Working Papers in Statistics
No 2016:4

Department of Statistics
School of Economics and Management
Lund University

Central limit theorems for functionals of large dimensional sample covariance matrix and mean vector in matrix-variate skewed model

TARAS BODNAR, STOCKHOLM UNIVERSITY

STEPAN MAZUR, LUND UNIVERSITY

NESTOR PAROLYA, LEIBNIZ UNIVERSITY OF HANNOVER

Central limit theorems for functionals of large dimensional sample covariance matrix and mean vector in matrix-variate skewed model

TARAS BODNAR[a,*], STEPAN MAZUR[b], NESTOR PAROLYA[c] [1]

[a] *Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden*

[b] *Department of Statistics, Lund University, SE-22007 Lund, Sweden*

[c] *Institute of Empirical Economics, Leibniz University of Hannover, D-30167 Hannover, Germany*

Abstract

In this paper we consider the asymptotic distributions of functionals of the sample covariance matrix and the sample mean vector obtained under the assumption that the matrix of observations has a matrix variate general skew normal distribution. The central limit theorem is derived for the product of the sample covariance matrix and the sample mean vector. Moreover, we consider the product of an inverse covariance matrix and the mean vector for which the central limit theorem is established as well. All results are obtained under the large dimensional asymptotic regime where the dimension p and sample size n approach to infinity such that $p/n \to c \in (0, 1)$.

ASM Classification: 62H10, 62E15, 62E20, 60F05, 60B20

Keywords: Skew normal distribution, large dimensional asymptotics, stochastic representation, random matrix theory.

[1*] Corresponding author. E-mail address: taras.bodnar@math.su.se. The second author appreciates the financial support of the Swedish Research Council Grant Dnr: 2013-5180 and Riksbankens Jubileumsfond Grant Dnr: P13-1024:1

1 Introduction

The functions of the sample covariance matrix and the sample mean vector appear in various statistical applications. The classical improvement techniques for the mean estimation have already been discussed by Stein (1956) and Jorion (1986). In particular, Efron (2006) constructed confidence regions of smaller volume than the standard spheres for the mean vector of a multivariate normal distribution. Fan et al. (2008), Bai and Shi (2011), Bodnar and Gupta (2011), Cai and Zhou (2012), Cai and Yuan (2012), Fan et al. (2013), Bodnar et al. (2014a), Bodnar et al. (2016a), Wang et al. (2015) among others suggested improved techniques for the estimation of covariance matrix and precision matrix (the inverse of covariance matrix).

In our work we introduce the family of matrix variate general skew normal (MVGSN) distributions which is a generalization of the models considered by Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Azzalini (2005), Liseo and Loperfido (2003, 2006), Bartoletti and Loperfido (2010), Loperfido (2010), Christiansen and Loperfido (2014), Adcock et al. (2015), De Luca and Loperfido (2015) and among others. Under the assumption of MVGSN we consider the expressions for the sample mean vector $\overline{\mathbf{x}}$ and the sample covariance matrix $\mathbf{S}$. In particulary, we deal with two products $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ and $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}$ where $\mathbf{l}$ is a non-zero vector of constants. It is noted that this kind of expressions has not been intensively considered in the literature, although they are present in numerous important applications. The first application of the products arises in the portfolio theory, where the vector of optimal portfolio weights is proportional to $\mathbf{S}^{-1}\overline{\mathbf{x}}$. The second application is in the discriminant analysis where the coefficients of the discriminant function are expressed as a product of the inverse sample covariance matrix and the difference of the sample mean vectors. In Bayesian context it is highly related to the product $\mathbf{S}\overline{\mathbf{x}}$.

Bodnar and Okhrin (2011) derived the exact distribution of the product of the inverse sample covariance matrix and the sample mean vector under the assumption of normality, while Kotsiuba and Mazur (2015) obtained its asymptotic distribution as well as its approximate density based on the Gaussian integral and the third order Taylor series expansion. Moreover, Bodnar et al. (2013, 2014b) analyzed the product of the sample (singular) covariance matrix and the sample mean vector. In the present paper, we contribute to the existing literature by deriving the central limit theorems (CLTs) under the introduced class of matrix variate distribution, i.e., general matrix variate skew normality in the case of the high-dimensional observation matrix. Under the considered family of distributions, the columns of the observation matrix are not independent anymore and, thus, the CLTs cover more general class of random matrices.

Nowadays, modern scientific data include large number of sample points which is often comparable to the number of features (dimension) and so the sample covariance matrix and the sample mean vector are not the efficient estimators anymore. For example, stock markets include a large number of companies which is often close to the number of available time points. In order to understand better the statistical properties of the traditional estimators and tests based on high-dimensional settings, it is of interest to study the asymptotic distribution of the above mentioned bilinear forms involving the sample covariance matrix and the sample mean vector.

The appropriate central limit theorems, which do not suffer from the "curse of dimensionality" and do not reduce the number of dimensions, are of great interest for high-dimensional statistics because more efficient estimators and tests may be constructed and applied in practice. The classical multivariate procedures are based on the central limit theorems assuming that the dimension p is fixed and the sample size n increases. However, numerous authors provide quite reasonable proofs that this assumption does not lead to precise distributional approximations for commonly used statistics, and that under increasing dimension asymptotics the better approximations can be obtained [see, e.g., Bai and Silverstein (2004) and references therein]. Technically speaking, under the high-dimensional asymptotics we understand the case when the sample size n and the dimension p tend to infinity, such that their ratio p/n converges to some positive constant c (here we assume that $c < 1$). Under this condition the well-known Marchenko-Pastur and Silverstein's equations were derived [see, Marčenko and Pastur (1967), Silverstein (1995)].

The rest of the paper is structured as follows. In Section 2 we introduce a semi-parametric matrix-variate family of skewed distributions. Main results are given in Section 3, where we derive the central limit theorems under high-dimensional asymptotic regime of the sample (inverse) covariance matrix and the sample mean vector under the MVGSN distribution. Section 4 presents a short numerical study in order to verify the obtained analytic results.

2 Semi-parametric matrix-variate family of skewed distributions

In this section we introduce a family of matrix-variate skewed distributions which generalizes the skew normal distribution.

Let

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{p1} & \dots & x_{pn} \end{pmatrix} = (\mathbf{x}_1, ..., \mathbf{x}_n) \, ,$$

be the $p \times n$ observation matrix where $\mathbf{x}_j$ is the j^{th} observation vector. In the following, we assume that the random matrix $\mathbf{X}$ possesses a stochastic representation given by

$$\mathbf{X} \stackrel{d}{=} \mathbf{Y} + \mathbf{B}\boldsymbol{\nu}\mathbf{1}_n^T, \tag{1}$$

where $\mathbf{Y} \sim \mathcal{N}_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^T, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)$ ($p \times n$-dimensional matrix-variate normal distribution with mean matrix $\boldsymbol{\mu}\mathbf{1}_n^T$ and covariance matrix $\boldsymbol{\Sigma} \otimes \mathbf{I}_n$), $\boldsymbol{\nu}$ is a q-dimensional random vector with continuous density function $f_{\boldsymbol{\nu}}(\cdot)$, $\mathbf{B}$ is a $p \times q$ matrix of constants. Further, it is assumed that $\mathbf{Y}$ and $\boldsymbol{\nu}$ are independently distributed. If random matrix $\mathbf{X}$ follows model (1) then we say that $\mathbf{X}$ is generalized matrix-variate skew-normal distributed with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\mathbf{B}$, and $f_{\boldsymbol{\nu}}(\cdot)$. The first three parameters are finite dimensional, while the third parameter is infinite dimensional. This makes model (1) to be of a semi-parametric type. The assertion we denote by $\mathbf{X} \sim \mathcal{SN}_{p,n;q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}})$. If $f_{\boldsymbol{\nu}}$ can be parametrized by finite dimensional parameter $\boldsymbol{\theta}$ then model (1) reduces to a parametrical model which is denoted by $\mathbf{X} \sim \mathcal{SN}_{p,n;q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{B}; \boldsymbol{\theta})$. If $n = 1$ then we use the notation $\mathcal{SN}_{p;q}(\cdot, \cdot, \cdot; \cdot)$ instead of $\mathcal{SN}_{p,1;q}(\cdot, \cdot, \cdot; \cdot)$.

From (1) the density function of $\mathbf{X}$ is expressed as

$$f_{\mathbf{X}}(\mathbf{Z}) = \int_{\mathbb{R}^q} f_{N_{p,n}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n)}(\mathbf{Z} - \mathbf{B}\boldsymbol{\nu}^* \mathbf{1}_n^T | \boldsymbol{\nu} = \boldsymbol{\nu}^*) f_{\boldsymbol{\nu}}(\boldsymbol{\nu}^*) \mathbf{d}\boldsymbol{\nu}^*. \tag{2}$$

Let $C = \Phi_q(\mathbf{0}; -\boldsymbol{\xi}, \boldsymbol{\Omega})$. In a special case when $\boldsymbol{\nu} = |\boldsymbol{\psi}|$ is the vector formed by the absolute values of every element in $\boldsymbol{\psi}$ where $\boldsymbol{\psi} \sim \mathcal{N}_q(\boldsymbol{\xi}, \boldsymbol{\Omega})$, i.e. $\boldsymbol{\nu}$ has a q-variate truncated normal distribution, we get

Proposition 1. *Assume model (1). Let $\boldsymbol{\nu} = |\boldsymbol{\psi}|$ with $\boldsymbol{\psi} \sim \mathcal{N}_q(\boldsymbol{\xi}, \boldsymbol{\Omega})$. Then the density function of $\mathbf{X}$ is given by*

$$f_{\mathbf{X}}(\mathbf{Z}) = \widetilde{C}^{-1} \Phi_q \left(\mathbf{0}; -\mathbf{D}[\mathbf{E}\mathbf{vec}(\mathbf{Z} - \boldsymbol{\mu}\mathbf{1}_n^T) + \boldsymbol{\Omega}^{-1}\boldsymbol{\xi}], \mathbf{D} \right) \phi_{pn} \left(\mathbf{vec}(\mathbf{Z} - \boldsymbol{\mu}\mathbf{1}_n^T); \mathbf{F}\mathbf{E}^T \mathbf{D}\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}, \mathbf{F} \right) \tag{3}$$

where $\mathbf{D} = (n\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B} + \boldsymbol{\Omega}^{-1})^{-1}$, $\mathbf{E} = \mathbf{1}_n^T \otimes \mathbf{B}^T\boldsymbol{\Sigma}^{-1}$, $\mathbf{F} = (\mathbf{I}_n \otimes \boldsymbol{\Sigma}^{-1} - \mathbf{E}^T\mathbf{D}\mathbf{E})^{-1}$, and

$$\widetilde{C}^{-1} = C^{-1} \frac{|\mathbf{F}|^{1/2}|\mathbf{D}|^{1/2}}{|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{ -\frac{1}{2} \left[\boldsymbol{\xi}^T\boldsymbol{\Omega}^{-1}(-\mathbf{D} + \boldsymbol{\Omega} - \mathbf{D}\mathbf{E}\mathbf{F}\mathbf{E}^T\mathbf{D})\boldsymbol{\Omega}^{-1}\boldsymbol{\xi} \right] \right\}.$$

The proof of Proposition 1 is presented in the Appendix.

It is remarkable that model (1) includes several skew-normal distributions considered by Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Azzalini (2005). For example, in case of $n = 1$, $q = 1$, $\boldsymbol{\mu} = \mathbf{0}$, $\mathbf{B} = \boldsymbol{\Delta}\mathbf{1}_p$, and $\boldsymbol{\Sigma} = (\mathbf{I}_p - \boldsymbol{\Delta}^2)^{1/2}\boldsymbol{\Psi}(\mathbf{I}_p - \boldsymbol{\Delta}^2)^{1/2}$ we get

$$\mathbf{X} \stackrel{d}{=} (\mathbf{I}_p - \boldsymbol{\Delta}^2)^{1/2}\mathbf{v}_0 + \boldsymbol{\Delta}\mathbf{1}_p|v_1|, \tag{4}$$

where $\mathbf{v}_0 \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi})$ and $v_1 \sim \mathcal{N}(0,1)$ are independently distributed; $\boldsymbol{\Psi}$ is a correlation matrix and $\boldsymbol{\Delta} = diag(\delta_1, ..., \delta_p)$ with $\delta_j \in (-1,1)$. Model (4) was previously introduced by Azzalini (2005).

3 CLTs for expressions involving the sample covariance matrix and the sample mean vector

The sample estimators for the mean vector and the covariance matrix are given by

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i = \frac{1}{n}\mathbf{X}\mathbf{1}_n \quad \text{and} \quad \mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T = \mathbf{X}\mathbf{V}\mathbf{X}^T,$$

where $\mathbf{V} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ is a symmetric idempotent matrix, i.e., $\mathbf{V} = \mathbf{V}^T$ and $\mathbf{V}^2 = \mathbf{V}$.

The following theorem shows that $\overline{\mathbf{x}}$ and $\mathbf{S}$ are independently distributed and presents their marginal distributions under model (1). The results of Theorem 1 show that the independence of $\overline{\mathbf{x}}$ and $\mathbf{S}$ could not be used as a characterization property of a multivariate normal distribution if the observation vectors in data matrix are dependent.

Theorem 1. *Let* $\mathbf{X} \sim \mathcal{SN}_{p,n;q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}})$ *with* $p < n-1$*. Then*

(a) $(n-1)\mathbf{S} \sim \mathcal{W}_p(n-1, \boldsymbol{\Sigma})$ *(p-dimensional Wishart distribution with* $(n-1)$ *degrees of freedom and covariance matrix* $\boldsymbol{\Sigma}$*),*

(b) $\overline{\mathbf{x}} \sim \mathcal{SN}_{p;q}\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}}\right)$*,*

(c) $\mathbf{S}$ *and* $\overline{\mathbf{x}}$ *are independently distributed.*

Proof. Let $\mathbf{X}^* \stackrel{d}{=} \mathbf{X}|\boldsymbol{\nu} = \boldsymbol{\nu}^*$, $\overline{\mathbf{x}}^* \stackrel{d}{=} \overline{\mathbf{x}}|\boldsymbol{\nu} = \boldsymbol{\nu}^*$, and $\mathbf{S}^* \stackrel{d}{=} \mathbf{S}|\boldsymbol{\nu} = \boldsymbol{\nu}^*$. Because $\mathbf{Y}$ and $\boldsymbol{\nu}$ are independent we get that

$$\mathbf{X}^* \stackrel{d}{=} \mathbf{X}|\boldsymbol{\nu} = \boldsymbol{\nu}^* \sim \mathcal{N}_{p,n}((\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*)\mathbf{1}_n^T, \boldsymbol{\Sigma} \otimes \mathbf{I}_n).$$

From Theorem 3.1.2 of Muirhead (1982) we obtain that

$$\begin{aligned}
\overline{\mathbf{x}}^* &\sim \mathcal{N}_p\left(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}^*, \frac{1}{n}\boldsymbol{\Sigma}\right),\\
\mathbf{S}^* &\sim \mathcal{W}_p(n-1,\boldsymbol{\Sigma}),
\end{aligned}$$

and $\mathbf{S}^*$ and $\overline{\mathbf{x}}^*$ are independent. Hence, it follows that

$$\begin{aligned}
f_{\overline{\mathbf{x}},\mathbf{S}}(\widetilde{\mathbf{x}},\widetilde{\mathbf{S}}) &= \int_{\mathbb{R}_+^q} f_{\mathbf{S}|\boldsymbol{\nu}=\boldsymbol{\nu}^*}(\widetilde{\mathbf{S}}) f_{\overline{\mathbf{x}}|\boldsymbol{\nu}=\boldsymbol{\nu}^*}(\widetilde{\mathbf{x}}) f_{\boldsymbol{\nu}}(\boldsymbol{\nu}^*) \mathbf{d}\boldsymbol{\nu}^*\\
&= f_{\mathbf{S}}(\widetilde{\mathbf{S}}) \int_{\mathbb{R}_+^q} f_{\overline{\mathbf{x}}|\boldsymbol{\nu}=\boldsymbol{\nu}^*}(\widetilde{\mathbf{x}}) f_{\boldsymbol{\nu}}(\boldsymbol{\nu}^*) \mathbf{d}\boldsymbol{\nu}^*, \qquad (5)
\end{aligned}$$

where the last equality follows from the fact that the density of $\mathbf{S}^*$ does not depend on $\boldsymbol{\nu}^*$.

From (5) we directly get that $\overline{\mathbf{x}}$ and $\mathbf{S}$ are independent; $\mathbf{S}$ is Wishart distributed with $(n-1)$ degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$; $\overline{\mathbf{x}}$ is generalized skew-normal distributed with parameters $\boldsymbol{\mu}$, $\frac{1}{n}\boldsymbol{\Sigma}$, $\mathbf{B}$ and $f_{\boldsymbol{\nu}}$. The theorem is proved. □

For the validity of the asymptotic results presented in Sections 3.1 and 3.2 we need the following two conditions

(A1) Let $(\lambda_i, \mathbf{u}_i)$ denote the set of eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$. We assume that there exist m_1 and M_1 such that

$$0 < m_1 \le \lambda_1 \le \lambda_2 \le ... \le \lambda_p \le M_1 < \infty$$

uniformly on p.

(A2) There exists M_2 such that

$$|\mathbf{u}_i^T\boldsymbol{\mu}| \le M_2 \;\text{ and }\; |\mathbf{u}_i^T\mathbf{b}_j| \le M_2 \;\text{ for all }\; i=1,...,p \;\text{ and }\; j=1,...,q$$

uniformly on p where $\mathbf{b}_j$, $j=1,...,q$, are the columns of $\mathbf{B}$.

In more general terms, we say that an arbitrary p-dimensional vector $\mathbf{l}$ satisfies the condition (A2) if $|\mathbf{u}_i^T\mathbf{l}| \le M_2$ for all $i=1,\dots,p$.

Assumption (A1) is a classical condition in random matrix theory (see, Bai and Silverstein (2004)), which bounds the spectrum of $\boldsymbol{\Sigma}$ from below as well as from above. Assumption (A2) is a technical one. In combination with (A1) this condition ensures that $p^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$, $p^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $p^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}$, $p^{-1}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-3}\boldsymbol{\mu}$, as well as that all the diagonal elements of $\mathbf{B}^T\boldsymbol{\Sigma}\mathbf{B}$, $\mathbf{B}^T\boldsymbol{\Sigma}^3\mathbf{B}$, and

$\mathbf{B}^T\mathbf{\Sigma}^{-1}\mathbf{B}$ are uniformly bounded. All these quadratic forms are used in the statements and the proofs of our results. Note that the constants appearing in the inequalities will be denoted by M_2 and may vary from one expression to another.

3.1 CLT for the product of sample covariance matrix and sample mean vector

In this section we present the central limit theorem for the product of the sample covariance matrix and the sample mean vector.

Theorem 2. *Assume* $\mathbf{X} \sim \mathcal{SN}_{p,n;q}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}})$, $p < n-1$, *with* $\mathbf{\Sigma}$ *positive definite and let* $p/n = c + o(n^{-1/2})$, $c \in [0,1)$ *as* $n \to \infty$. *Let* $\mathbf{l}$ *be a* p*-dimensional vector of constants that satisfies condition (A2). Then, under (A1) and (A2) it holds that*

$$\sqrt{n}\sigma_{\boldsymbol{\nu}}^{-1}\left(\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} - \mathbf{l}^T\mathbf{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{for } p/n \to c \in [0,1) \quad \text{as} \quad n \to \infty\,, \tag{6}$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\nu}} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}, \tag{7}$$

$$\sigma^2_{\boldsymbol{\nu}} = \left[\boldsymbol{\mu}_{\boldsymbol{\nu}}^T\mathbf{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}} + c||\mathbf{\Sigma}||_F^2\right]\mathbf{l}^T\mathbf{\Sigma}\mathbf{l} + (\mathbf{l}^T\mathbf{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}})^2\,. \tag{8}$$

Proof. Since $\mathbf{S}$ and $\overline{\mathbf{x}}$ are independently distributed, the conditional distribution $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}|(\overline{\mathbf{x}} = \overline{\mathbf{x}}^*)$ equals to the distribution of $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}^*$. Let $\mathbf{L}^* = (\mathbf{l}, \overline{\mathbf{x}}^*)^T$ and define $\widetilde{\mathbf{S}} = \mathbf{L}^*\mathbf{S}\mathbf{L}^{*T} = \{\widetilde{\mathbf{S}}_{ij}\}_{i,j=1,2}$ with $\widetilde{S}_{11} = \mathbf{l}^T\mathbf{S}\mathbf{l}$, $\widetilde{S}_{12} = \mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}^*$, $\widetilde{S}_{21} = \overline{\mathbf{x}}^{*T}\mathbf{S}\mathbf{l}$, and $\widetilde{S}_{22} = \overline{\mathbf{x}}^{*T}\mathbf{S}\overline{\mathbf{x}}^*$. Similarly, let $\widetilde{\mathbf{\Sigma}} = \mathbf{L}^*\mathbf{\Sigma}\mathbf{L}^{*T} = \{\widetilde{\mathbf{\Sigma}}_{ij}\}_{i,j=1,2}$ with $\widetilde{\Sigma}_{11} = \mathbf{l}^T\mathbf{\Sigma}\mathbf{l}$, $\widetilde{\Sigma}_{12} = \mathbf{l}^T\mathbf{\Sigma}\overline{\mathbf{x}}^*$, $\widetilde{\Sigma}_{21} = \overline{\mathbf{x}}^{*T}\mathbf{\Sigma}\mathbf{l}$, and $\widetilde{\Sigma}_{22} = \overline{\mathbf{x}}^{*T}\mathbf{\Sigma}\overline{\mathbf{x}}^*$.

Using $\mathbf{S} \sim \mathcal{W}_p\left(n-1, \frac{1}{n-1}\mathbf{\Sigma}\right)$ and *rank* $\mathbf{L}^* = 2 \le p$ we get from Theorem 3.2.5 of Muirhead (1982) that $\widetilde{\mathbf{S}} \sim \mathcal{W}_2\left(n-1, \frac{1}{n-1}\mathbf{\Sigma}\right)$. As a result, applying Theorem 3.2.10 of Muirhead (1982) we obtain

$$\widetilde{S}_{12}|\widetilde{S}_{22}, \overline{\mathbf{x}} = \overline{\mathbf{x}}^* \sim \mathcal{N}_k\left(\widetilde{\Sigma}_{12}\widetilde{\Sigma}_{22}^{-1}\widetilde{S}_{22}, \frac{1}{n-1}\widetilde{\Sigma}_{11\cdot 2}\widetilde{S}_{22}\right),$$

where $\widetilde{\Sigma}_{11\cdot 2} = \widetilde{\Sigma}_{11} - \widetilde{\Sigma}_{12}^2/\widetilde{\Sigma}_{22}$ is the Schur complement.

Let $\xi = (n-1)\widetilde{S}_{22}/\widetilde{\Sigma}_{22}$, then

$$\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}|\xi, \overline{\mathbf{x}} \sim \mathcal{N}\left(\frac{\xi}{n-1}\mathbf{l}^T\mathbf{\Sigma}\overline{\mathbf{x}}, \frac{\xi}{(n-1)^2}[\overline{\mathbf{x}}^T\mathbf{\Sigma}\overline{\mathbf{x}}\mathbf{l}^T\mathbf{\Sigma}\mathbf{l} - (\overline{\mathbf{x}}^T\mathbf{\Sigma}\mathbf{l})^2]\right).$$

From Theorem 3.2.8 of Muirhead (1982) follows that ξ and $\overline{\mathbf{x}}$ are independently distributed and $\xi \sim \chi^2_{n-1}$. Hence, the stochastic representation of $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ is given by

$$\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} \stackrel{d}{=} \frac{\xi}{n-1}\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} + \sqrt{\frac{\xi}{n-1}}(\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l} - (\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}})^2)^{1/2}\frac{z_0}{\sqrt{n-1}}, \tag{9}$$

where $\xi \sim \chi^2_n$, $z_0 \sim \mathcal{N}(0,1)$, $\overline{\mathbf{x}} \sim \mathcal{SN}_{p;q}\left(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}}\right)$; ξ, z_0 and $\overline{\mathbf{x}}$ are mutually independent.

From the properties of χ^2-distribution we immediately receive

$$\sqrt{n}\left(\frac{\xi}{n} - 1\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,2) \text{ as } n \to \infty\,. \tag{10}$$

We further get that $\sqrt{n}(z_0/\sqrt{n}) \sim \mathcal{N}(0,1)$ for all n and, consequently, it is also its asymptotic distribution.

Next, we show that $\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}$ and $\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}$ are jointly asymptotically normally distributed given $\boldsymbol{\nu} = \boldsymbol{\nu}^*$. For any a_1 and a_2, we consider

$$a_1\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} + 2a_2\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} = a_1\left(\overline{\mathbf{x}} + \frac{a_2}{a_1}\mathbf{l}\right)^T\boldsymbol{\Sigma}\left(\overline{\mathbf{x}} + \frac{a_2}{a_1}\mathbf{l}\right) - \frac{a_2^2}{a_1}\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l} = a_1\tilde{\mathbf{x}}^T\boldsymbol{\Sigma}\tilde{\mathbf{x}} - \frac{a_2^2}{a_1}\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l}\,,$$

where $\tilde{\mathbf{x}}|\boldsymbol{\nu} = \boldsymbol{\nu}^* \sim \mathcal{N}_p\left(\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}, \frac{1}{n}\boldsymbol{\Sigma}\right)$ with $\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^* + \dfrac{a_2}{a_1}\mathbf{l}$. By Provost and Rudiuk (1996) the random variable $\tilde{\mathbf{x}}^T\boldsymbol{\Sigma}\tilde{\mathbf{x}}$ can be expressed as

$$\tilde{\mathbf{x}}^T\boldsymbol{\Sigma}\tilde{\mathbf{x}} \stackrel{d}{=} \frac{1}{n}\sum_{i=1}^{p}\lambda_i^2\xi_i\,,$$

where

$$\xi_i \stackrel{i.i.d.}{\sim} \chi^2_1(\delta_i^2) \text{ with } \delta_i = \sqrt{n}\lambda_i^{-1/2}\mathbf{u}_i^T\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}\,.$$

The symbol $\chi^2_d(\delta_i^2)$ denotes the chi-squared distribution with d degrees of freedom and non-centrality parameter δ_i^2.

Now, we use the Lindeberg CLT to the i.i.d. random variables $V_i = \lambda_i^2\xi_i/n$. For that reason, we need first to verify the Lindeberg's condition. Denoting $\sigma_n^2 = \mathbb{V}(\sum_{i=1}^p V_i)$ we get

$$\sigma_n^2 = \sum_{i=1}^{p}\mathbb{V}\left[\frac{\lambda_i^2}{n}\xi_i\right] = \sum_{i=1}^{p}\frac{\lambda_i^4}{n^2}2(1+2\delta_i^2) = \frac{1}{n^2}\left(2\mathrm{tr}(\boldsymbol{\Sigma}^4) + 4n\boldsymbol{\mu}'_{\mathbf{a},\boldsymbol{\nu}^*}\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}\right) \tag{11}$$

We need to check if for any small $\varepsilon > 0$ it holds that

$$\lim_{n\to\infty}\frac{1}{\sigma_n^2}\sum_{i=1}^{p}\mathbb{E}\left[(V_i - \mathbb{E}(V_i))^2\,\mathbb{1}_{\{|V_i-\mathbb{E}(V_i)|>\varepsilon\sigma_n\}}\right] \longrightarrow 0\,.$$

First, we get

$$\begin{aligned}
\sum_{i=1}^{p} \mathbb{E}\left[(V_i - \mathbb{E}(V_i))^2 \, \mathbb{1}_{\{|V_i-\mathbb{E}(V_i)|>\varepsilon\sigma_n\}}\right] &\overset{Cauchy-Schwarz}{\leq} \sum_{i=1}^{p} \mathbb{E}^{1/2}\left[(V_i - \mathbb{E}(V_i))^4\right] \mathbb{P}^{1/2}\{|V_i - \mathbb{E}(V_i)| > \varepsilon\sigma_n\} \\
&\overset{Chebyshev}{\leq} \sum_{i=1}^{p} \frac{\lambda_i^4}{n^2} \sqrt{12(1+2\delta_i^2)^2 + 48(1+4\delta_i^2)} \frac{\sigma_i}{\varepsilon\sigma_n}
\end{aligned}$$

with $\sigma_i = \mathbb{V}(V_i)$ and, thus,

$$\begin{aligned}
\frac{1}{\sigma_n^2}\sum_{i=1}^{p} \mathbb{E}\left[(V_i - \mathbb{E}(V_i))^2 \, \mathbb{1}_{\{|V_i-\mathbb{E}(V_i)|>\varepsilon\sigma_n\}}\right] &\leq \frac{1}{\varepsilon} \frac{\sum_{i=1}^{p} \lambda_i^4 \sqrt{12(1+2\delta_i^2)^2 + 48(1+4\delta_i^2)} \frac{\sigma_i}{\sigma_n}}{2\mathrm{tr}(\mathbf{\Sigma}^4) + 4n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}}} \\
&= \frac{\sqrt{3}}{\varepsilon} \frac{\sum_{i=1}^{p} \lambda_i^4 \sqrt{(5+2\delta_i^2)^2 - 20} \frac{\sigma_i}{\sigma_n}}{\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}} \\
&\leq \frac{\sqrt{3}}{\varepsilon} \frac{\sum_{i=1}^{p} \lambda_i^4 (5+2\delta_i^2) \frac{\sigma_i}{\sigma_n}}{\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}} \\
&= \frac{\sqrt{3}}{\varepsilon} \frac{5\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}}{\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}} \frac{\sigma_{max}}{\sigma_n} \\
&\leq \frac{\sqrt{3}}{\varepsilon} \left(\frac{4}{1 + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}/\mathrm{tr}(\mathbf{\Sigma}^4)} + 1 \right) \frac{\sigma_{max}}{\sigma_n} \\
&\leq \frac{5\sqrt{3}}{\varepsilon} \frac{\sigma_{max}}{\sigma_n} \, .
\end{aligned}$$

Finally, Assumptions (A1) and (A2) yield

$$\frac{\sigma_{max}^2}{\sigma_n^2} = \frac{\sup_i \sigma_i^2}{\sigma_n^2} = \frac{\sup_i \lambda_i^4 (1+2\delta_i^2)}{\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}} = \frac{\sup_i \lambda_i^4 + 2n\lambda_i^3 (\mathbf{u}_i^T \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*})^2}{\mathrm{tr}(\mathbf{\Sigma}^4) + 2n\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma}^3 \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}} \longrightarrow 0 \, , \tag{12}$$

which verifies the Lindeberg condition since

$$\begin{aligned}
(\mathbf{u}_i^T \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*})^2 &= \left(\mathbf{u}_i^T \boldsymbol{\mu} + \mathbf{u}_i^T \mathbf{B}\boldsymbol{\nu}^* + \mathbf{u}_i^T \mathbf{l} \frac{a_2}{a_1} \right)^2 \\
&= (\mathbf{u}_i' \boldsymbol{\mu})^2 + \left(\mathbf{u}_i^T \mathbf{l} \frac{a_2}{a_1} \right)^2 + (\mathbf{u}_i^T \mathbf{B}\boldsymbol{\nu}^*)^2 + 2\mathbf{u}_i^T \boldsymbol{\mu} \cdot \mathbf{u}_i^T \mathbf{B}\boldsymbol{\nu}^* + 2\frac{a_2}{a_1} \mathbf{u}_i^T \mathbf{l} (\mathbf{u}_i^T \boldsymbol{\mu} + \mathbf{u}_i^T \mathbf{B}\boldsymbol{\nu}^*) \\
&\overset{(A2)}{\leq} M_2^2 + qM_2^2 \boldsymbol{\nu}^{*\,\prime} \boldsymbol{\nu}^* + M_2^2 \frac{a_2^2}{a_1^2} + 2M_2^2 \sqrt{q\boldsymbol{\nu}^{*\,\prime}\boldsymbol{\nu}^*} + 2M_2^2 \frac{a_2}{a_1} (1 + \sqrt{q\boldsymbol{\nu}^{*\,T}\boldsymbol{\nu}^*}) \\
&= M_2^2 \left(1 + \sqrt{q\boldsymbol{\nu}^{*\,\prime}\boldsymbol{\nu}^*} + \frac{a_2}{a_1} \right)^2 < \infty.
\end{aligned} \tag{13}$$

Thus, using (11) and

$$\sum_{i=1}^{p} \mathbb{E}(V_i) = \sum_{i=1}^{p} \frac{\lambda_i^2}{n} (1 + \delta_i^2) = \mathrm{tr}(\mathbf{\Sigma}^2)/n + \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T \mathbf{\Sigma} \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*} \tag{14}$$

we get the following CLT

$$\sqrt{n}\frac{\tilde{\mathbf{x}}^T\boldsymbol{\Sigma}\tilde{\mathbf{x}} - \mathrm{tr}(\boldsymbol{\Sigma}^2)/n - \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}}{\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}^4)/n + 2\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}'\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}}} \xrightarrow{d} \mathcal{N}(0,2)$$

and for $a_1\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} + 2a_2\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}$ we have

$$\sqrt{n}\frac{a_1\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} + 2a_2\mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} - a_1\left(\mathrm{tr}(\boldsymbol{\Sigma}^2)/n + \boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}\right) + \dfrac{a_2^2}{a_1}\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l}}{\sqrt{a_1^2\left(\mathrm{tr}(\boldsymbol{\Sigma}^4)/n + 2\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}'\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\mathbf{a},\boldsymbol{\nu}^*}\right)}} \xrightarrow{d} \mathcal{N}(0,2)\,.$$

Denoting $\mathbf{a} = (a_1, 2a_2)^T$ and $\boldsymbol{\mu}_{\boldsymbol{\nu}^*} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*$ we can rewrite it as

$$\sqrt{n}\left[\mathbf{a}^T\begin{pmatrix}\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}\\ \mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}\end{pmatrix} - \mathbf{a}^T\begin{pmatrix}\boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*} + c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{p}\\ \mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*}\end{pmatrix}\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{a}^T\begin{pmatrix}2c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{p} + 4\boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*} & 2\mathbf{l}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*}\\ 2\mathbf{l}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*} & \mathbf{l}^T\boldsymbol{\Sigma}^3\mathbf{l}\end{pmatrix}\mathbf{a}\right) \quad (15)$$

which implies that the vector $\sqrt{n}\left(\overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} - \boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*} - c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{p},\ \mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}} - \mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*}\right)^T$ has asymptotically multivariate normal distribution because the vector $\mathbf{a}$ is arbitrary.

Taking into account (15), (10) and the fact that ξ, z_0 and $\overline{\mathbf{x}}$ are mutually independent we get the following CLT

$$\sqrt{n}\left[\begin{pmatrix}\frac{\xi}{n}\\ \overline{\mathbf{x}}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}\\ \mathbf{l}^T\boldsymbol{\Sigma}\overline{\mathbf{x}}\\ \frac{z_0}{\sqrt{n}}\end{pmatrix} - \begin{pmatrix}1\\ \boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*} + c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{p}\\ \mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*}\\ 0\end{pmatrix}\right] \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}2 & 0 & 0 & 0\\ 0 & 2c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^4)}{p} + 4\boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*} & 2\mathbf{l}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*} & 0\\ 0 & 2\mathbf{l}^T\boldsymbol{\Sigma}^3\boldsymbol{\mu}_{\boldsymbol{\nu}^*} & \mathbf{l}^T\boldsymbol{\Sigma}^3\mathbf{l} & 0\\ 0 & 0 & 0 & 1\end{pmatrix}\right).$$

The application of the multivariate delta method leads to

$$\sqrt{n}\sigma_{\boldsymbol{\nu}^*}^{-1}\left(\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} - \mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*}\right) \xrightarrow{d} \mathcal{N}(0,1) \quad (16)$$

where

$$\sigma_{\boldsymbol{\nu}^*}^2 = (\mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*})^2 + \mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l}\left[\boldsymbol{\mu}_{\boldsymbol{\nu}^*}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}^*} + c\frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{p}\right]$$

The asymptotic distribution does not depend on $\boldsymbol{\nu}^*$ and, thus, it is also the unconditional asymptotic distribution. □

Theorem 2 shows that properly normalized bilinear form $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ itself can be accurately approximated by a mixture of normal distributions with both mean and variance depending on $\boldsymbol{\nu}$. Moreover, this central limit theorem delivers the following approximation for the distribution of

$\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$, namely for large n and p we have

$$p^{-1}\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} \approx \mathcal{CN}\left(p^{-1}\mathbf{l}^T\mathbf{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}}, \frac{p^{-2}\sigma^2_{\boldsymbol{\nu}}}{n}\right), \tag{17}$$

i.e., it has compound normal distribution with random mean and variance. The usual asymptotic normality can be recovered as a special case of our result, i.e., taking $\boldsymbol{\nu}$ as a deterministic vector (e.g., zero).

The proof of Theorem 2 shows, in particular, that its key point is a stochastic representation of the product $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ which can be presented using a χ^2 distributed random variable, a standard normally distributed random variable, and the random vector which follows the generalized skew normal distribution. Assumption (A1) ensures that the spectrum of matrix $\mathbf{\Sigma}$ is uniformly bounded away from zero, i.e., maximum eigenvalue is bounded from above and minimum eigenvalue from below by positive constants. The second assumption (A2) is a technical one, it stands, in particular, for boundedness of normalized Frobenius norm of the population covariance matrix. Both (A1) and (A2) guarantee that the asymptotic mean $p^{-1}\mathbf{l}^T\mathbf{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}}$ and variance $p^{-2}\sigma^2_{\boldsymbol{\nu}}$ stay bounded and the covariance matrix $\mathbf{\Sigma}$ is invertible as the dimension p increases. Note that the case of standard asymptotics can be easily recovered from our result if we set $c \to 0$.

3.2 CLT for the product of inverse sample covariance matrix and sample mean vector

In this section we consider the distributional properties of the product of the inverse sample covariance matrix $\mathbf{S}^{-1}$ and the sample mean vector $\overline{\mathbf{x}}$. Again we prove that proper weighted bilinear forms involving $\mathbf{S}^{-1}$ and $\overline{\mathbf{x}}$ have asymptotically a normal distribution. This result is summarized in Theorem 3.

Theorem 3. *Assume $\mathbf{X} \sim \mathcal{SN}_{p,n;q}(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{B}; f_{\boldsymbol{\nu}})$, $p < n-1$, with $\mathbf{\Sigma}$ positive definite and let $p/n = c + o(n^{-1/2})$, $c \in [0,1)$ as $n \to \infty$. Let $\mathbf{l}$ be a p-dimensional vector of constants such that $p^{-1}\mathbf{l}^T\mathbf{\Sigma}^{-1}\mathbf{l} \leq M_2 < \infty$. Then, under (A1) and (A2) it holds that*

$$\sqrt{n}\tilde{\sigma}^{-1}_{\boldsymbol{\nu}}\left(\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}} - \frac{1}{1-c}\mathbf{l}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \tag{18}$$

where $\boldsymbol{\mu}_{\boldsymbol{\nu}} = \boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}$ and

$$\tilde{\sigma}^2_{\boldsymbol{\nu}} = \frac{1}{(1-c)^3}\left(\left(\mathbf{l}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right)^2 + \mathbf{l}^T\mathbf{\Sigma}^{-1}\mathbf{l}(1+\boldsymbol{\mu}^T_{\boldsymbol{\nu}}\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}})\right).$$

Proof. From the properties of the Wishart distribution (see Muirhead (1982)) and Theorem 1 it holds that

$$\mathbf{S}^{-1} \sim \mathcal{IW}_p\left(n+p, (n-1)\boldsymbol{\Sigma}^{-1}\right).$$

Since $\mathbf{S}^{-1}$ and $\overline{\mathbf{x}}$ are independently distributed we get that the conditional distribution of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}|\overline{\mathbf{x}} = \overline{\mathbf{x}}^*$ equals to the distribution of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*$ with

$$\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^* = (n-1)\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*\frac{\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*}{\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*}\frac{\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*}{(n-1)\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*}.$$

Using Theorem 3.2.12 of Muirhead (1982) we get that

$$(n-1)\frac{\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*}{\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*} \sim \chi^2_{n-p}$$

and it is independent of $\overline{\mathbf{x}}^*$. Hence,

$$\widetilde{\xi} = (n-1)\frac{\overline{\mathbf{x}}^{T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}}{\overline{\mathbf{x}}^{T}\mathbf{S}^{-1}\overline{\mathbf{x}}} \sim \chi^2_{n-p} \tag{19}$$

and it is independent of $\overline{\mathbf{x}}$.

Applying Theorem 3 of Bodnar and Okhrin (2008) it follows that $\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*$ is independent of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*/\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*$ for given $\overline{\mathbf{x}}^*$. Moreover, as a result, it is also independent of $\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^* \cdot \mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*/\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*$ and, respectively, of $\overline{\mathbf{x}}^{T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}} \cdot \mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}/\overline{\mathbf{x}}^{T}\mathbf{S}^{-1}\overline{\mathbf{x}}$.

From the proof of Theorem 1 of Bodnar and Schmid (2008) we obtain

$$(n-1)\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*\frac{\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*}{\overline{\mathbf{x}}^{*T}\mathbf{S}^{-1}\overline{\mathbf{x}}^*} \sim t\left(n-p+1; (n-1)\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*; (n-1)^2\frac{\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*}{n-p+1}\mathbf{l}^T\mathbf{R}_{\overline{\mathbf{x}}^*}\mathbf{l}\right), \tag{20}$$

where $\mathbf{R}_{\mathbf{a}} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{a}\mathbf{a}^T\boldsymbol{\Sigma}^{-1}/\mathbf{a}^T\boldsymbol{\Sigma}^{-1}\mathbf{a}$, $\mathbf{a} \in I\!R^p$, and the symbol $t(k, \mu, \sigma^2)$ denotes t distribution with k degrees of freedom, mean μ and variance σ^2.

Combining (19) and (20), we get that the stochastic representation of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^*$ is given by

$$\begin{aligned}
\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^* &\stackrel{d}{=} \widetilde{\xi}^{-1}(n-1)\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^* + t_0\sqrt{\frac{\overline{\mathbf{x}}^{*T}\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^*}{n-p+1}\cdot \mathbf{l}^T\mathbf{R}_{\overline{\mathbf{x}}^*}\mathbf{l}}\right) \\
&= \widetilde{\xi}^{-1}(n-1)\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\overline{\mathbf{x}}^* + \frac{t_0}{\sqrt{n-p+1}}\sqrt{\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}}\sqrt{\overline{\mathbf{x}}^{*T}\mathbf{R}_{\mathbf{l}}\overline{\mathbf{x}}^*}\right),
\end{aligned}$$

where $\overline{\mathbf{x}}^* \sim \mathcal{N}_p\left(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*, \frac{1}{n}\boldsymbol{\Sigma}\right)$, $\widetilde{\xi} \sim \chi^2_{n-p}$, and $t_0 \sim t(n-p+1, 0, 1)$; $\widetilde{\xi}$, $\overline{\mathbf{x}}^*$ and t_0 are mutually independent.

Since $\mathbf{R_l \Sigma R_l} = \mathbf{R_l}$, $tr(\mathbf{R_l \Sigma}) = p-1$, and $\mathbf{R_l \Sigma \Sigma^{-1} l} = \mathbf{0}$, the application of Corollary 5.1.3a and Theorem 5.5.1 in Mathai and Provost (1992) leads to

$$\mathbf{l}^T \mathbf{\Sigma}^{-1} \overline{\mathbf{x}}^* \sim \mathcal{N}\left(\mathbf{l}^T \mathbf{\Sigma}^{-1}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*), \frac{1}{n}\mathbf{l}^T \mathbf{\Sigma}^{-1}\mathbf{l}\right)$$

and

$$\overline{\mathbf{x}}^{*T}\mathbf{R_l}\overline{\mathbf{x}}^* \sim \chi^2_{p-1}(n\delta^2(\boldsymbol{\nu}^*)) \quad \text{with} \quad \delta^2(\boldsymbol{\nu}^*) = (\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*)^T \mathbf{R_l}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*)$$

as well as $\mathbf{l}^T \mathbf{\Sigma}^{-1} \overline{\mathbf{x}}^*$ and $\overline{\mathbf{x}}^{*T}\mathbf{R_l}\overline{\mathbf{x}}^*$ are independent. Finally, using the stochastic representation of a t-distributed random variable, we get

$$\mathbf{l}^T \mathbf{S}^{-1} \overline{\mathbf{x}}^* \overset{d}{=} \tilde{\xi}^{-1}(n-1)\left(\mathbf{l}^T \mathbf{\Sigma}^{-1}(\boldsymbol{\mu} + \mathbf{B}\boldsymbol{\nu}^*) + \sqrt{\mathbf{l}^T \mathbf{\Sigma}^{-1}\mathbf{l}}\sqrt{1 + \frac{p-1}{n-p+1}\eta^*}\frac{z_0}{\sqrt{n}}\right), \qquad (21)$$

where $\tilde{\xi} \sim \chi^2_{n-p}$, $z_0 \sim \mathcal{N}(0,1)$, and $\eta^* \sim F_{p-1,n-p+1}(n\delta^2(\boldsymbol{\nu}^*))$ (non-central F-distribution with $p-1$ and $n-p+1$ degrees of freedom and non-centrality parameter $n\delta^2(\boldsymbol{\nu}^*)$); $\tilde{\xi}$, z_0 and η^* are mutually independent.

From Lemma 6.4.(b) in Bodnar et al. (2016b) we get

$$\sqrt{n}\left(\begin{pmatrix} \tilde{\xi}/(n-p) \\ \eta^* \\ z_0/\sqrt{n} \end{pmatrix} - \begin{pmatrix} 1 \\ 1 + \delta^2(\boldsymbol{\nu}^*)/c \\ 0 \end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2/(1-c) & 0 & 0 \\ 0 & \sigma^2_\eta & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

for $p/n = c + o(n^{-1/2})$, $c \in (0,1)$ as $n \to \infty$ with

$$\sigma^2_\eta = \frac{2}{c}\left(1 + 2\frac{\delta^2(\boldsymbol{\nu}^*)}{c}\right) + \frac{2}{1-c}\left(1 + \frac{\delta^2(\boldsymbol{\nu}^*)}{c}\right)^2$$

Consequently,

$$\sqrt{n}\left(\begin{pmatrix} \tilde{\xi}/(n-1) \\ (p-1)\eta^*/(n-p+1) \\ z_0/\sqrt{n} \end{pmatrix} - \begin{pmatrix} (1-c) \\ (c + \delta^2(\boldsymbol{\nu}^*))/(1-c) \\ 0 \end{pmatrix}\right)$$
$$\xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 2(1-c) & 0 & 0 \\ 0 & c^2\sigma^2_\eta/(1-c)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

for $p/n = c + o(n^{-1/2})$, $c \in (0,1)$ as $n \to \infty$.

Finally, the application of the delta-method (c.f. DasGupta (2008, Theorem 3.7)) leads to

$$\sqrt{n}\left(\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^* - \frac{1}{1-c}\mathbf{l}^T\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}^*)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\sigma}^2_{\boldsymbol{\nu}^*})$$

for $p/n = c + o(n^{-1/2})$, $c \in (0,1)$ as $n \to \infty$ with

$$\tilde{\sigma}^2_{\boldsymbol{\nu}^*} = \frac{1}{(1-c)^3}\left(2\left(\mathbf{l}^T\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}^*)\right)^2 + \mathbf{l}^T\mathbf{\Sigma}^{-1}\mathbf{l}(1+\delta^2(\boldsymbol{\nu}^*))\right).$$

Consequently,

$$\sqrt{n}\tilde{\sigma}^{-1}_{\boldsymbol{\nu}^*}\left(\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}^* - \frac{1}{1-c}\mathbf{l}^T\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}^*)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,,$$

where the asymptotic distribution does not depend on $\boldsymbol{\nu}^*$. Hence, it is also the unconditional asymptotic distribution.

$\square$

Again, Theorem 3 shows that the distribution of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}$ can be approximated by a mixture of normal distributions. Indeed,

$$p^{-1}\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}} \approx \mathcal{CN}\left(\frac{p^{-1}}{1-c}\mathbf{l}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}}, \frac{p^{-2}\tilde{\sigma}^2_{\boldsymbol{\nu}}}{n}\right). \tag{22}$$

In the proof of Theorem 3 we can read out that the stochastic representation for the product of the inverse sample covariance matrix and the sample mean vector is presented by using a χ^2 distributed random variable, a general skew normally distributed random vector and a standard t-distributed random variable. This result is itself very useful and allows to generate the values of $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}$ by just generating three random variables from the standard univariate distributions and a random vector $\boldsymbol{\nu}$ which determines the family of the skew normal matrix variate distribution. The assumptions about the boundedness of the quadratic and bilinear forms involving $\mathbf{\Sigma}^{-1}$ plays here the same role as in Theorem 2. Note that in this case we need no assumption either on the Frobenius norm of the covariance matrix or its inverse.

4 Numerical study

In this section we provide a Monte Carlo simulation study to investigate the performance of the suggested CLTs for the products of the sample (inverse) covariance matrix and the sample mean vector.

In our simulations we put $\mathbf{l} = \mathbf{1}_p$, each element of the vector $\boldsymbol{\mu}$ is uniformly distributed on $[-1, 1]$ while each element of the matrix $\mathbf{B}$ is uniformly distributed on $[0, 1]$. Also, we take $\boldsymbol{\Sigma}$ as a diagonal matrix where each diagonal element is uniformly distributed on $[0, 1]$. It can be checked that in such a setting the assumptions (A1) and (A2) are satisfied. Indeed, the population covariance matrix satisfies the condition (A1) because the probability of getting exactly zero eigenvalue equals to zero. On the other hand, the condition (A2) is obviously valid too because the ith eigenvector of $\boldsymbol{\Sigma}$ is $\mathbf{u}_i = \mathbf{e}_i = (0, \ldots, \underset{\text{ith place}}{1}, 0, \ldots, 0)'$.

In order to define the distribution for the random vector $\boldsymbol{\nu}$, we consider two special cases. In the first case we take $\boldsymbol{\nu} = |\boldsymbol{\psi}|$, where $\boldsymbol{\psi} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}_q)$, i.e. $\boldsymbol{\nu}$ has a q-variate truncated normal distribution. In the second case we put $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{I}_q, \mathbf{1}_q, 10)$, i.e. $\boldsymbol{\nu}$ has a q-variate generalized asymmetric Laplace distribution (c.f., Kozubowski et al. (2013)). Also, we put $q = 10$.

We compare the results for several values of $c \in \{0.1, 0.5, 0.8, 0.95\}$. The simulated data consists of $N = 10^4$ independent realizations which are used to fit the corresponding kernel density estimators with Gaussian density. The bandwith parameters are determined via cross-validation for every sample. The asymptotic distributions are simulated using the results of Theorems 2 and 3. The corresponding algorithm is given next:

a) generate $\boldsymbol{\nu} = |\boldsymbol{\psi}|$, where $\boldsymbol{\psi} \sim \mathcal{N}_q(\mathbf{0}_q, \mathbf{I}_q)$, or generate $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{I}_q, \mathbf{1}_q, 10)$;

b) generate $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ by using the stochastic representation (9) obtained in the proof of Theorem 2, namely

$$\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} \overset{d}{=} \frac{\xi}{n-1}\mathbf{l}^T\boldsymbol{\Sigma}(\mathbf{y}+\mathbf{B}\boldsymbol{\nu}) + \frac{\sqrt{\xi}}{n-1}((\mathbf{y}+\mathbf{B}\boldsymbol{\nu})^T\boldsymbol{\Sigma}(\mathbf{y}+\mathbf{B}\boldsymbol{\nu})\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l} - (\mathbf{l}^T\boldsymbol{\Sigma}(\mathbf{y}+\mathbf{B}\boldsymbol{\nu}))^2)^{1/2}z_0,$$

where $\xi \sim \chi^2_n$, $z_0 \sim \mathcal{N}(0, 1)$, $\mathbf{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$; ξ, z_0, $\mathbf{y}$, and $\boldsymbol{\nu}$ are mutually independent

b') generate $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}$ by using the stochastic representation (21) obtained in the proof of Theorem 3, namely

$$\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}} \overset{d}{=} \widetilde{\xi}^{-1}(n-1)\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}) + \sqrt{\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}}\sqrt{1+\frac{p-1}{n-p+1}\eta}\frac{z_0}{\sqrt{n}}\right),$$

where $\widetilde{\xi} \sim \chi^2_{n-p}$, $z_0 \sim \mathcal{N}(0, 1)$, and $\eta \sim F_{p-1,n-p+1}(n\delta^2(\boldsymbol{\nu}))$ with $\delta^2(\boldsymbol{\nu}) = (\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu})^T\mathbf{R}_\mathbf{l}(\boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu})$, $\mathbf{R}_\mathbf{l} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}/\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$; $\widetilde{\xi}$, z_0 and $(\eta, \boldsymbol{\nu})$ are mutually independent.

c) compute

$$\sqrt{n}\sigma_{\boldsymbol{\nu}}^{-1}\left(\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}} - \mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right)$$

and

$$\sqrt{n}\tilde{\sigma}_{\boldsymbol{\nu}}^{-1}\left(\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}-\frac{1}{1-c}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right)$$

where

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\nu}} &= \boldsymbol{\mu}+\mathbf{B}\boldsymbol{\nu}\\
\sigma_{\boldsymbol{\nu}}^2 &= \left[\boldsymbol{\mu}_{\boldsymbol{\nu}}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}}+c||\boldsymbol{\Sigma}||_F^2\right]\mathbf{l}^T\boldsymbol{\Sigma}\mathbf{l}+(\mathbf{l}^T\boldsymbol{\Sigma}\boldsymbol{\mu}_{\boldsymbol{\nu}})^2\\
\tilde{\sigma}_{\boldsymbol{\nu}}^2 &= \frac{1}{(1-c)^3}\left(2\left(\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_{\boldsymbol{\nu}}\right)^2+\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}(1+\delta^2(\boldsymbol{\nu}))\right)
\end{aligned}$$

with $\delta^2(\boldsymbol{\nu})=\boldsymbol{\mu}_{\boldsymbol{\nu}}^T\mathbf{R}_{\mathbf{l}}\boldsymbol{\mu}_{\boldsymbol{\nu}}$, $\mathbf{R}_{\mathbf{l}}=\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\mathbf{l}\mathbf{l}^T\boldsymbol{\Sigma}^{-1}/\mathbf{l}^T\boldsymbol{\Sigma}^{-1}\mathbf{l}$.

d) repeat a)-c) N times.

It is remarkable that for generating $\mathbf{l}^T\mathbf{S}\overline{\mathbf{x}}$ and $\mathbf{l}^T\mathbf{S}^{-1}\overline{\mathbf{x}}$ only random variables from the standard distributions are need. Neither the data matrix $\mathbf{X}$ nor the sample covariance matrix $\mathbf{S}$ are used.

$$[\ Figures\ \ 1-8\ ]$$

In Figures 1-4 we present the results of simulations for the asymptotic distribution that is given in Theorem 2 while the asymptotic distribution as given in Theorem 3 is presented in Figures 5-8 for different values of $c=\{0.1,0.5,0.8,0.95\}$. The suggested asymptotic distributions are shown as a a dashed black line, while the standard normal distribution is a solid black line. All results demonstrate a good performance of both asymptotic distributions for all considered values of c. Even in the extreme case $c=0.95$ our asymptotic results seem to produce a quite reasonable approximation. Moreover, we observe a good robustness of our theoretical results for different distributions of $\boldsymbol{\nu}$. Also, we observe that all asymptotic distributions are slightly skewed to the right for the finite dimensions. This effect is even more significant in the case of the generalized asymmetric Laplace distribution. Nevertheless, the skewness disappears with growing dimension and sample size, i.e., the distribution becomes symmetric one and converges to its asymptotic counterpart.

5 Summary

In this paper we introduce the family of the matrix-variate generalized skew normal (MVGSN) distribution that generalizes a large number of the existing skew normal models. Under the

MVGSN distribution we derive the distributions of the sample mean vector and the sample covariance matrix. Moreover, we show that they are independently distributed. Furthermore, we derive the CLTs under high-dimensional asymptotic regime for the products of the sample (inverse) covariance matrix and the sample mean vector. In the numerical study, we document the good finite sample performance of both asymptotic distributions.

6 Appendix

Proof of Proposition 1.

Proof. Straightforward but tedious calculations give

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{Z}) &= C^{-1}\int_{\mathbb{R}_+^q} f_{N_{p,n}(\boldsymbol{\mu}\mathbf{1}_n^T,\boldsymbol{\Sigma}\otimes\mathbf{I}_n)}(\mathbf{Z}-\mathbf{B}\boldsymbol{\nu}^*|\boldsymbol{\nu}=\boldsymbol{\nu}^*) f_{N_q(\boldsymbol{\xi},\boldsymbol{\Omega})}(\boldsymbol{\nu}^*)\mathbf{d}\boldsymbol{\nu}^*\\
&= C^{-1}\frac{(2\pi)^{-(np+q)/2}}{|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Sigma}|^{n/2}}\int_{\mathbb{R}_+^q}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}^*-\boldsymbol{\xi})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\nu}^*-\boldsymbol{\xi})\right\}\\
&\times \exp\left\{-\frac{1}{2}\mathbf{vec}\left(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T-\mathbf{B}\boldsymbol{\nu}^*\mathbf{1}_n^T\right)^T(\mathbf{I}_n\otimes\boldsymbol{\Sigma})^{-1}\mathbf{vec}\left(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T-\mathbf{B}\boldsymbol{\nu}^*\mathbf{1}_n^T\right)\right\}\mathbf{d}\boldsymbol{\nu}^*\\
&= C^{-1}\frac{(2\pi)^{-(np+q)/2}}{|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Sigma}|^{n/2}}\exp\left\{-\frac{1}{2}\left[\mathbf{vec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)^T(\mathbf{I}_n\otimes\boldsymbol{\Sigma})^{-1}\mathbf{vec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\xi}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right]\right\}\\
&\times \exp\left\{\frac{1}{2}\left[\left(\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right)^T\mathbf{D}\left(\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right)\right]\right\}\\
&\times \int_{\mathbb{R}_+^q}\exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{\nu}^*-\mathbf{D}\left(\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right)\right)^T\right.\right.\\
&\times \left.\left.\mathbf{D}^{-1}\left(\boldsymbol{\nu}^*-\mathbf{D}\left(\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right)\right)\right]\right\}\mathbf{d}\boldsymbol{\nu}^*\\
&= C^{-1}\frac{|\mathbf{F}|^{1/2}|\mathbf{D}|^{1/2}}{|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Sigma}|^{n/2}}\exp\left\{-\frac{1}{2}\left[\boldsymbol{\xi}^T\boldsymbol{\Omega}^{-1}(-\mathbf{D}+\boldsymbol{\Omega}-\mathbf{DEFE}^T\mathbf{D})\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right]\right\}\\
&\times \Phi_q\left(\mathbf{0};-\mathbf{D}[\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}],\mathbf{D}\right)\\
&\times \phi_{pn}\left(\mathbf{vec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T);\mathbf{FE}^T\mathbf{D}\boldsymbol{\Omega}^{-1}\boldsymbol{\xi},\mathbf{F}\right)\\
&= \widetilde{C}^{-1}\Phi_q\left(\mathbf{0};-\mathbf{D}[\mathbf{Evec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T)+\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}],\mathbf{D}\right)\phi_{pn}\left(\mathbf{vec}(\mathbf{Z}-\boldsymbol{\mu}\mathbf{1}_n^T);\mathbf{FE}^T\mathbf{D}\boldsymbol{\Omega}^{-1}\boldsymbol{\xi},\mathbf{F}\right)
\end{aligned}$$

where $\mathbf{D}=(n\mathbf{B}^T\boldsymbol{\Sigma}^{-1}\mathbf{B}+\boldsymbol{\Omega}^{-1})^{-1}$, $\mathbf{E}=\mathbf{1}_n^T\otimes\mathbf{B}^T\boldsymbol{\Sigma}^{-1}$, $\mathbf{F}=(\mathbf{I}_n\otimes\boldsymbol{\Sigma}^{-1}-\mathbf{E}^T\mathbf{DE})^{-1}$, and

$$\widetilde{C}^{-1}=C^{-1}\frac{|\mathbf{F}|^{1/2}|\mathbf{D}|^{1/2}}{|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Sigma}|^{n/2}}\exp\left\{-\frac{1}{2}\left[\boldsymbol{\xi}^T\boldsymbol{\Omega}^{-1}(-\mathbf{D}+\boldsymbol{\Omega}-\mathbf{DEFE}^T\mathbf{D})\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}\right]\right\}.$$

□

References

Adcock, C., Eling, M., and Loperfido, N. (2015). Skewed distributions in finance and actuarial science: a review. The European Journal of Finance, 21:1253–1281.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. Scandinavian Journal of Statistics, 32:159–188.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. Journal of the Royal Statistical Society: Series B, 61:579–602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. Biometrika, 83:715–726.

Bai, J. and Shi, S. (2011). Estimating high dimensional covariance matrices and its applications. Annals of Economics and Finance, 12:199–215.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. Annals of Probability, 32:553–605.

Bartoletti, S. and Loperfido, N. (2010). Modelling air polution data by the skew-normal distribution. Stochastic Environmental Research and Risk Assessment, 24:513–517.

Bodnar, T. and Gupta, A. K. (2011). Estimation of the precision matrix of multivariate elliptically contoured stable distribution. Statistics, 45:131–142.

Bodnar, T., Gupta, A. K., and Parolya, N. (2014a). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. Journal of Multivariate Analysis, 132:215–228.

Bodnar, T., Gupta, A. K., and Parolya, N. (2016a). Direct shrinkage estimation of large dimensional precision matrix. Journal of Multivariate Analysis, page to appear.

Bodnar, T., Hautsch, N., and Parolya, N. (2016b). Consistent estimation of the high dimensional efficient frontier. Technical report.

Bodnar, T., Mazur, S., and Okhrin, Y. (2013). On the exact and approximate distributions of the product of a wishart matrix with a normal vector. Journal of Multivariate Analysis, 125:176–189.

Bodnar, T., Mazur, S., and Okhrin, Y. (2014b). Distribution of the product of singular wishart matrix and normal vector. Theory of Probability and Mathematical Statistics, 91:1–14.

Bodnar, T. and Okhrin, Y. (2008). Properties of the singular, inverse and generalized inverse partitioned wishart distributions. Journal of Multivariate Analysis, 99:2389–2405.

Bodnar, T. and Okhrin, Y. (2011). On the product of inverse wishart and normal distributions with applications to discriminant analysis and portfolio theory. Scandinavian Journal of Statistics, 38:311–331.

Bodnar, T. and Schmid, W. (2008). A test for the weights of the global minimum variance portfolio in an elliptical model. Metrika, 67(2):127–143.

Cai, T. T. and Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. Ann. Statist., 40:2014–2042.

Cai, T. T. and Zhou, H. H. (2012). Optimal rates of convergence for sparse covariance matrix estimation. Ann. Statist., 40:2389–2420.

Christiansen, M. and Loperfido, N. (2014). Improved approximation of the sum of random vectors by the skew normal distribution. Journal of Applied Probability, 51(2):466–482.

DasGupta, A. (2008). Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics. Springer.

De Luca, G. and Loperfido, N. (2015). Modelling multivariate skewness in financial returns: a SGARCH approach. The European Journal of Finance, 21:1113–1131.

Efron, B. (2006). Minimum volume confidence regions for a multivariate normal mean vector. Journal of the Royal Statistical Society: Series B, 68:655–670.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. Journal of Econometrics, 147:186–197.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B, 75:603–680.

Jorion, P. (1986). Bayes-stein estimation for portfolio analysis. Journal of Financial and Quantative Analysis, 21:279–292.

Kotsiuba, I. and Mazur, S. (2015). On the asymptotic and approximate distributions of the product of an inverse Wishart matrix and a Gaussian random vector. Theory of Probability and Mathematical Statistics, 93:96–105.

Kozubowski, T. J., Podgórski, K., and Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields. Journal of Multivariate Analysis, 113:59–72.

Liseo, B. and Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. Statistics & Probability Letters, 61:395–401.

Liseo, B. and Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. Journal of Statistical Planning and Inference, 136:373–389.

Loperfido, N. (2010). Canonical transformations of skew-normal variates. TEST, 19:146–165.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. Sbornik: Mathematics, 1:457–483.

Mathai, A. and Provost, S. B. (1992). Quadratic Forms in Random Variables. Marcel Dekker.

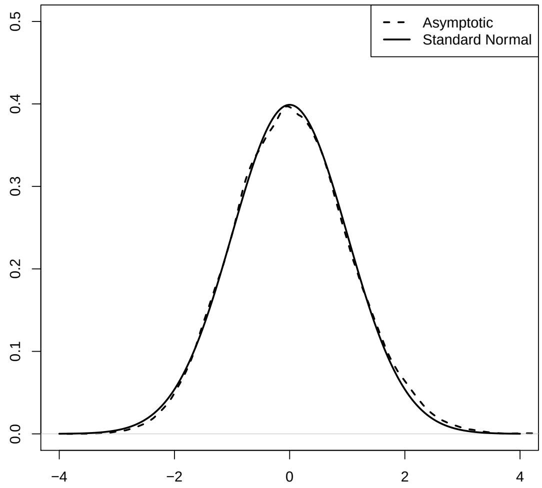
Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. Wiley, New York.

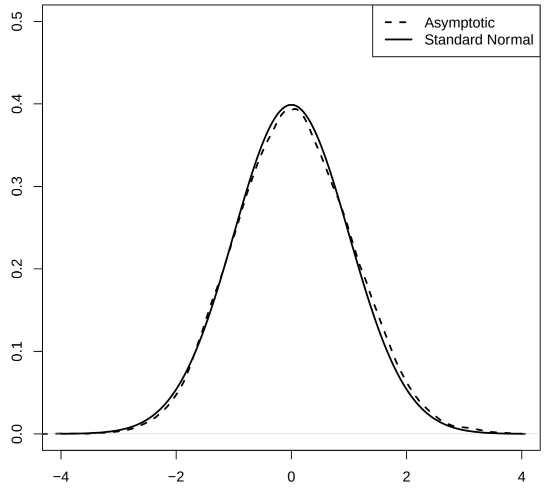
Provost, S. and Rudiuk, E. (1996). The exact distribution of indefinite quadratic forms in noncentral normal vectors. Annals of the Institute of Statistical Mathematics, 48:381–394.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. Journal of Multivariate Analysis, 55:331–339.

Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. In Neyman, J., editor, Proceedings of the third Berkeley symposium on mathematical and statistical probability. University of California, Berkley.

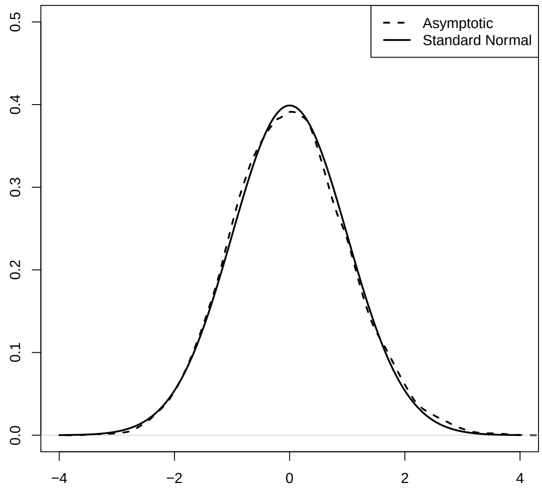
Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. Statistica Sinica, 25:993–1008.

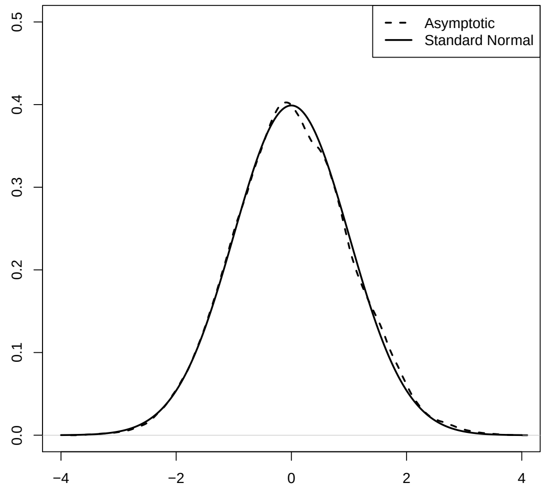




(a) $p = 50$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

(b) $p = 100$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

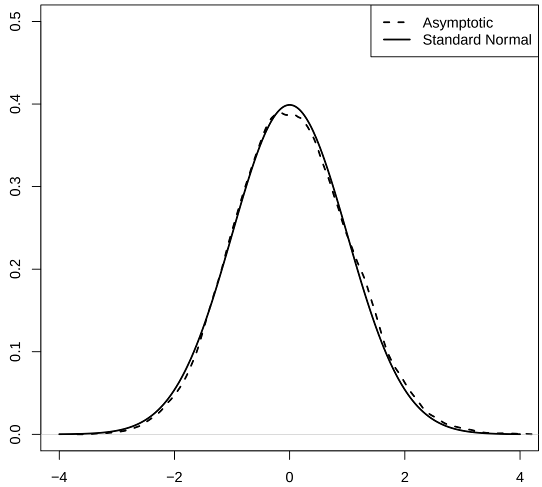




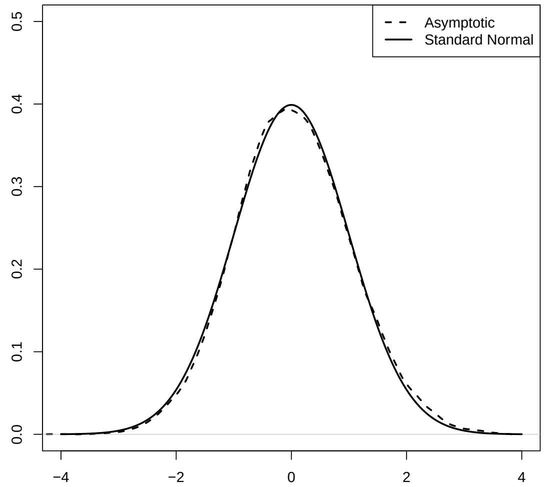
(c) $p = 50$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

(d) $p = 100$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

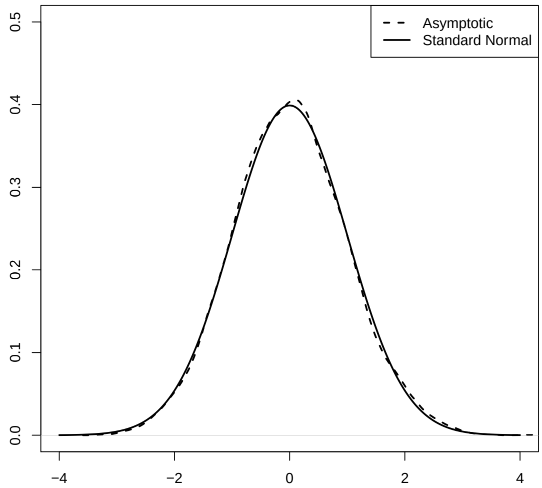
Figure 1: The kernel density estimator of the asymptotic distribution as given in Theorem 2 for $c = 0.1$.



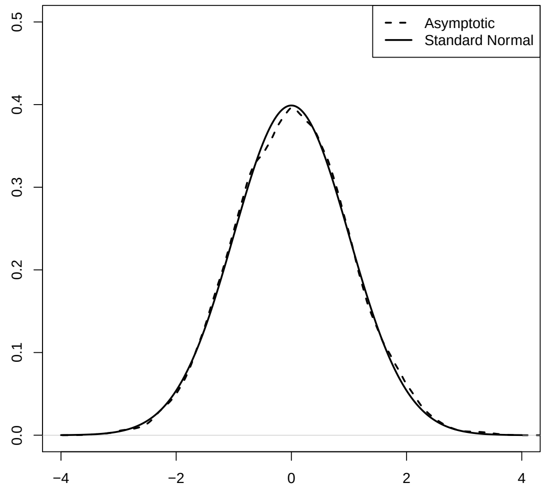
(a) $p = 250$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 500$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

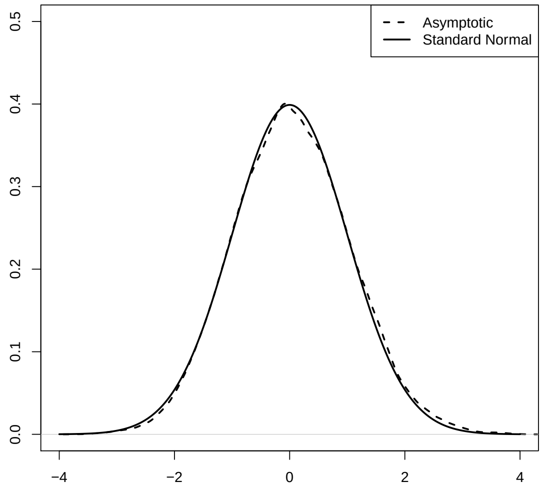


(c) $p = 250$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

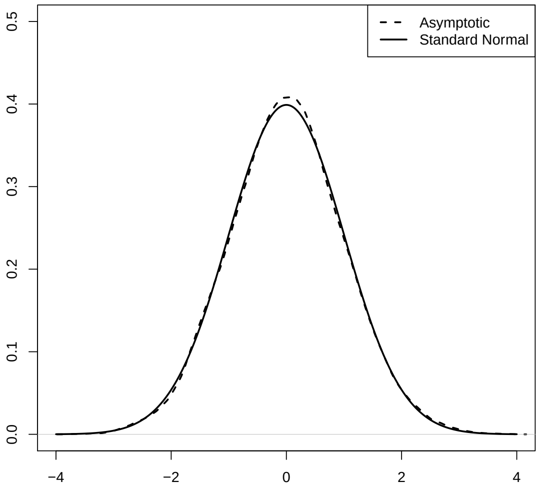


(d) $p = 500$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

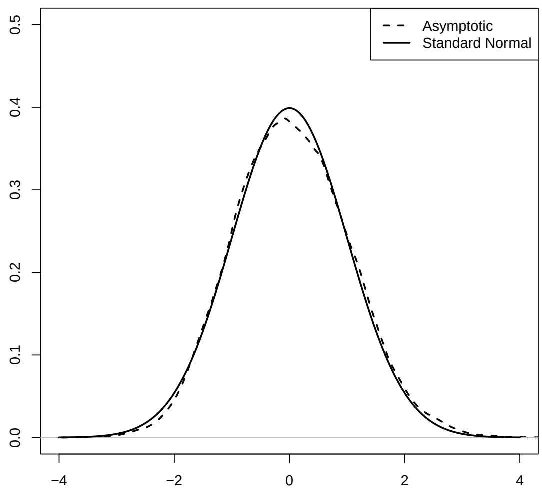
Figure 2: The kernel density estimator of the asymptotic distribution as given in Theorem 2 for $c = 0.5$.



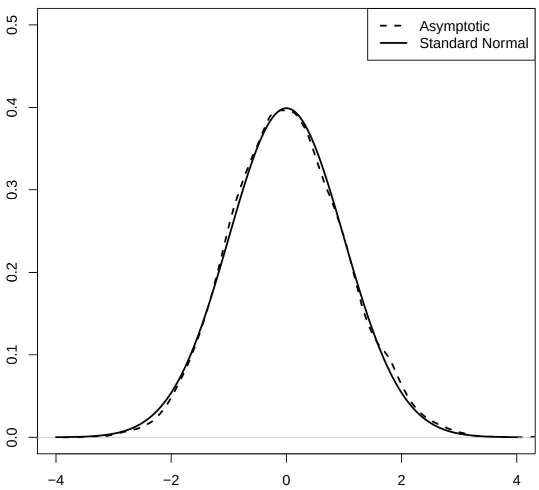
(a) $p = 400$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 800$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

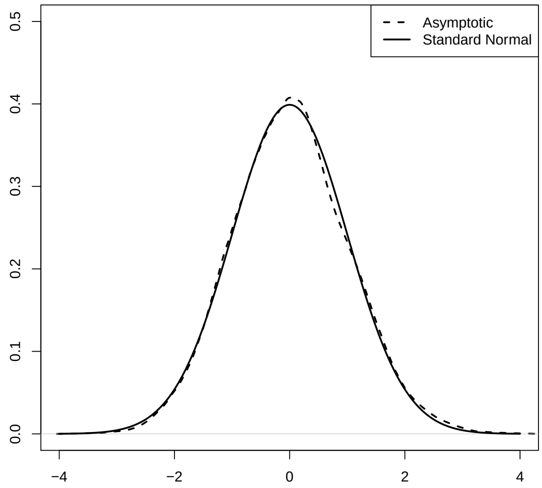


(c) $p = 400$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

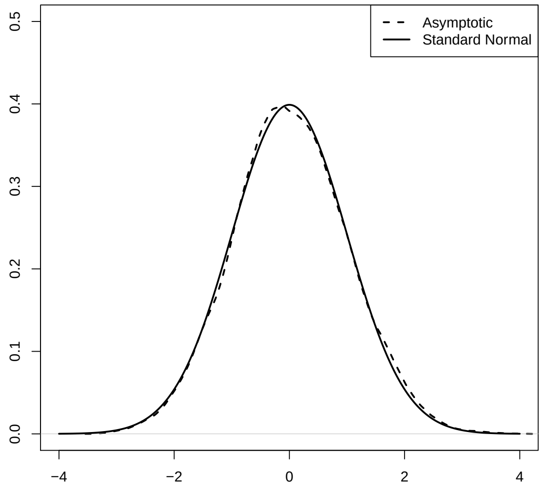


(d) $p = 800$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

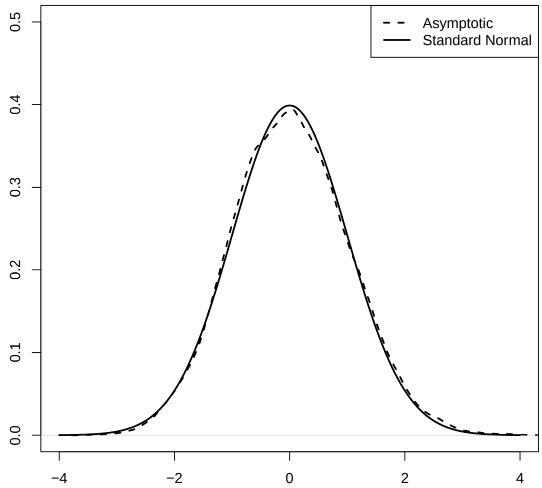
Figure 3: The kernel density estimator of the asymptotic distribution as given in Theorem 2 for $c = 0.8$.



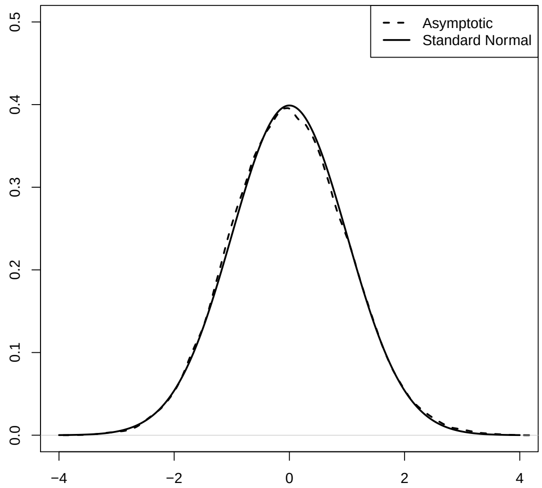
(a) $p = 475$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 950$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

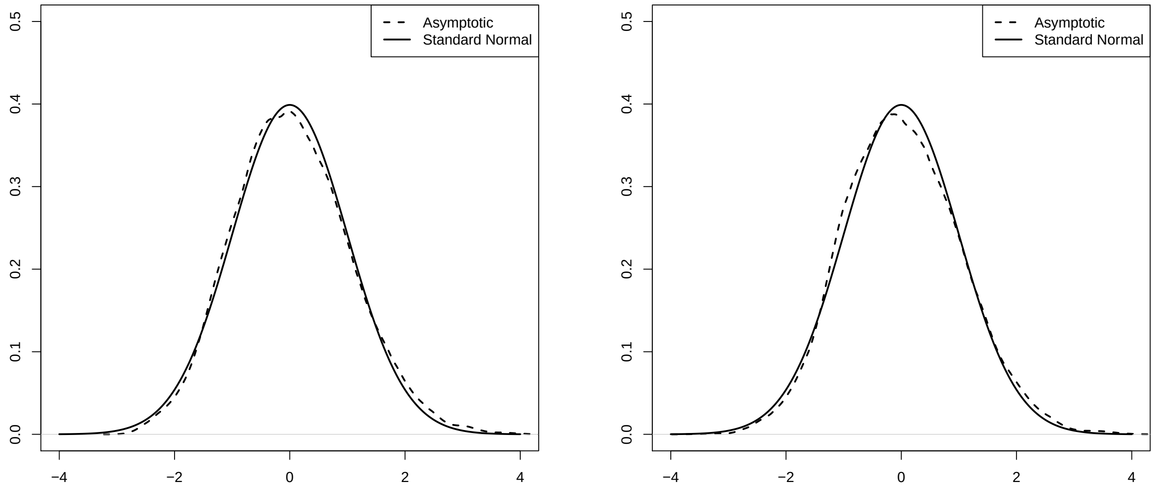


(c) $p = 475$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

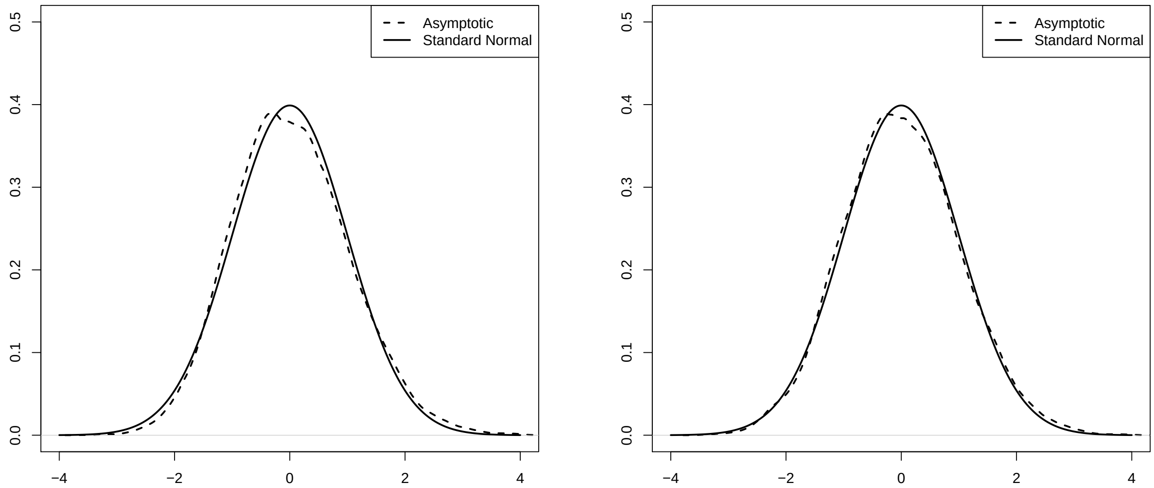


(d) $p = 950$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

Figure 4: The kernel density estimator of the asymptotic distribution as given in Theorem 2 for $c = 0.95$.

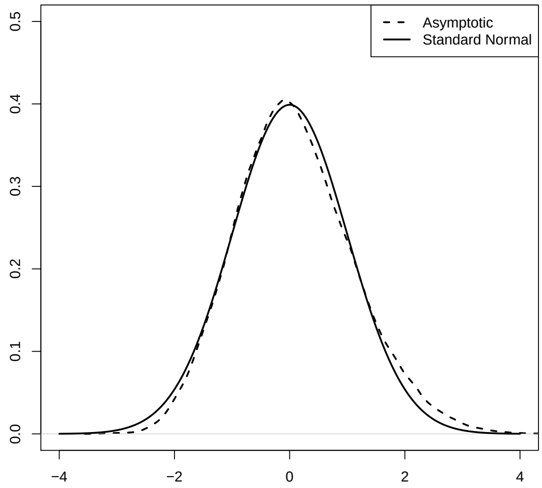


(a) $p = 50$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$. (b) $p = 100$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

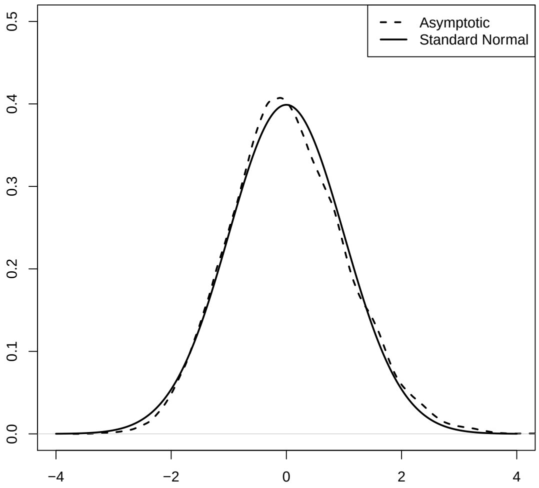


(c) $p = 50$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$. (d) $p = 100$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

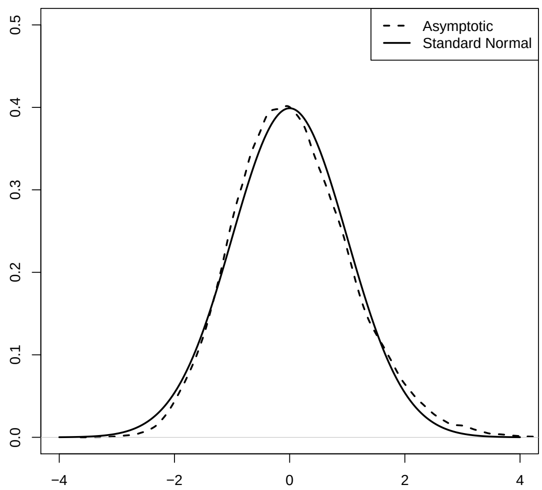
Figure 5: The kernel density estimator of the asymptotic distribution as given in Theorem 3 for $c = 0.1$.



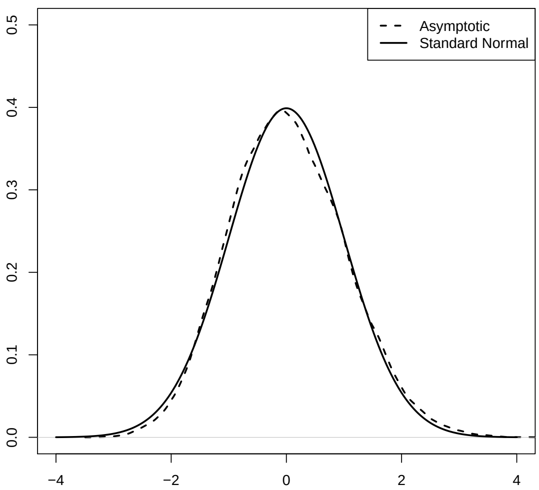
(a) $p = 250$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 500$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

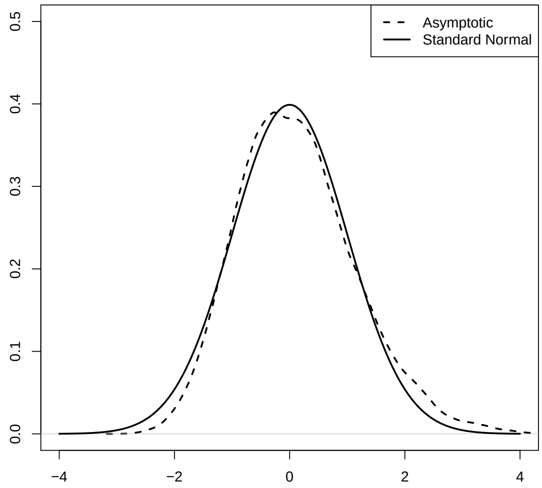


(c) $p = 250$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

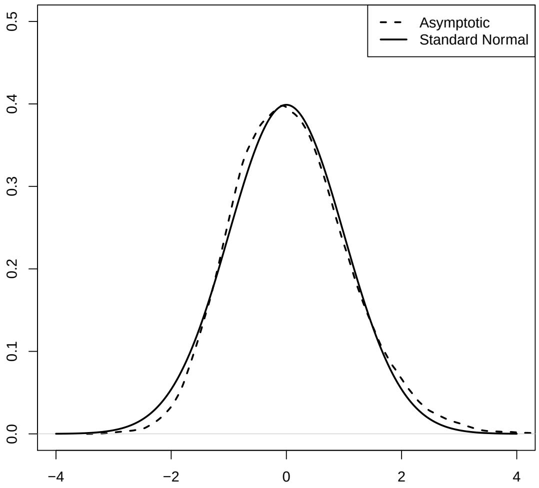


(d) $p = 500$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

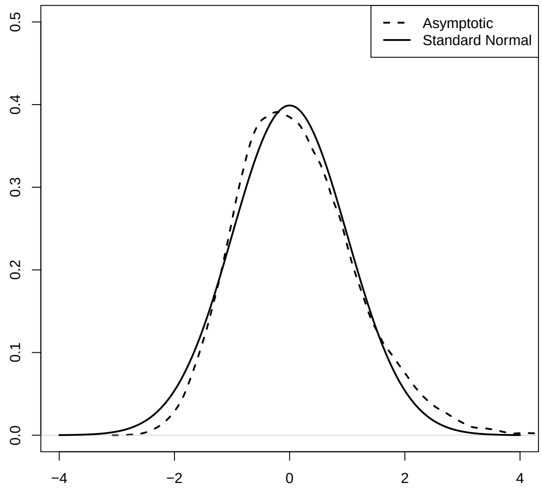
Figure 6: The kernel density estimator of the asymptotic distribution as given in Theorem 3 for $c = 0.5$.



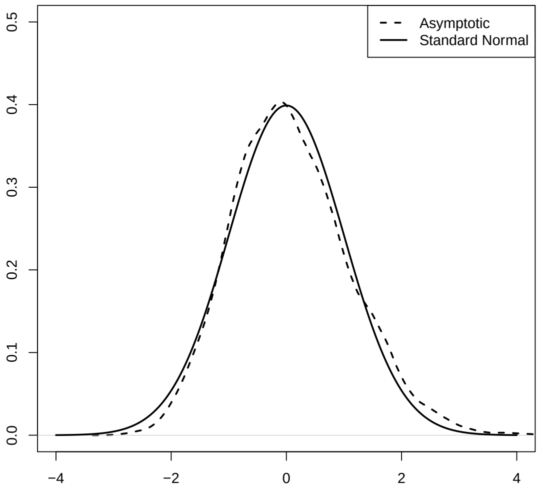
(a) $p = 400$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 800$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.

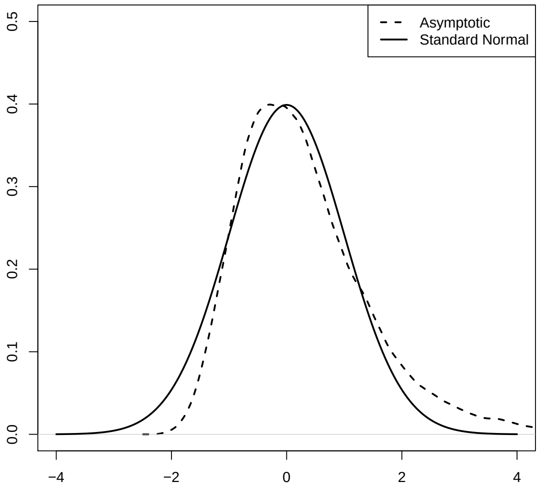


(c) $p = 400$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

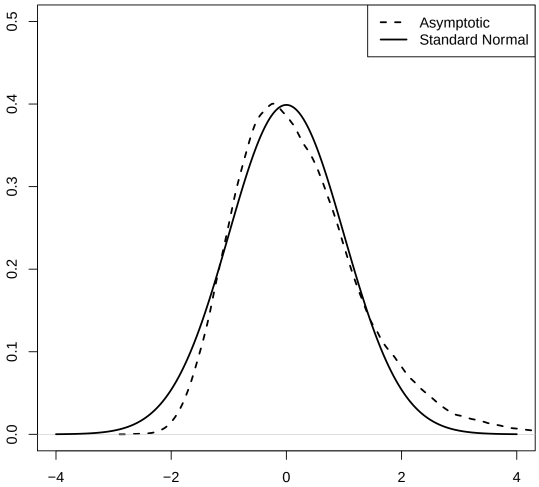


(d) $p = 800$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

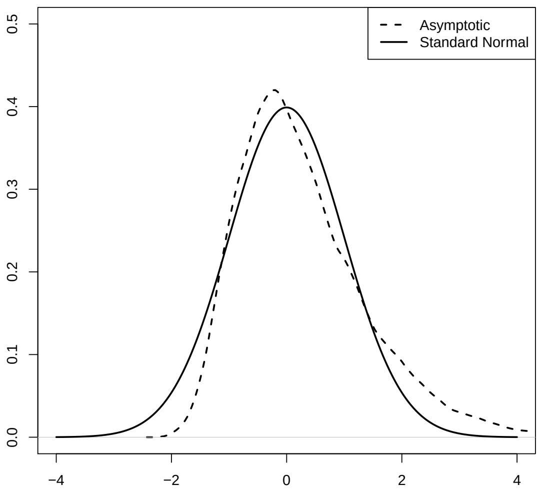
Figure 7: The kernel density estimator of the asymptotic distribution as given in Theorem 3 for $c = 0.8$.



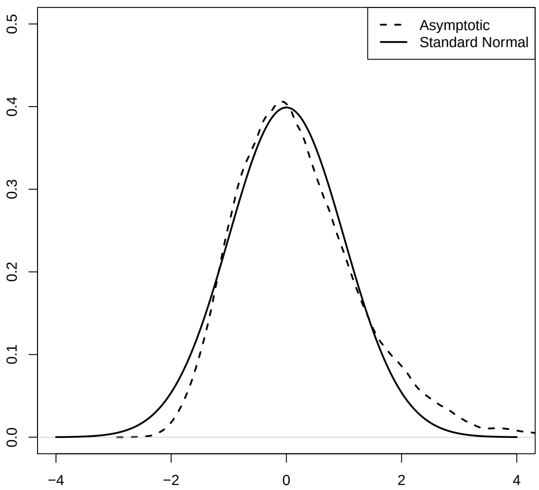
(a) $p = 475$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(b) $p = 950$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{TN}_q(\mathbf{0}, \mathbf{I}_q)$.



(c) $p = 475$, $n = 500$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.



(d) $p = 950$, $n = 1000$, $\boldsymbol{\nu} \sim \mathcal{GAL}_q(\mathbf{1}_q, \mathbf{I}_q, 10)$.

Figure 8: The kernel density estimator of the asymptotic distribution as given in Theorem 3 for $c = 0.95$.

http://journals.lub.lu.se/stat

LUND UNIVERSITY
School of Economics and Management

Working Papers in Statistics 2016
LUND UNIVERSITY
SCHOOL OF ECONOMICS AND MANAGEMENT
Department of Statistics
Box 743
220 07 Lund, Sweden