end{pmatrix}' \\ &:= \begin{pmatrix}x & \mathbf{y} \\ \mathbf{y} & \mathbf{Z}\end{pmatrix},\end{aligned}$$

where

$$\begin{aligned}x &= \mathbf{h}'\mathbf{C}\mathbf{h} + \mathbf{l}'\mathbf{D}\mathbf{h} + \mathbf{h}'\mathbf{D}\mathbf{l} + \mathbf{l}'\mathbf{E}\mathbf{l} \\ \mathbf{y} &= \mathbf{h}'\mathbf{C}\mathbf{M}' + \mathbf{l}'\mathbf{D}\mathbf{M}' + \mathbf{h}'\mathbf{D}\mathbf{O}' + \mathbf{l}'\mathbf{E}\mathbf{O}' + \mathbf{h}'\,\mathbb{C}[\mathbf{a},\mathbf{b}]'\mathbf{F}' + \mathbf{l}'\,\mathbb{C}[\mathbf{a},\mathbf{b}^2]'\mathbf{F}' \\ \mathbf{Z} &= \mathbf{M}\mathbf{C}\mathbf{M}' + \mathbf{O}\mathbf{D}\mathbf{M}' + \mathbf{F}\,\mathbb{C}[\mathbf{a},\mathbf{b}]\mathbf{M}' + \mathbf{M}\mathbf{D}\mathbf{O}' + \mathbf{O}\mathbf{E}\mathbf{O}' + \mathbf{F}\,\mathbb{C}[\mathbf{a},\mathbf{b}^2]\mathbf{O}' \\ &\quad + \mathbf{M}\,\mathbb{C}[\mathbf{a},\mathbf{b}]'\mathbf{F}' + \mathbf{O}\,\mathbb{C}[\mathbf{a},\mathbf{b}^2]'\mathbf{F}' + \mathbf{F}\,\mathbb{C}[\mathbf{a},\mathbf{a}]\mathbf{F}'\end{aligned}$$

Inserting the values from Equation (10) and Equations (11) to (16) and carrying out the computations, we finally obtain, for the elements in x, $\mathbf{y}$ and $\mathbf{Z}$, with $u, v = 1, \ldots, s$,

$$\begin{aligned}
x &= \frac{2}{p^2}\sum_{i=1}^{p}\left[\left(\alpha^{(i)}\right)^2 + \alpha^{(i)}\right] \\
y_u &= \sum_{i=1}^{p}\left[\frac{2}{p\alpha\sigma_{ii}^{(i)}}(1+\alpha^{(i)})\,\mathbb{C}\left[a_{h(u)}, b_{ii}\right] - \frac{\mathbb{C}\left[a_{h(u)}, b_{ii}^2\right]}{p\alpha(\sigma_{ii}^{(i)})^2} - \frac{2\sigma_{h(u)}\alpha^{(i)}(\alpha^{(i)}+1)}{p^2\alpha}\right]. \\
z_{uv} &= \frac{\mathbb{C}[a_{h(u)}, a_{h(v)}]}{\alpha^2} + \sum_{i=1}^{p}\frac{2\sigma_{h(u)}\sigma_{h(v)}}{p^2\alpha^2}(\alpha^{(i)}+1)\alpha^{(i)} \\
&\quad - \sum_{i=1}^{p}\frac{2(\alpha^{(i)}+1)}{p\alpha^2\sigma_{ii}^{(i)}}\left(\sigma_{h(v)}\,\mathbb{C}\left[a_{h(u)}, b_{ii}\right] + \sigma_{h(u)}\,\mathbb{C}\left[a_{h(v)}, b_{ii}\right]\right) \\
&\quad + \sum_{i=1}^{p}\frac{1}{p\alpha^2\left(\sigma_{ii}^{(i)}\right)^2}\left(\sigma_{h(v)}\,\mathbb{C}\left[a_{h(u)}, b_{ii}^2\right] + \sigma_{h(u)}\,\mathbb{C}\left[a_{h(v)}, b_{ii}^2\right]\right),
\end{aligned}$$

which was to be shown. Finally, since

$$\mathbf{m}\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix} \xrightarrow{\text{p}} \mathbb{E}\left[\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix}\right],$$

we have through the continuous mapping theorem that

$$\begin{aligned}
\mathbf{g}\left(\mathbf{m}\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix}\right) &\xrightarrow{\text{p}} \mathbf{g}\left(\mathbb{E}\left[\begin{pmatrix}\mathbf{b}\\ \mathbf{b}^2\\ \mathbf{a}\end{pmatrix}\right]\right) \\
\begin{pmatrix}\hat{\alpha}_b\\ \text{vech}(\hat{\mathbf{\Sigma}}_b)\end{pmatrix} &\xrightarrow{\text{p}} \begin{pmatrix}\alpha\\ \text{vech}(\mathbf{\Sigma})\end{pmatrix},
\end{aligned}$$

showing that $(\hat{\alpha}_b, \text{vech}(\hat{\mathbf{\Sigma}}_b))$ is a consistent estimator for $(\alpha, \text{vech}(\mathbf{\Sigma}))$, which completes the proof. □

References

Alfelt, G. (2018). Stein-Haff Identity for the Exponential Family. *Theory of Probability and Mathematical Statistics*, 99(2):7–18.

Andersen, T. W. (2003). *An Introduction to Mulivariate Statistical Analysis.* Wiley.

Golub, G. and Loan, C. V. (2013). *Matrix Computations. 4th Edition.* Johns Hopkins University Press, Baltimore.

Gupta, A. K. and Nagar, D. K. (2000). *Matrix variate distributions.* CRC Press.

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

Vani Lakshmi, R. and Vaidyanathan, V. (2015). Parameter Estimation in Multivariate Gamma Distribution. *Statistics, Optimization and Information Computing*, 3:147–159.

Ye, Z.-S. and Chen, N. (2017). Closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician*, 71(2):177–181.