# Probability Distributions and GAS Models for Realized Covariance Matrices


Michael Stollenwerk

Heidelberg University


July 18, 2022


**Abstract**

Realized covariance matrices (RCs) are an important input to asses the risks involved in different investment allocations and it is thus useful to model and forecast them. To this end generalized autoregressive score (GAS) models are employed in this paper. These models are ideal for comparing different probability distributions in terms of their ability to model and forecast RCs, since the dynamic parameters of the conditional observation density are updated by incorporating the shape of the distribution itself (via the scaled score of the log-likelihood). All probability distributions so far applied to time series of RCs in the literature are compared and it is shown how they are related to each other. Furthermore a novel family of probability distribution, which has a property called "tail homogeneity", is derived and added to the comparison. The necessary inputs for the GAS models (Fisher information matrix and score) are derived for all distributions. An in-sample fit comparison confirms previous results that "fat-tailed" distributions outperform others and shows that the novel distribution family achieves very good fit. Out-of-sample forecasting comparisons further corroborate the excellent performance of the novel distribution family.

**Key words:** Realized Covariance Matrices, Generalized Autoregressive Score Models, Matrix Distributions, Time-Series Models

## 1 Introduction

In recent years there has been a lot of focus on the modeling and forecasting of time-series of realized covariance measures. In this strand of the literature the realized covariance measures are typically treated as "observed" measures of asset price variability, that can be modeled directly (see e.g. Chiriac and Voev 2011, Golosnoy, Gribisch, and Liesenfeld 2012, Opschoor et al. 2018, Gribisch and Stollenwerk 2020). The approach to treat realized measures created from high-frequency price observations as observations of volatility has been pioneered by Andersen et al. 2001 and can be likened to the macroeconomic literature where measurement-error prone variables of interest (e.g. GDP) are also treated as observables and directly modeled.[1] This view allows disregarding any distributional properties of the realized measures that might be inherited from an underlying assumption on the price processes (e.g. semi-martingale) and enables researchers to assume probability distributions on the realized measures directly.

Many different probability distributions with support on realized covariance measures (i.e. symmetric positive semi-definite matrices) have been proposed in the literature. These are the (non-central)-Wishart (Golosnoy, Gribisch, and Liesenfeld 2012, Gorgi et al. 2019, Yu, Li, and Ng 2017), inverse Wishart (Gourieroux, Jasiak, and Sufana 2009, Asai and So 2013), matrix-F (Opschoor et al. 2018, Zhou et al. 2019), Riesz (Gribisch and Hartkopf 2022), inverse Riesz and F-Riesz (both Blasques et al. 2021) distributions. All mentioned distributions are related to each other. The inverse Wishart, inverse Riesz, F and F-Riesz distributions accommodate the loosely defined stylized fact of "fat-tailedness" of realized measures. The recently proposed F-Riesz by Blasques et al. 2021 additionally allows for heterogeneous tails, i.e. it assumes that the variances of the realized variances can differ across assets.

In this paper, we compare, by showing explicitly how they are related to each other, all hitherto used probability distributions [2] in terms of their fit to different data sets

1. See also Andersen et al. 2003, Andersen et al. 2006 and McAleer and Medeiros 2008.

2. Excluding the noncentral Wishart distribution, which gives only slight improvements compared to the Wishart in terms of fit and forecasting ability and is not applicable to dimensions higher than five due to computational difficulties involving the matrix-variate hypergeometric function.

of time series of realized covariance measures and assess their forecasting performance. Furthermore, we propose a new family of distributions, also related to the other ones mentioned above and included in the comparison, which in contrast to Blasques et al. 2021 assumes tail homogeneity and show that, especially in times of high market volatility, this distribution represents a more realistic assumption to the data. It is also easier to handle analytically and numerically. It can be rooted in the assumption of a joint $t$-distribution on the vectors of intra-day returns, which generate realized covariance measures.

The rest of this paper is structured as follows. The next section presents all hitherto used probability distributions and the new family of probability distributions. Section 3 presents the GAS models for all distributions. Section 4 presents the data, Section 5 contains the empirical application. Section 6 concludes.

## 2 Probability Distributions

Let $\mathbf{R}$ denote a $p \times p$ realized covariance matrix. Any hitherto considered probability distribution can be characterized by its degree of freedom parameter(s), which we denote by $\boldsymbol{\theta}_d$ and a real symmetric positive semi-definite $p \times p$ parameter matrix $\boldsymbol{\Omega}_d$. Throughout the rest of this paper we will omit the subscript $d$ for all parameters, they are always to be understood as specific to the chosen distribution. This eases notation considerably.

If a probability distributions real symmetric positive semi-definite $p \times p$ expected value matrix

$$\mathbb{E}\left[\mathbf{R}\right] =: \boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top \tag{1}$$

exists, where $\mathbf{C}$ denotes the lower Cholesky factor, the distribution can be equivalently characterized by $\boldsymbol{\Sigma}$ and $\boldsymbol{\theta}$. In this paper we assume that $\boldsymbol{\Sigma}$ always exists[3], such that we

3. Existence depends on the degree of freedom parameter(s), but is not a restrictive assumption since for all estimated distributions in this paper, the mean exists.

can write

$$\mathbf{R} \sim d(\mathbf{\Sigma}, \boldsymbol{\theta}), \tag{2}$$

where $d \in (\mathcal{W}, i\mathcal{W}, t\mathcal{W}, it\mathcal{W}, \mathcal{F}, \mathcal{R}, i\mathcal{R}, t\mathcal{R}, it\mathcal{R}, \mathcal{FR}, i\mathcal{FR})$[4] indexes the different distributions. The probability density functions,

$$p_d(\mathbf{R}|\mathbf{\Sigma}, \boldsymbol{\theta}), \tag{3}$$

for all considered probability distributions are given in table 1. The composition of $\boldsymbol{\theta}$ for each distribution also becomes clear in this table.

4. See column 1, table 1 for all considered distributions

| Distribution | Probability Density Function, $p(\mathbf{R}\mid\boldsymbol{\Sigma},\boldsymbol{\theta})$ | | | |
|---|---|---|---|---|
| Wishart | $\frac{n^{np/2}}{2^{np/2}}\frac{1}{\Gamma_p(n/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert^{\frac{n}{2}}$ | $\operatorname{etr}\left(-\frac{1}{2}n\mathbf{Z}\right)$ |
| Riesz | $\frac{\prod_{i=1}^p n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}}\frac{1}{\Gamma_p(\mathbf{n}/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{\frac{\mathbf{n}}{2}}$ | $\operatorname{etr}\left(-\frac{1}{2}\operatorname{dg}(\mathbf{n})\,\mathbf{Z}\right)$ |
| Inverse Wishart | $\frac{(\nu-p-1)^{\nu p/2}}{2^{\nu p/2}}\frac{1}{\Gamma_p(\nu/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert^{-\frac{\nu}{2}}$ | $\operatorname{etr}\left(-\frac{1}{2}(\nu-p-1)\mathbf{Z}^{-1}\right)$ |
| Inverse Riesz | $\frac{\prod_{i=1}^p m_i^{-\nu_i/2}}{2^{p\bar{\boldsymbol{\nu}}/2}}\frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{-\frac{\boldsymbol{\nu}}{2}}$ | $\operatorname{etr}\left(-\frac{1}{2}\operatorname{dg}(\mathbf{m})^{-1}\mathbf{Z}^{-1}\right)$ |
| t-Wishart | $\left(\frac{n}{\nu-2}\right)^{pn/2}\frac{\Gamma((\nu+pn)/2)}{\Gamma_p(n/2)\Gamma(\nu/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert^{\frac{n}{2}}$ | $\left(1+\frac{n}{\nu-2}\operatorname{tr}(\mathbf{Z})\right)^{-\frac{\nu+pn}{2}}$ |
| t-Riesz | $\frac{\prod_{i=1}^p n_i^{n_i/2}}{(\nu-2)^{p\bar{\mathbf{n}}/2}}\frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{\frac{\mathbf{n}}{2}}$ | $\left(1+\frac{1}{\nu-2}\operatorname{tr}\left(\operatorname{dg}(\mathbf{n})\mathbf{Z}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}$ |
| Inverse t-Wishart | $\left(\frac{\nu-p-1}{n}\right)^{\nu p/2}\frac{\Gamma((n+p\nu)/2)}{\Gamma_p(\nu/2)\Gamma(n/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert^{-\frac{\nu}{2}}$ | $\left(1+\frac{\nu-p-1}{n}\operatorname{tr}(\mathbf{Z}^{-1})\right)^{-\frac{n+p\nu}{2}}$ |
| Inverse t-Riesz | $\frac{\prod_{i=1}^p (m_i^{i\mathcal{R}''})^{-\nu_i/2}}{n^{p\bar{\boldsymbol{\nu}}/2}}\frac{\Gamma((n+p\bar{\boldsymbol{\nu}})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma(n/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{-\frac{\boldsymbol{\nu}}{2}}$ | $\left(1+\frac{1}{n}\operatorname{tr}\left(\operatorname{dg}(\mathbf{m}^{i\mathcal{R}''})^{-1}\mathbf{Z}^{-1}\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}}$ |
| F | $\left(\frac{n}{\nu-p-1}\right)^{np/2}\frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert^{\frac{n}{2}}$ | $\left\lvert\mathbf{I}+\frac{n}{\nu-p-1}\mathbf{Z}\right\rvert^{-\frac{\nu+n}{2}}$ |
| F-Riesz | $\prod_{i=1}^p (m_i^{\mathcal{FR}'})^{n_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{\frac{\mathbf{n}}{2}}$ | $\left\lvert\mathbf{I}+\operatorname{dg}(\mathbf{m}^{\mathcal{FR}'})^{1/2}\mathbf{Z}\operatorname{dg}(\mathbf{m}^{\mathcal{FR}'})^{1/2}\right\rvert_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$ |
| Inverse F-Riesz | $\prod_{i=1}^p (m_i^{i\mathcal{R}''})^{-\nu_i/2}\frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)}$ | $\lvert\mathbf{R}\rvert^{-\frac{p+1}{2}}$ | $\lvert\mathbf{Z}\rvert_{-\frac{\boldsymbol{\nu}}{2}}$ | $\left\lvert\left(\mathbf{I}+\operatorname{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}\mathbf{Z}^{-1}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}\right)^{-1}\right\rvert_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}$ |

Table 1: Probability density functions of all considered distributions. Recall that $\mathbf{Z}=\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}$, where $\mathbf{C}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}$. The degrees of freedom parameters $\mathbf{n}=(n_1,n_2,\ldots,n_p)^\top$ and $\boldsymbol{\nu}=(\nu_1,\nu_2,\ldots,\nu_p)^\top$ are real parameter vectors, $n$ and $\nu$ are scalars. A bar on top of a vector denotes the average of its entries, e.g. $\bar{\mathbf{n}}=p^{-1}\sum_{i=1}^p n_i$, left arrow on top of a vector denotes the original vector in reverse order, e.g. $\overleftarrow{\mathbf{n}}=(n_p,n_{p-1},\ldots,n_1)^\top$. The well known gamma function $\Gamma(\cdot)$, the multivariate gamma function $\Gamma_p(\cdot)$ and $\Gamma_p(\cdot)$ with scalar argument are defined in equations (5.2.1), (35.3.5) and (35.3.6) of the NIST Digital Library of Mathematical Functions, respectively. The determinant with subscript, e.g. $\lvert\mathbf{Z}\rvert_{\frac{\mathbf{n}}{2}}$, denotes the power weighted determinant as in Blasques et al. 2021, also know as highest weight vector or generalized power function. For the definition of $\mathbf{m}^{i\mathcal{R}''}$, $\mathbf{m}^{\mathcal{FR}'}$ and $\mathbf{m}^{i\mathcal{FR}''}$ see equations (274), (399) and (417), respectively. See the appendix for further information and derivations.

Finally, let us denote

$$\mathbf{Z} = \mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top} \tag{4}$$

as the standardized realized covariance matrix.

## 2.1 Stochastic Representations

At the basis of all considered distributions lie the $p \times p$ random triangular matrices

$$\underline{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{n_1-1+1}} & 0 & \dots & 0 \\ \mathcal{N}(0,1) & \ddots & 0 & \vdots \\ \vdots & \mathcal{N}(0,1) & \ddots & 0 \\ \mathcal{N}(0,1) & \dots & \mathcal{N}(0,1) & \sqrt{\chi^2_{n_p-p+1}} \end{bmatrix} \tag{5}$$

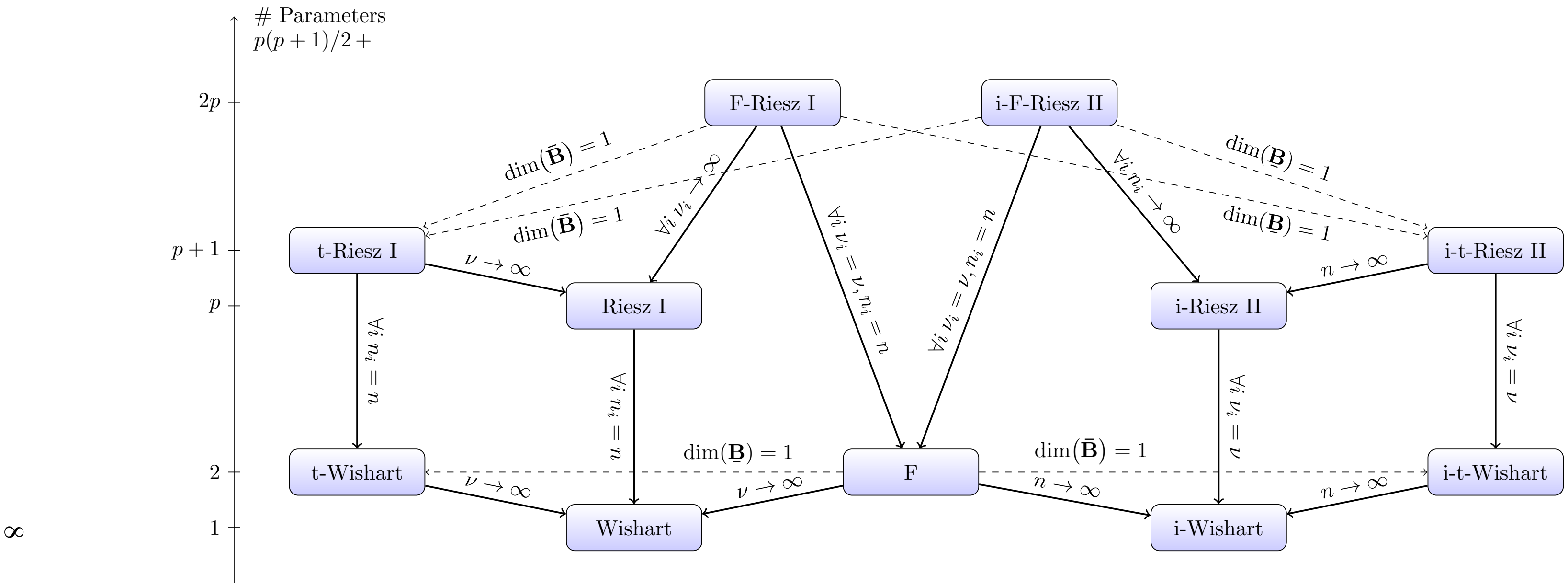
and/or

$$\bar{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{\nu_1-p+1}} & \mathcal{N}(0,1) & \dots & \mathcal{N}(0,1) \\ 0 & \ddots & \mathcal{N}(0,1) & \vdots \\ \vdots & 0 & \ddots & \mathcal{N}(0,1) \\ 0 & \dots & 0 & \sqrt{\chi^2_{\nu_p-p+p}}, \end{bmatrix}, \tag{6}$$

where all random variables inside the matrices are independent of each other. [5] For the distributions to exist it is thus necessary and sufficient that we restrict $n_i > i - 1$ and $\nu_i > p - i$, since otherwise the $\chi^2$ distributions on the main diagonals would not exist.[6] We denote the special cases where for all $i$, $n_i = n$ with $n > p - 1$ and $\nu_i = \nu$ with $\nu > p - 1)$ as $\underline{\mathcal{B}}$ and $\bar{\mathcal{B}}$, respectively.

5. We choose to name them with the letter "B" due to their resemblance of the "Barlett" decomposition.

6. Note that this does not imply existence of $\mathbb{E}[\mathbf{R}]$. For example the inverse Wishart distribution is based on $\left(\bar{\mathcal{B}}\bar{\mathcal{B}}^\top\right)^{-1}$ and its mean only exists if in fact $\nu > p + 1$.




| Stochastic Representations | | | |
|---|---|---|---|
| Wishart | $\frac{1}{n}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top$ | Riesz $I$ | $\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}$ |
| Inverse Wishart | $(n-p-1)\bar{\mathcal{B}}^{-\top}\bar{\mathcal{B}}^{-1}$ | Inverse Riesz $II$ | $\mathrm{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}$ |
| t-Wishart | $\frac{\nu-2}{\nu n}\Gamma^{-1}_{\left(\frac{\nu}{2},\frac{2}{\nu}\right)}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top$ | t-Riesz $I$ | $\frac{\nu-2}{\nu}\Gamma^{-1}_{\left(\frac{\nu}{2},\frac{2}{\nu}\right)}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}$ |
| Inverse t-Wishart | $(n-p-1)\Gamma_{\left(\frac{n}{2},\frac{2}{n}\right)}\bar{\mathcal{B}}^{-\top}\bar{\mathcal{B}}^{-1}$ | Inverse t-Riesz $II$ | $\Gamma_{\left(\frac{n}{2},\frac{2}{n}\right)}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}''})^{-\frac{1}{2}}$ |
| F | $\frac{\nu-p-1}{n}\bar{\mathcal{B}}^{-\top}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top\bar{\mathcal{B}}^{-1}$ | F-Riesz $I$ | $\mathrm{dg}(\mathbf{m}^{\mathcal{F}\mathcal{R}'})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{\mathcal{F}\mathcal{R}'})^{-\frac{1}{2}}$ |
| F | $\frac{\nu-p-1}{n}\underline{\mathcal{B}}\bar{\mathcal{B}}^{-\top}\bar{\mathcal{B}}^{-1}\underline{\mathcal{B}}^\top$ | Inverse F-Riesz $II$ | $\mathrm{dg}(\mathbf{m}^{i\mathcal{F}\mathcal{R}''})^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{m}^{i\mathcal{F}\mathcal{R}''})^{-\frac{1}{2}}$ |

Figure 1: Stochastic representations and the relationships between the considered probability distributions. Every arrow indicates a generalization, where the distribution further down is a nested by the one further up. The dashed arrows indicate a relation, where the distribution at the end of the arrow is related to but not nested by the other one. Note that the stochastic representations are for the distributions with expectation equal to the identity matrix. They have to be pre-multiplied by $\mathbf{C}$ and post-multiplied by $\mathbf{C}^\top$ in order to arrive at the general standardized distributions as in table 1.

The stochastic representations of and the relationships between the probability distributions are depicted in figure 1. The t-named distributions are novel distributions that will be explained in more detail in the section 2.2. Note that the stochastic representations in figure 1 are for the mean-$\mathbf{I}$ distributions, that is, all have expectation equal to the identity matrix. They have to be pre-multiplied by $\mathbf{C}$ and post-multiplied by $\mathbf{C}^\top$ in order to arrive at the distributions in table 1.

On the vertical axis of figure 1 we can see the number of parameters of the respective distributions. All have $p(p+1)/2$ distinct parameters in the symmetric positive semi-definite matrix $\mathbf{\Sigma}$ plus the number of degree of freedom parameters. Every Wishart-based distribution is a special case of its Riesz-named counterpart and is obtained by setting the entries in the degree of freedom parameter vector all equal to each other. Furthermore we can see that the (inverse) Riesz and (inverse) Wishart distributions are limiting cases of the other distributions, that have relatively more probability mass on extremely "large" RCs and are thus labeled "fat-tailed", if we let the parameters governing the "fat-tailedness" go to infinity. Note that the dashed arrows indicate that the distribution at the end of the arrow is not nested by but merely related to the one at the beginning of the arrow.

For each Riesz-named distribution the ordering of the assets in the RCs matters. That is, given an initial ordering, applying a different one to the rows and columns of the RCs and the expected value matrix, as well as to the degree of freedom parameter vectors, changes the probability distribution. This implies that in maximum likelihood estimation, we also have to optimize over the order of the assets, as suggested in Blasques et al. 2021.

Furthermore for all Riesz-named distributions there are two types, one based on $\underline{\mathbf{B}}$, the other on $\bar{\mathbf{B}}$. For example the Riesz type $II$ distribution has stochastic representation $\bar{\mathbf{B}}\bar{\mathbf{B}}^\top$ as opposed to $\underline{\mathbf{B}}\underline{\mathbf{B}}^\top$ for the Riesz type $I$ distribution. It can however be shown that a Riesz type $I$ distribution is the same as a Riesz type $II$ distribution, where the order of assets in the data matrix and in $\mathbf{\Sigma}$ and in the entries of the parameter vectors are reversed. The same holds for the standardized versions of all distributions.[7]. Thus when optimizing over the asset order, only one of the two types needs to be considered. Finally, note that an inverse F distribution is again an F distribution[8], but an inverse F-Riesz distribution is not again an F-Riesz distribution.

### 2.2 The t-Riesz Distribution Family

Recall the stochastic representation for the t-Riesz and the inverse t-Riesz distribution,

$$\frac{\nu-2}{\nu}\Gamma^{-1}_{\left(\frac{\nu}{2},\frac{2}{\nu}\right)}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \text{ and } \Gamma_{\left(\frac{n}{2},\frac{2}{n}\right)}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}, \tag{7}$$

respectively. Note that, given $\underline{\mathbf{B}}$ ($\bar{\mathbf{B}}$), a tail realization of the inverse gamma (gamma) distribution yields a tail realization of the t-Riesz (inverse t-Riesz) distribution. This in stark contrast to the F-Riesz and inverse F-Riesz distribution, which have stochastic representations

$$\mathrm{dg}(\mathbf{m}^{\mathcal{F}\mathcal{R}^{I}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{\mathcal{F}\mathcal{R}^{I}})^{-\frac{1}{2}} \text{ and } \mathrm{dg}(\mathbf{m}^{i\mathcal{F}\mathcal{R}^{II}})^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{m}^{i\mathcal{F}\mathcal{R}^{II}})^{-\frac{1}{2}}, \tag{8}$$

respectively. Here, given $\underline{\mathbf{B}}$ ($\bar{\mathbf{B}}$), a tail realization of one of the entries on the main diagonal of $\bar{\mathbf{B}}$ ($\underline{\mathbf{B}}$), while the other diagonal entries do not lie in the tail, yields a tail realization in some of the entries the (inverse) F-Riesz matrix, but has a decreasing effect on the other entries the further away they are from the index where the tail observation occurred. Blasques et al. 2021 call this property of the F-Riesz distribution "tail heterogeneity", which leads us to calling the t-Riesz distribution family being "tail homogeneous". The natural conjecture would be that the t-Riesz distribution family might work better in times of market-wide crises, while the F-Riesz distribution has advantages when one asset or subsections of the market are in distress.

The t-Riesz, inverse t-Riesz, inverse t-Wishart and inverse F-Riesz distributions are to the best of our knowledge novel distributions that we derive in the appendix. The

7. See Theorem 7.2 in the appendix.
8. With the degrees of freedom parameters switched and the expected value matrix inverted, as is easy to see from their stochastic representations.

t-Wishart distribution is a standardized version of the distribution introduced in Sutradhar and Ali 1989. They prove an interesting result that can be readily applied to the realized covariance estimator, namely that if we assume a joint t-distribution with block diagonal scale matrix, where the blocks are all equal, on all intraday-return vectors, then the realized covariance follows a t-Wishart distribution where the degree of freedom $n$ equals the number of intraday-return vectors. The Wishart distribution can be based on a similar assumption, but with a Normal distribution instead of a t-distribution. As such, the Wishart and t-Wishart distributions are the only ones of the considered distributions for $\mathbf{R}$, which can be grounded in an assumption on the underlying intraday-return vectors, but it is common knowledge, that the t-distribution assumption is much more realistic that the Normal distribution assumption for return vectors. This is mirrored in the much superior performance of the t-Wishart compared to the Wishart as will be visible in the empirical part of this paper.

Other advantages of the novel distributions are that they have less parameters than the F(-Riesz) and that their evaluation is numerically more stable as their pdfs depend on the trace, rather than the (power weighted) determinant of $\mathbf{Z}$.

## 2.3 Static Estimation

Now lets add subscripts for the days in our sample, $t = 1, \ldots, T$. In a first step we will assume a static distribution on the time series of RCs, that is

$$\mathbf{R}_t \overset{iid}{\sim} d(\boldsymbol{\Sigma}, \boldsymbol{\theta}), \tag{9}$$

where $\boldsymbol{\Sigma}$ in theory is allowed to differ across distributions. In practice, however, we choose to estimate $\boldsymbol{\Sigma}$ with the obvious method of moments estimator to avoiding the curse of dimensionality for large cross-sections,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T} \sum_{t=1}^{T} \mathbf{R}_t, \tag{10}$$

which of course is the same across distributions. Then, in a second step we estimate the degrees of freedom parameters via standard numerical maximum likelihood estimation, conditional on our estimate for $\mathbf{\Sigma}$. I follow the algorithm proposed in Blasques et al. 2021 to optimize over the asset ordering. The seed for the random generation of permutations to try, is the same for all Riesz-named distributions.

# 3 GAS Models

In the literature on time series of RCs it is standard to assume time-variation in the mean, i.e. in the $\mathbf{\Sigma}$ matrix of the underlying distribution, while leaving the degree of freedom parameters fixed over time, that is

$$\mathbf{R}_t|\mathbb{F}_{t-1} \sim d(\mathbf{\Sigma}_t, \boldsymbol{\theta}), \tag{11}$$

where $\mathbb{F}_{t-1}$ represents all information up to time $t-1$. There are different proposals for the updating mechanism of $\mathbf{\Sigma}_t$ in the literature. For example one could simply assume a multivariate GARCH(1,1)-type recursion, $\mathbf{\Sigma}_t = \mathbf{\Xi} + a\mathbf{R}_{t-1} + b\mathbf{\Sigma}_{t-1}$, where $\mathbf{\Xi}$ is a symmetric positive semi-definite $p \times p$ parameter matrix. A natural choice for the updating mechanism in order to compare different probability distributions is the Generalized Autoregressive Score (GAS) framework introduced by Creal, Koopman, and Lucas 2013 and Creal, Koopman, and Lucas 2011, because it incorporates directly information about the shape of the distributions into the updating process. The first example of such a model applied to time-series of RCs is given by Gorgi et al. 2019, who apply the GAS framework to a Wishart distribution on a set of two-dimensional time series of RCs. The standard recursion in a GAS framework is

$$\mathbf{\Sigma}_t = (1-b)\mathbf{\Xi} + a\mathbf{S}_{t-1} + b\mathbf{\Sigma}_{t-1}, \tag{12}$$

where $\mathbf{S}_t$ is the scaled score of the assumed probability distribution $d$ at time $t$. One popular choice for the scaling of the score,

$$\nabla_t = \frac{\partial \log p_d(\boldsymbol{\Sigma}_t, \boldsymbol{\theta})}{\partial \text{vech}\left(\boldsymbol{\Sigma}_t\right)}, \tag{13}$$

is the inverse of the Fisher information matrix,

$$\mathcal{I}_t = -\mathbb{E}\left[\frac{\partial \log p_d(\boldsymbol{\Sigma}_t, \boldsymbol{\theta})}{\partial \text{vech}\left(\boldsymbol{\Sigma}_t\right)} \frac{\partial \log p_d(\boldsymbol{\Sigma}_t, \boldsymbol{\theta})}{\partial \text{vech}\left(\boldsymbol{\Sigma}_t\right)^\top}\right], \tag{14}$$

such that

$$\mathbf{S}_t = \text{ivech}\left(\mathcal{I}_t^{-1} \nabla_t\right), \tag{15}$$

where $p_d$ represents the probability density function of distribution $i$. In this paper we derive the scores $\nabla_t$ and Fisher information matrices $\mathcal{I}_t$ for all probability distributions[9] considered above. These can be found, omitting the subscripts, in table 2 and 3, respectively.

Note that how we parameterize the distributions makes a difference in how the GAS models are defined. GAS models which use the scores w.r.t. $\boldsymbol{\Omega}$ are different from the ones above. There are several advantages in using the score w.r.t. $\boldsymbol{\Sigma}$. First, unlike $\boldsymbol{\Omega}$, the $\boldsymbol{\Sigma}$ parameter has the same meaning across distributions and an intuitive interpretation of yielding a mean-shifting process, which is in line with existing RC models. This makes the recursion parameters $a$ and $b$ easily comparable across the different GAS models. Second, the standardized distributions nest each other according to the distribution family tree in figure 1 and this nesting translates over to the GAS models. This makes comparisons of likelihood values via information criteria valid and opens the possibility of likelihood ratio tests. One could also use the optimized parameters of a nested GAS model as starting points for the estimation of the nesting GAS model. Third, the targeting estimation of the intercept matrix becomes very easy. In fact, if we introduce

9. Except for the Fisher information matrix of the (inverse) F-Riesz distribution.

| Distribution | Score, $\nabla = \mathbf{G}^\top \operatorname{vec}(\triangle)$ |
|---|---|
| Wishart | $\frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left(n\,\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1} - n\,\boldsymbol{\Sigma}^{-1}\right)$ |
| Inverse Wishart | $-\frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left((\nu-p-1)\mathbf{R}^{-1} - \nu\boldsymbol{\Sigma}^{-1}\right)$ |
| t-Wishart | $\frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left(n\,\frac{\nu+pn}{\nu-2+n\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{R})}\boldsymbol{\Sigma}^{-1}\mathbf{R}\boldsymbol{\Sigma}^{-1} - n\,\boldsymbol{\Sigma}^{-1}\right)$ |
| Inverse t-Wishart | $-\frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left(\frac{(n+p\nu)(\nu-p-1)}{n+(\nu-p-1)\operatorname{tr}(\boldsymbol{\Sigma}\mathbf{R}^{-1})}\mathbf{R}^{-1} - \nu\boldsymbol{\Sigma}^{-1}\right)$ |
| F | $-\frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left((n+\nu)\left(\boldsymbol{\Sigma} + \frac{n}{\nu-p-1}\mathbf{R}\right)^{-1} - \nu\boldsymbol{\Sigma}^{-1}\right)$ |
| Riesz | $\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n})\,\mathbf{Z} - \mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n})\right)\right)\mathbf{C}^{-1}\right)$ |
| Inverse Riesz | $-\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\mathbf{R}^{-1}\mathbf{C}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} - \mathbf{C}^{-\top}\operatorname{dg}(\boldsymbol{\nu})\right)\right)\mathbf{C}^{-1}\right)$ |
| t-Riesz | $\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\frac{\nu+p\bar{\mathbf{n}}}{\nu-2+\operatorname{tr}(\operatorname{dg}(\mathbf{n})\mathbf{Z})}\mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n})\,\mathbf{Z} - \mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n})\right)\right)\mathbf{C}^{-1}\right)$ |
| Inverse t-Riesz | $-\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\frac{n+p\bar{\boldsymbol{\nu}}}{n+\operatorname{tr}(\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{Z}^{-1})}\mathbf{R}^{-1}\mathbf{C}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} - \mathbf{C}^{-\top}\operatorname{dg}(\boldsymbol{\nu})\right)\right)\mathbf{C}^{-1}\right)$ |
| F-Riesz | $\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\mathbf{C}^{-\top}\operatorname{dg}(\boldsymbol{\nu}) - \mathbf{C}_{\mathbf{B}}^{-\top}\operatorname{dg}(\boldsymbol{\nu}+\mathbf{n})\mathbf{C}_{\mathbf{B}}^{-1}\mathbf{C}\operatorname{dg}(\mathbf{m}^{\mathcal{FR}^{I}})^{-1}\right)\right)\mathbf{C}^{-1}\right)$ |
| Inverse F-Riesz | $-\mathbf{G}^\top \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}\left(\mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n}) - \mathbf{C}^{-\top}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\mathbf{C}_{\mathbf{B}_2}\operatorname{dg}(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}_{\mathbf{B}_2}^\top\mathbf{C}^{-\top}\right)\right)\mathbf{C}^{-1}\right)$ |

Table 2: Scores w.r.t. $\boldsymbol{\Sigma}$ of all considered distributions. The subscripts are omitted for readability. $\mathbf{G}$ denotes the duplication matrix. $\mathbf{C}_{\mathbf{B}}$ denotes the lower Cholesky factor of $\mathbf{B} = \mathbf{C}\operatorname{dg}(\mathbf{m}^{\mathcal{FR}^{I}})^{-1}\mathbf{C}^\top + \mathbf{R}$. $\mathbf{C}_{\mathbf{B}_2}$ denotes the lower Cholesky factor of $\mathbf{B}_2 = \left(\mathbf{C}^{-\top}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^{-1} + \mathbf{R}^{-1}\right)^{-1}$. The tril($\mathbf{X}$) function returns a lower triangluar matrix by setting all elements of $\mathbf{X}$ above the main diagonal equal to zero, the $\Phi(\mathbf{X})$ function is defined in 7.2. It returns a lower triangluar matrix by setting all elements of $\mathbf{X}$ above the main diagonal equal to zero and halving the entries on the main diagonal. See the appendix for the derivations.

| Distribution | Fisher Information Matrix $\mathcal{I}$ |
|---|---|
| Wishart | $\frac{n}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G}$ |
| Inverse Wishart | $-\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G}$ |
| t-Wishart | $\frac{n}{2}\mathbf{G}^\top\left(\frac{\nu+pn}{\nu+pn+2}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right) - \frac{n}{(\nu+pn+2)}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}$ |
| Inverse t-Wishart | $-\frac{\nu}{2}\mathbf{G}^\top\left(\frac{n+p\nu}{n+p\nu+2}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right) - \frac{\nu}{(n+p\nu+2)}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}$ |
| F | $\frac{1}{2}\mathbf{G}^\top\left((\nu+(n+\nu)(c_3+c_4))\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right) + (n+\nu)c_4\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}$ |

Table 3: Fisher information matrices of all considered Wishart-based distributions. The subscripts $i$ and $t$ are omitted for readability. $\mathbf{G}$ denotes the duplication matrix. For the derivations and the definitions of $c_3$ and $c_4$ see equations (500) and (499), respectively. The Fisher information matrices of the (inverse) Riesz and t-Riesz distributions are derived in the appendix as well.

time variation in the degrees of freedoms, it would not be possible anymore to target the intercept matrix if we model $\mathbf{\Omega}$ as in Blasques et al. 2021, but if we model $\mathbf{\Sigma}$ the targeting stays the same.

Unfortunately, the Fisher information matrices of Riesz-named distributions involve matrix multiplications and inversions of $p^2 \times p^2$ matrices, which causes computation of the scaled scores in equation (15) to be prohibitively slow for estimation of GAS models of dimension say $p > 10$. Luckily, the most important part of the scaled scores are the scores themselves, which at time $t$ define the steepest ascend direction of the log-likelihood, in which the parameters should be updated. The scaling is of secondary importance. If we scale the Riesz-named distribution scores with the Fisher information matrices of their Wishart-based counterparts, i.e. we set all parameters in a degree of freedom parameter vector equal to the average, then theorem 3.1 shows that the computation of $\mathbf{S}_t$ reduces to $p \times p$ matrix operations for all distributions.[10] The scaling of the scores is in fact of minor importance and there is no standardized way to do it, Opschoor et al. 2018 simply scale with the Fisher information matrix of the Wishart distribution. Furthermore, our proposed scaling is in line with Blasques, Francq, and Laurent 2022, who show that it is possible to disentangle the distributional assumption from the score dynamics, i.e. one could update the parameters with a scaled score from a different distribution that the conditional distributional assumption on the data.

**Theorem 3.1.** *Consider, omitting the subscript $t$,*

$$\mathbf{S} = \operatorname{ivech}\left(\mathcal{I}^{-1}\nabla\right), \tag{16}$$

*as defined in equation (15). For any Riesz-named distribution use $\mathcal{I}$ of its Wishart-based counterpart instead of its own, by setting the degree(s) of freedom equal to the average*

10. Apart from making the GAS models applicable to high dimensions this also circumvents the difficulties of obtaining the F-Riesz Fisher information matrix as its Wishart-based nested version has been derived in this paper.

*of the corresponding degree of freedom parameter vector(s). Then*

$$\mathbf{S} = \frac{\alpha_{\boldsymbol{\theta}}}{2}\boldsymbol{\Sigma}\left(\triangle + \triangle^{\top}\right)\boldsymbol{\Sigma} + \beta_{\boldsymbol{\theta}}\,\mathrm{tr}\left(\boldsymbol{\Sigma}\triangle\right)\boldsymbol{\Sigma}, \tag{17}$$

*where $\triangle$ is the score matrix w.r.t. $\boldsymbol{\Sigma}$, ignoring symmetry,*[11] *and $\alpha$ and $\beta$ depend only on the degree of freedom parameters of the respective distribution.*

Using Theorem 3.1 in the standard GAS recursion (12) and we arrive at

$$\boldsymbol{\Sigma}_t = (1-b)\,\boldsymbol{\Xi} + a\frac{\alpha_{\boldsymbol{\theta}}}{2}\,\boldsymbol{\Sigma}_t\left(\triangle_t + \triangle_t^{\top}\right)\boldsymbol{\Sigma}_t + a\beta_{\boldsymbol{\theta}}\,\mathrm{tr}\left(\boldsymbol{\Sigma}_t\triangle_t\right)\boldsymbol{\Sigma}_t + b\,\boldsymbol{\Sigma}_{t-1}. \tag{18}$$

From here it is a small cost to further generalize the model to

$$\boldsymbol{\Sigma}_t = (1-c)\,\boldsymbol{\Xi} + a\,\boldsymbol{\Sigma}_t\left(\triangle_t + \triangle_t^{\top}\right)\boldsymbol{\Sigma}_t + b\,\mathrm{tr}\left(\boldsymbol{\Sigma}_t\triangle_t\right)\boldsymbol{\Sigma}_t + c\,\boldsymbol{\Sigma}_{t-1}, \tag{19}$$

where $a$, $b$ and $c$ are now stand-alone parameters, independent of the degree of freedom parameters. Equation (19) is the recursion we choose in all our empirical applications.

For "fat-tailed" distributions GAS models down-weight the impact of extreme realizations of $\mathbf{R}_t$ on the updating process of $\boldsymbol{\Sigma}_t$, since extreme realizations of $\mathbf{R}_t$ are less unexpected by a model with a "fat-tailed" distribution and thus yield less extreme scaled score realizations $\mathbf{S}_t$. This type of modeling behavior has also been advocated for in the literature by e.g. Bollerslev, Patton, and Quaedvlieg 2018, whos "Dynamic Attenuation Model" down-weights the impact of extreme realizations by incorporating the fact that they are relatively more inaccurate estimates of integrated covariance, "endogenously shrinking the influence of past realized covariances based on dynamically varying weights determined by an estimate of the reliability of the realized covariances".

As a final interesting point it holds that when we scale the (inverse) Riesz or Wishart distributions by their inverse Fisher information matrix, we obtain simple the GARCH-type dynamics mentioned before.[12]

11. $\nabla = \frac{\partial \log p(\boldsymbol{\Sigma},\boldsymbol{\theta})}{\partial \mathrm{vec}(\boldsymbol{\Sigma})^{\top}}\frac{\partial \mathrm{vec}(\boldsymbol{\Sigma})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}} = \mathrm{vec}(\triangle)^{\top}\mathbf{G} \Rightarrow \triangle = \mathrm{ivec}\left(\frac{\partial \log p(\boldsymbol{\Sigma},\boldsymbol{\theta})}{\partial \mathrm{vec}(\boldsymbol{\Sigma})^{\top}}\right) = \frac{\partial \log p(\boldsymbol{\Sigma},\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}}$

12. See appendix.

### 3.1 Estimation of the GAS Models

Since $\mathbb{E}[\mathbf{S}_t] = \mathbf{0}$, it is easy to show that for equation (12) it holds that

$$\mathbb{E}[\mathbf{R}_t] = \boldsymbol{\Xi}. \tag{20}$$

We thus choose the well established two step estimation method where we "target" $\boldsymbol{\Xi}$ in the first step, that is we simply apply the obvious method of moments estimator for it

$$\widehat{\boldsymbol{\Xi}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t, \tag{21}$$

and we estimate $a$, $b$ and the degree of freedom parameters in a second step via standard numerical maximum likelihood estimation, conditional on our estimate for $\boldsymbol{\Xi}$ in the first step. Again, we follow the algorithm proposed in Blasques et al. 2021 to optimize over the asset ordering, where the two-step maximum likelihood estimation described above, goes into step 2 of their 4-step estimation algorithm.

## 4 Data

Our original data are one-minute close prices from all trading days from 1 January 1998 to 05 February 2021 for every stock that was a constituent of the S&P 500 index during the sample period. A close price is defined as the latest observed trade price of the respective one-minute interval and as such is we have previous tick interpolation on a fixed one-minute grid. We acquired the data from Quantquote[13], who combine, clean and process data directly obtained different exchanges, where the biggest are NYSE, NASDAQ and AMEX[14]. The data include observations from official trading hours as well as before- and after hour trading observations.

The aim is to produce the longest possible time series of accurately estimated daily

13. The company is recommended by the Caltech Quantitative Finance Group, see http://quant.caltech.edu/historical-stock-data.html.

14. AMEX was bought by NYSE in 2008, and handled only 10% of trades at its height

integrated covariance estimators. We exclude dates before 1 January 2002, because the NYSE fully implemented decimal pricing in 2001[15] and there are numerous other trading irregularities during this year[16]. This leaves 4808 trading days. To be consistent across trading days we only keep observations from official trading hours. We then exclude all stocks that on at least one of the remaining trading days have missing observations on more than 25% minutes. From an initial sample of 983 stocks this leaves 99 stocks.[17] Excluding illiquid stocks is common practice in creating time-series of integrated covariance estimators (see e.g. Lunde, Shephard, and Sheppard 2016). While this procedure biases the sample towards stocks which where very liquid over the entire sample period[18] it does ensure that for those stocks included the integrated covariance estimates are accurate.

I follow Opschoor et al. 2018 and Blasques et al. 2021 and construct realized covariance matrices of the 99 assets using five-minute returns with subsampling[19]. Then we randomly choose two 5- and 10-dimensional principal submatrices and one 25- and 50-dimensional principal submatrix for a total of six datasets.

For a view of the data see figure 2, which shows the annualized realized volatility for Cisco (csco) and Goldman Sachs (gs), as well as their realized correlation and the log-determinant of the 99-asset realized covariance matrices. We see that the spikes in volatility are of similar magnitude in the recent COVID-19 induced market turmoil for both assets, while the global financial crises of 2008/2009 as expected caused volatility to spike much higher for Goldman Sachs than for Cisco. The dot-com crisis on the other hand causes more volatility for Cisco, also as expected. We see that correlations are mainly positive and more stable around crises periods. Finally we see that the log-determinant of $\mathbf{R}_t$, as a measure of the size of the RCs, does indeed spike in the aforementioned market turmoil periods (dot-com, COVID, global financial crisis).

15. On 29 January 2001 to be precise.

16. e.g. the days surrounding the terrorist attacks on 11 September 2001 and "computer systems connectivity problems" on 8 June 2001.

17. 465 are left after excluding those that do not have an observation at all on at least one trading day.

18. Relatively young firms (e.g. Facebook or Tesla) are excluded.

19. The subsampling estimator was first proposed by Zhang, Mykland, and Aït-Sahalia 2005

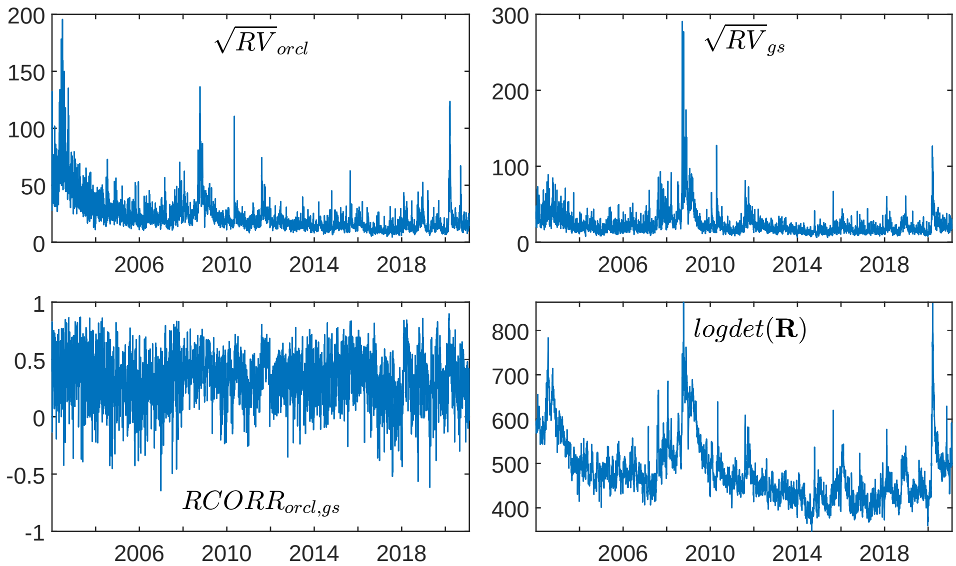



Figure 2: Top row: Realized volatilities of Cisco (csco) and Goldman Sachs (gs), i.e. the square root of two of the elements on the main diagonal of the 99 assset $\mathbf{R}_t$ for the complete sample (1 January 2002 to 05 February 2021). Bottom row:

## 5 Empirical Application

In this section we estimate the static distributions and GAS models based all distributions to the data described above.

### 5.1 In-Sample Fit

Notice, as is easily seen from the stochastic representations, that for $c > 0$,

$$\mathbf{R} \sim d(\mathbf{\Sigma}, \boldsymbol{\theta}) \Leftrightarrow c\mathbf{R} \sim d(c\mathbf{\Sigma}, \boldsymbol{\theta}). \tag{22}$$

From table 1 it is however obvious that, since $\mathbf{Z}$ is unaffected, the scaling with $c$ changes the probability density functions of all distributions by the same amount through the term $|\mathbf{R}|^{-(p+1)/2}$. That is, for a different scaling of the RCs[20] the overall log-likelihoods of all distributions differ by a factor of $Tp(p+1)/2\log(c)$, which obviously has an impact

20. For example some researches like to scale in terms of annualized volatility in percentage terms, i.e. $c = 252 * 100 * 100$ relative to trading-daily volatility.

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $-\frac{p+1}{2}\sum_{t=1}^{T}\log\lvert\mathbf{R}_t\rvert =$ | -23825 + | -21055 + | 32476 + | 42134 + | 29571 + | |
| Wishart | -78548 | -77402 | -229019 | -189898 | -634572 | |
| Riesz | -66259 | -63468 | -169237 | -143186 | -325345 | |
| iWishart | -53481 | -52655 | -128888 | -99982 | -88294 | |
| iRiesz | -49171 | -46884 | -102347 | -77643 | 32107 | |
| tWishart | -38774 | -35641 | -64340 | -56199 | 14672 | |
| tRiesz | -30551 | -29581 | -36222 | -32247 | 178615 | |
| itWishart | -36261 | -32226 | -41612 | -21886 | 309817 | |
| itRiesz | -29247 | -28467 | -20331 | -4094 | 375739 | |
| F | -52474 | -51592 | -121011 | -93515 | -31855 | |
| FRiesz | -34377 | -30180 | -43343 | -24965 | 287955 | |
| iFRiesz | -39018 | -34944 | -60301 | -39011 | 231719 | |

Table 4: Log-likelihood values for the estimated static distributions and different datasets. The background shades are to be read column-wise, with the lowest log-likelihood value being shaded black and the highest one being shaded white, with a linear scaling in between.

on relative comparisons between log-likelihood and information criterion values. For this reason the part

$$-\frac{p+1}{2}\sum_{t=1}^{T}\log|\mathbf{R}_t| =$$

of the log-likelihoods is displayed separately in tables 4 and 5[21].

It is no surprise that distributions that are nested by others[22] obtain a lower log-likelihood value than the nesting ones. It is however not clear how big the differences are and how non-nested distributions, e.g. the (inverse) F-Riezs and the (inverse) t-Riesz distributions compare.

In table 4[23] we see the log-likelihood values of the estimated static distributions, table 5[24] displays the log-likelihood values of the estimated GAS models for the different

21. A similar term has been separated out for the Bayes information criterion tables in the appendix
22. see again figure 1
23. For the Bayes information criterion see table 13 in the appendix
24. For the Bayes information criterion see table 14 in the appendix

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $-\frac{p+1}{2}\sum_{t=1}^{T}\log\lvert\mathbf{R}_t\rvert =$ | -23825 + | -21055 + | 32476 + | 42134 + | 29571 + | |
| Wishart | -29555 | -25819 | -34377 | -38397 | 140907 | |
| Riesz | -24713 | -22815 | -18507 | -20851 | 245176 | |
| iWishart | -10729 | -10588 | 34724 | 30197 | 608212 | |
| iRiesz | -8850 | -9246 | 39021 | 35790 | 632471 | |
| tWishart | -15604 | -14160 | 4047 | -2067 | 273354 | |
| tRiesz | -13401 | -12252 | 14258 | 9117 | 352243 | |
| itWishart | -6691 | -6162 | 49386 | 44115 | 659591 | |
| itRiesz | -5181 | -4954 | 53406 | 48993 | 680302 | |
| F | -10409 | -10418 | 34904 | 30626 | 611107 | |
| FRiesz | -4778 | -4516 | 52417 | 50165 | 683462 | |
| iFRiesz | -6596 | -6527 | 46209 | 45236 | 663824 | |

Table 5: Log-likelihood values for the estimated GAS models and different datasets. The background shades are to be read column-wise, with the lowest log-likelihood value being shaded black and the highest one being shaded white, with a linear scaling in between. NaN values are caused by prohibitively long computing times.

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $10^{-2}\times$ | | | | | | |
| Wishart | 1.446 | 1.105 | 0.600 | 0.609 | 0.106 | |
| Riesz | 0.740 | 0.938 | 0.346 | 0.315 | 0.079 | |
| iWishart | 0.443 | 0.611 | 0.336 | 0.342 | 0.127 | |
| iRiesz | 0.472 | 0.541 | 0.293 | 0.293 | 0.113 | |
| tWishart | 0.699 | 0.719 | 0.375 | 0.426 | 0.115 | |
| tRiesz | 0.549 | 0.607 | 0.264 | 0.270 | 0.085 | |
| itWishart | 0.594 | 0.639 | 0.351 | 0.371 | 0.138 | |
| itRiesz | 0.517 | 0.557 | 0.314 | 0.320 | 0.122 | |
| F | 0.644 | 0.715 | 0.365 | 0.393 | 0.146 | |
| FRiesz | 0.367 | 0.330 | 0.211 | 0.232 | 0.103 | |
| iFRiesz | 0.501 | 0.459 | 0.272 | 0.261 | 0.113 | |

Table 6: Score parameter 1.

datasets. Note that the Bayes information criterion (BIC) rankings and distances are very close to those of the log-likelihood, since the BIC penalty term for the number of parameters is dominated by the number of parameters in $\mathbf{\Sigma}$, which is common to all distributions and are of order $n^2$.

In the static setting we see a clear pattern in favor of the t-named distributions and the (inverse) F-Riesz distribution. The best fit is achieved by the tail homogeneous inverse t-Riesz distribution for all datasets.

In the dynamic GAS setting the picture is less clear. It is apparent that the mean shifting achieved by time-variation in $\mathbf{\Sigma}_t$ drastically improves the fit as compared to the static setting for all distributions. It stands out that the inverse t-Riesz and F-Riesz distributions achieve the best fit and are very close to each other. The former achieves the best fit for four datasets and the latter for one dataset. Thus, once we have accounted for the time-varying mean, it seems that the difference between "tail homogeneity" and "tail heterogeneity" are slim.

Finally, it should be noted that simple scalar dynamics with one lagged value for $\mathbf{\Sigma}_t$ might seem restrictive. One could, for example, easily extend to HAR-type dynamics (see

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $10^{-2}\times$ | | | | | | |
| Wishart | 0.554 | 0.690 | 0.332 | 0.279 | 0.073 | |
| Riesz | 0.753 | 0.810 | 0.438 | 0.325 | 0.089 | |
| iWishart | 0.534 | 0.518 | 0.250 | 0.229 | 0.068 | |
| iRiesz | 0.644 | 0.577 | 0.275 | 0.253 | 0.077 | |
| tWishart | 5.090 | 4.796 | 4.516 | 4.693 | 4.479 | |
| tRiesz | 4.977 | 4.956 | 4.657 | 4.411 | 4.321 | |
| itWishart | 2.942 | 3.357 | 2.727 | 2.427 | 1.606 | |
| itRiesz | 3.049 | 3.579 | 2.919 | 2.537 | 1.669 | |
| F | 0.605 | 0.604 | 0.262 | 0.249 | 0.076 | |
| FRiesz | 1.080 | 1.175 | 0.472 | 0.449 | 0.127 | |
| iFRiesz | 1.012 | 1.101 | 0.414 | 0.410 | 0.112 | |

Table 7: Score parameter 2.

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| Wishart | 0.9832 | 0.9880 | 0.9926 | 0.9848 | 0.9969 | |
| Riesz | 0.9926 | 0.9882 | 0.9957 | 0.9924 | 0.9980 | |
| iWishart | 0.9969 | 0.9945 | 0.9968 | 0.9939 | 0.9974 | |
| iRiesz | 0.9969 | 0.9956 | 0.9974 | 0.9952 | 0.9980 | |
| tWishart | 0.9946 | 0.9940 | 0.9966 | 0.9915 | 0.9972 | |
| tRiesz | 0.9967 | 0.9950 | 0.9978 | 0.9953 | 0.9982 | |
| itWishart | 0.9956 | 0.9942 | 0.9968 | 0.9932 | 0.9972 | |
| itRiesz | 0.9969 | 0.9958 | 0.9975 | 0.9948 | 0.9980 | |
| F | 0.9955 | 0.9938 | 0.9966 | 0.9931 | 0.9971 | |
| FRiesz | 0.9982 | 0.9984 | 0.9989 | 0.9968 | 0.9987 | |
| iFRiesz | 0.9973 | 0.9977 | 0.9986 | 0.9966 | 0.9984 | |

Table 8: Garch Parameter.

Corsi 2009 and Opschoor et al. 2018) with full or diagonal parameter matrices. Furthermore one could let $\boldsymbol{\theta}$ vary over time by using similar GAS-HAR dynamics. In practice however, these ideas provide only minor improvements. In particular we applied these rich dynamic specifications to the most ill-fitted distribution, i.e. the Wishart distributions, which should benefit most from further extensions to the parameter dynamics. For the first five-dimensional dataset, the shift from estimating a static distribution to our baseline GAS model provides a large increase in the relevant part of the estimated log-likelihood from $-78548$ to $-29555$.[25] The further increase to $-28040$, achieved by employing HAR-type dynamics with diagonal parameter matrices[26], is comparatively low, as is the further increase to $-27639$ achieved by also adding GAS-HAR dynamics to the degree of freedom parameter $n$.

### 5.1.1 Tail Homogeneity vs Tail Heterogeneity

Now we investigate the differences in fit between the inverse t-Riesz and the F-Riesz distribution, i.e. a tail homogeneous versus a tail heterogeneous distribution. Figure 3 shows the difference in log-likelihood contributions between the two distributions depending on the log determinant of the RDs for the first five-dimensional dataset. We clearly see, that the inverse t-Riesz distribution gains its advantage in static fit mainly from the "larger" RCs. This is in line with our expectation that tail heterogeneity is disadvantageous for crises periods. It also fits better for very small RCs, which can be rationalized by the fact that in times of a very calm market the probability that one of the $\chi^2$ distributions on the main diagonal of $\underline{\mathbf{B}}$ lies in the tail (which is disadvantageous for the fit in market wide calm periods) is higher than only the probability that the $\Gamma$ distribution lies in the tail.

As we see in figure 4, as soon as we introduce dynamics in $\boldsymbol{\Sigma}$, the clear advantage of the inverse t-Riesz disappears although there still seems to be a relationship between bigger RCs being better captured by the inverse t-Riesz distribution, especially for the

25. See tables 4 and 5

26. In each parameter matrix, the estimated parameters were very close to each other, which is further motivation for scalar dynamics.

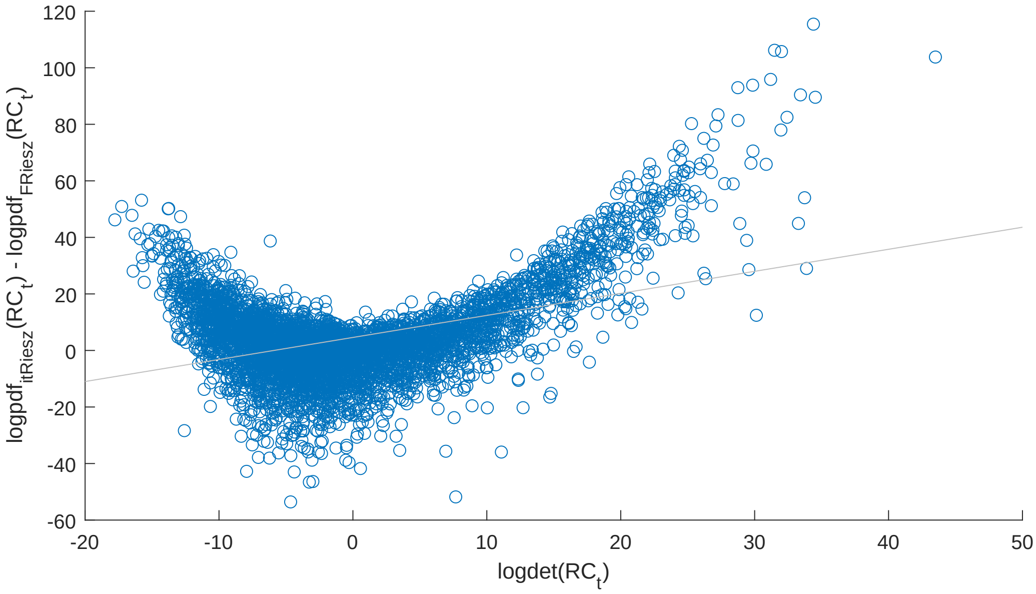



Figure 3: Difference in log likelihood contributions between the fitted static inverse t-Riesz and F-Riesz distributions, depending on the log-determinant of the realized covariance matrices.

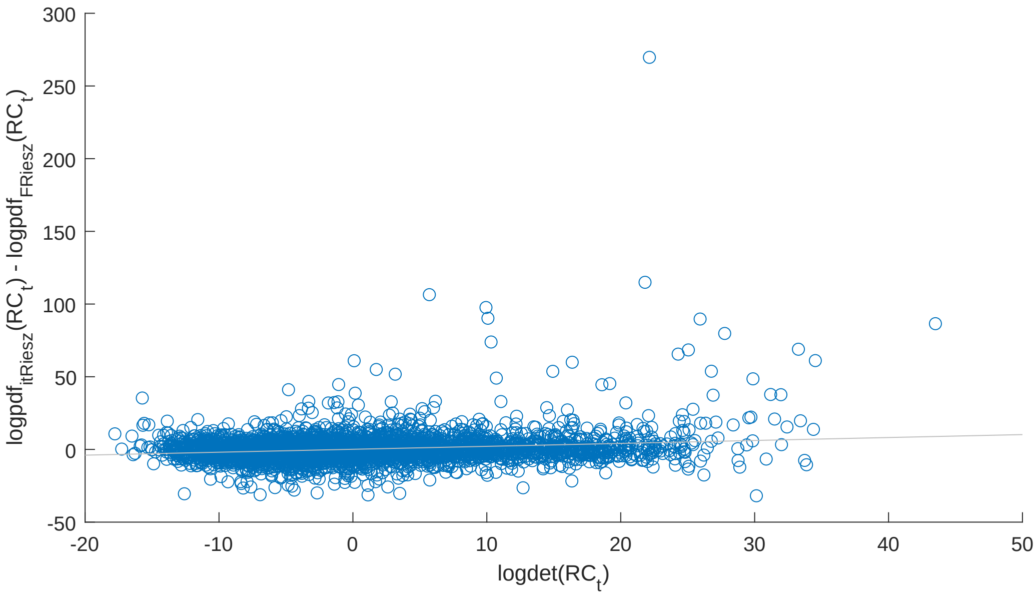



Figure 4: Difference in log likelihood contributions between the fitted GAS models using the inverse t-Riesz and F-Riesz distributions, depending on the log-determinant of the realized covariance matrices.

largest RCs. It seems that mean shifting dynamics of the GAS models do a very good job at .

Finally a look at our estimated mixture model weights in figure **??** for e.g. the first five-dimensional dataset reveals that, when we allow the model to choose the weights between the inverse t-Riesz and the F-Riesz distribution, it tends to put a higher weight on the inverse t-Riesz distribution in times of large RCs, as we expected.

## 5.2 Out-of-Sample Forecasting Ability

We use the 10-dimensional dataset and starting from 2007 re-estimate the GAS models for the different distributions daily with a moving window of 1250 (roughly 5 years) trailing observations, make one-day ahead forecasts and evaluate the forecasting ability with thre different loss functions. For the Riezs-named distributions we do not optimize over the ordering of the assets, but simply take the optimal ordering of the full-sample estimations in table 5. Note that the forecasted $\widehat{\mathbf{\Sigma}}_{t+1}$ is easy to get by simply plugging $\mathbf{R}_t$ and $\widehat{\mathbf{\Sigma}}_t$, which is obtained from the estimation, into equation (12).

The first loss function is the simple mean squared error. The second one is the log-score, where $\mathbf{R}_{t+1}$ is plugged into the time-$t$ forecasted log probability density function,

$$log\, p_d(\mathbf{R}_{t+1}|\widehat{\mathbf{\Sigma}}_{t+1}, \widehat{\boldsymbol{\theta}}; \mathcal{F}_t).$$

For the third loss-function we want to minimize portfolio variance, which is an economically relevant objective as is seen by a quick search through Morningstar listed US equity funds, which yields a total of 14 volatility related funds, see table **??**. For this we take the interpretation that $\widehat{\mathbf{\Sigma}}_{t+1}$ is the predicted RC and we want to minimize the predicted portfolio variance

$$\mathbf{RV}_{t+1}^{pred} = \mathbf{w}_{t+1}^{\top}\, \widehat{\mathbf{\Sigma}}_{t+1}\, \mathbf{w}_{t+1}.$$

| | Mean Squared Error | | | | - Log-Score | | | | GMVP Variances | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Assets: | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 |
| Wishart | 506 | 773 | 853 | 215 | 9.2 | 8.5 | -9.2 | -11.8 | 0.895 | 1.124 | 0.585 | 0.610 |
| Riesz | 516 | 778 | 844 | 240 | 8.0 | 7.9 | -12.4 | -14.6 | 0.903 | 1.133 | 0.598 | 0.627 |
| iWishart | 519 | 791 | 853 | 252 | 5.3 | 5.3 | -23.6 | -24.6 | 0.950 | 1.154 | 0.597 | 0.623 |
| iRiesz | 521 | 797 | 859 | 259 | 4.8 | 5.1 | -24.5 | -25.6 | 0.907 | 1.158 | 0.607 | 0.630 |
| F | 512 | 789 | 843 | 241 | 5.1 | 5.2 | -23.7 | -24.7 | 0.901 | 1.143 | 0.595 | 0.619 |
| FRiesz | 523 | 807 | 896 | 266 | 3.9 | 4.0 | -27.5 | -28.8 | 0.915 | 1.167 | 0.607 | 0.637 |
| iFRiesz | 510 | 794 | 840 | 254 | 4.2 | 4.4 | -26.2 | -28.0 | 0.900 | 1.150 | 0.611 | 0.636 |
| tWishart | 504 | 785 | 823 | 226 | 5.9 | 5.7 | -18.7 | -19.8 | 0.893 | 1.125 | 0.585 | 0.615 |
| tRiesz | 503 | 782 | 822 | 259 | 5.4 | 5.4 | -20.6 | -21.9 | 0.900 | 1.135 | 0.612 | 0.626 |
| itWishart | 498 | 767 | 814 | 219 | 4.2 | 4.2 | -27.1 | -27.9 | 0.893 | 1.134 | 0.591 | 0.616 |
| itRiesz | 500 | 769 | 813 | 224 | 3.8 | 3.9 | -27.9 | -28.7 | 0.894 | 1.137 | 0.593 | 0.623 |

Table 9: Forecasting performance for the entire forecasting window using one-step ahead forecasts, where each model is re-estimated every 10 trading days. 90% model confidence sets in gray.

To do so an investor must set optimal portfolio weights

$$\mathbf{w}^*_{t+1} = \frac{\left(\widehat{\mathbf{\Sigma}}_{t+1}\right)^{-1} \mathbf{1}}{\mathbf{1}^\top \left(\widehat{\mathbf{\Sigma}}_{t+1}\right)^{-1} \mathbf{1}}.$$

The actually realized portfolio variance

$$\left(\mathbf{w}^*_{t+1}\right)^\top \mathbf{R}_{t+1}\, \mathbf{w}^*_{t+1} \tag{23}$$

is then the loss for $t+1$.[27]

We evaluate the predictive ability by constructing model confidence sets as proposed in Hansen, Lunde, and Nason 2011. These sets contain the model with the best predictive ability for a given confidence level.

Table 9 shows the results of this forecasting exercise. First of all it is striking to see that the inverse t-Riesz distribution has the lowest loss for all loss-function-dataset

27. Note that the variance for "oracle" minimum variance portfolio obtains as $\left(\frac{\mathbf{R}_t^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{R}_t^{-1}\mathbf{1}}\right)^\top \mathbf{R}_t \left(\frac{\mathbf{R}_t^{-1}\mathbf{1}}{\mathbf{1}^\top\mathbf{R}_t^{-1}\mathbf{1}}\right) = \left(\mathbf{1}^\top\mathbf{R}_t^{-1}\mathbf{1}\right)^{-1}$.

| | - Log-Score | | | |
|---|---|---|---|---|
| # Assets: | 5 | 5 | 10 | 10 |
| Wishart | 18.0 | 18.0 | 30.7 | 15.6 |
| Riesz | 17.3 | 17.6 | 26.6 | 12.7 |
| iWishart | 15.5 | 15.3 | 15.4 | 2.7 |
| iRiesz | 14.7 | 15.1 | 14.7 | 1.6 |
| F | 15.0 | 15.3 | 15.3 | 2.5 |
| FRiesz | 13.8 | 13.9 | 11.2 | -1.2 |
| iFRiesz | 14.1 | 14.4 | 12.6 | -0.3 |
| tWishart | 15.3 | 15.4 | 19.0 | 6.7 |
| tRiesz | 14.9 | 15.2 | 17.7 | 5.0 |
| itWishart | 13.8 | 14.1 | 11.1 | -1.1 |
| itRiesz | 13.5 | 13.9 | 10.4 | -2.0 |

Table 10: Log-score forecasting performance from 01 January 2007 until 31 December 2010 using one-step ahead forecasts, where each model is reestimated every 10 trading days. 90% model confidence sets in gray.

| | Mean Squared Error | | | | - Log-Score | | | | GMVP Variances | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Assets: | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 | 5 | 5 | 10 | 10 |
| Wishart | 580 | 874 | 994 | 278 | 11.9 | 11.2 | -0.1 | -4.2 | 0.923 | 1.159 | 0.610 | 0.643 |
| Riesz | 577 | 879 | 991 | 293 | 10.1 | 10.6 | -3.7 | -8.8 | 0.926 | 1.175 | 0.619 | 0.646 |
| iWishart | 602 | 899 | 1033 | 339 | 7.1 | 7.4 | -16.7 | -18.0 | 0.966 | 1.171 | 0.612 | 0.641 |
| iRiesz | 599 | 899 | 1031 | 339 | 6.6 | 7.1 | -17.8 | -19.5 | 0.925 | 1.176 | 0.616 | 0.647 |
| F | 599 | 898 | 1028 | 335 | 7.0 | 7.4 | -16.8 | -18.2 | 0.920 | 1.169 | 0.611 | 0.639 |
| FRiesz | 598 | 909 | 1063 | 339 | 5.2 | 5.3 | -23.5 | -25.1 | 0.930 | 1.184 | 0.619 | 0.651 |
| iFRiesz | 592 | 901 | 1028 | 334 | 5.7 | 5.9 | -21.6 | -24.0 | 0.916 | 1.170 | 0.614 | 0.648 |
| tWishart | 594 | 902 | 1036 | 322 | 6.9 | 7.0 | -15.5 | -16.7 | 0.916 | 1.156 | 0.610 | 0.642 |
| tRiesz | 587 | 893 | 1023 | 334 | 6.3 | 6.5 | -18.2 | -19.8 | 0.922 | 1.167 | 0.632 | 0.642 |
| itWishart | 584 | 874 | 984 | 313 | 5.4 | 5.5 | -23.7 | -24.5 | 0.913 | 1.162 | 0.611 | 0.637 |
| itRiesz | 581 | 873 | 982 | 316 | 5.0 | 5.2 | -24.8 | -25.8 | 0.914 | 1.165 | 0.607 | 0.643 |

Table 11: Forecasting performance for the entire forecasting window, where each model is reestimated every 10 trading days, using the resulting one- to ten-step ahead forecasts. 90% model confidence sets in gray.

| | - Log-Score | | | |
|---|---|---|---|---|
| # Assets: | 5 | 5 | 10 | 10 |
| Wishart | 20.1 | 20.3 | 40.0 | 21.2 |
| Riesz | 19.1 | 20.2 | 37.5 | 17.2 |
| iWishart | 17.5 | 17.4 | 22.4 | 7.8 |
| iRiesz | 16.6 | 17.2 | 21.5 | 6.5 |
| F | 16.9 | 17.3 | 22.2 | 7.6 |
| FRiesz | 15.2 | 15.2 | 15.7 | 2.2 |
| iFRiesz | 15.7 | 15.8 | 17.6 | 3.4 |
| tWishart | 16.1 | 16.3 | 21.4 | 8.3 |
| tRiesz | 15.6 | 16.0 | 19.9 | 6.4 |
| itWishart | 14.9 | 15.2 | 14.0 | 1.1 |
| itRiesz | 14.5 | 15.1 | 13.2 | 0.1 |

Table 12: Log-score forecasting performance from 01 January 2007 until 31 December 2010, where each model is reestimated every 10 trading days, using the resulting one- to ten-step ahead forecasts. 90% model confidence sets in gray.

combinations except one. Furthermore, it is contained in the 90% model confidence set in all loss-function-dataset combinations except two. Finally, if we restrict the forecasting sample to the volatile market window from 01 January 2007 to 31 December 2010, then table 10 reveals that it is clearly preferred in terms of forecasting performance to the benchmark model, the F-Riesz distribution.

# 6 Conclusion

In this paper all probability distributions so far applied to time series of RCs in the literature are compared and it is shown how they are related to each other. A novel family of probability distribution, which has an intuitive property called "tail homogeneity", is derived and added to the comparison. Generalized autoregressive score (GAS) models are derived for all distributions. The empirical application shows a similar fit of some of the novel distributions to the so-far best competitor distribution, the F-Riesz. The finding of Blasques et al. 2021 that there is "strong heterogeneity of tail behavior of realized covariance matrices" cannot be confirmed, as the "tail-homogeneous" inverse t-Riesz distribution has similar fit and forecasting performance to the F-Riesz distribution.

In a portfolio risk minimizing the novel distribution family performs significantly better than all other distributions.

## References


**(https://mathoverflow.net/users/2570/steven-pav), Steven Pav.** *The derivative of the Cholesky Factor.* MathOverflow. https://mathoverflow.net/q/167719.

**Andersen, Torben G, Tim Bollerslev, Peter F Christoffersen, and Francis X Diebold.** 2006. "Chapter 15 Volatility and Correlation Forecasting," edited by G Elliott, C W J Granger, and A B T - Handbook of Economic Forecasting Timmermann, 1:777–878. Elsevier. ISBN: 1574-0706. https://doi.org/https://doi.org/10.1016/S1574-0706(05)01015-3. https://www.sciencedirect.com/science/article/pii/S1574070605010153.

**Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens.** 2001. "The distribution of realized stock return volatility." *Journal of Financial Economics* 61 (1): 43–76. ISSN: 0304405X. https://doi.org/10.1016/S0304-405X(01)00055-1.

**Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys.** 2003. "Modeling and forecasting realized volatility." *Econometrica* 71 (2): 579–625.

**Asai, Manabu, and Mike K. P. So.** 2013. "Stochastic Covariance Models." *Journal of the Japan Statistical Society* 43 (2): 127–162. https://doi.org/10.2139/ssrn.1673764.

**Blasques, F, Christian Francq, and Sébastien Laurent.** 2022. "Quasi score-driven models." *Journal of Econometrics,* ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2021.12.005. https://www.sciencedirect.com/science/article/pii/S030440762200001X.

**Blasques, Francisco, André Lucas, Anne Opschoor, and Luca Rossini.** 2021. "Tail Heterogeneity for Dynamic Covariance-Matrix-Valued Random Variables : the F-Riesz Distribution."

**Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg.** 2018. "Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions." *Journal of Econometrics* 207 (1): 71–91. ISSN: 18726895. https://doi.org/10.1016/j.jeconom.2018.05.004. https://doi.org/10.1016/j.jeconom.2018.05.004.

**Catania, Leopoldo.** 2021. "Dynamic Adaptive Mixture Models with an Application to Volatility and Risk." *Journal of Financial Econometrics* 19, no. 4 (December): 531–564. ISSN: 1479-8409. https://doi.org/10.1093/jjfinec/nbz018. https://doi.org/10.1093/jjfinec/nbz018.

**Chiriac, Roxana, and Valeri Voev.** 2011. "Modelling and forecasting multivariate realized volatility." *Journal of Applied Econometrics* 26 (6): 922–947. https://doi.org/10.1002/jae.1152. https://doi.org/10.1002/jae.1152.

**Corsi, Fulvio.** 2009. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7, no. 2 (March): 174–196. ISSN: 1479-8409. https://doi.org/10.1093/jjfinec/nbp001. https://doi.org/10.1093/jjfinec/nbp001.

**Creal, Drew, Siem J. Koopman, and André Lucas.** 2013. "Generalized autoregressive score models with applications." *Journal of Applied Econometrics* 28, no. 5 (August): 777–795. ISSN: 0883-7252. https://doi.org/10.1002/jae.1279. https://doi.org/10.1002/jae.1279.

**Creal, Drew, Siem Jan Koopman, and André Lucas.** 2011. "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations." *Journal of Business and Economic Statistics* 29 (4): 552–563. ISSN: 07350015. https://doi.org/10.1198/jbes.2011.10070. https://doi.org/10.1198/jbes.2011.10070.

**Díaz-García, José A.** 2013. "A note on the moments of the Riesz distribution." *Journal of Statistical Planning and Inference* 143 (11): 1880–1886. ISSN: 03783758. https://doi.org/10.1016/j.jspi.2013.07.012. http://dx.doi.org/10.1016/j.jspi.2013.07.012.

———. 2014. "On Riesz distribution." *Metrika* 77 (4): 469–481. ISSN: 1435926X. https://doi.org/10.1007/s00184-013-0449-5.

———. 2016. "A kotz-Riesz-type distribution." *Journal of the Iranian Statistical Society* 15 (1): 71–90. ISSN: 17264057. https://doi.org/10.7508/jirss.2016.01.005.

**NIST Digital Library of Mathematical Functions.** http://dlmf.nist.gov/, Release 1.1.4 of 2022-01-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. http://dlmf.nist.gov/.

**Faraut, Jacques, and Adam Korányi.** 1994. *Analysis on Symmetric Cones.* 382. Oxford: Clarendon Press. ISBN: 0-19-853477-9.

**Golosnoy, Vasyl, Bastian Gribisch, and Roman Liesenfeld.** 2012. "The conditional autoregressive Wishart model for multivariate stock market volatility." *Journal of Econometrics* 167 (1): 211–223. ISSN: 03044076. https://doi.org/10.1016/j.jeconom.2011.11.004. http://dx.doi.org/10.1016/j.jeconom.2011.11.004.

**Gorgi, P., Peter R. Hansen, Pawel Janus, and Siem J. Koopman.** 2019. "Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility model." *Journal of Financial Econometrics* 17 (1): 1–32. ISSN: 14798417. https://doi.org/10.1093/jjfinec/nby007.

**Gourieroux, C., J. Jasiak, and R. Sufana.** 2009. "The Wishart autoregressive process of multivariate stochastic volatility." *Journal of Econometrics* 150 (2): 167–181. ISSN: 03044076. https://doi.org/10.1016/j.jeconom.2008.12.016. http://dx.doi.org/10.1016/j.jeconom.2008.12.016.

**Gribisch, Bastian, and Jan Patrick Hartkopf.** 2022. "Modeling realized covariance measures with heterogeneous liquidity: A generalized matrix-variate Wishart state-space model." *Journal of Econometrics,* ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.01.007. https://www.sciencedirect.com/science/article/pii/S0304407622000392.

**Gribisch, Bastian, and Michael Stollenwerk.** 2020. "Dynamic principal component CAW models for high-dimensional realized covariance matrices." *Quantitative Finance* 20 (5): 799–821. ISSN: 14697696. https://doi.org/10.1080/14697688.2019.1701197.

**Gupta, Arjun K., and D K Nagar.** 2000. *Matrix variate distributions.* Chapman / Hall/CRC, May. ISBN: 9780203749289. https://doi.org/10.1201/9780203749289. https://www.taylorfrancis.com/books/9780203749289.

**Hansen, Peter R., Asger Lunde, and James M. Nason.** 2011. "The model confidence set." *Econometrica* 79, no. 2 (March): 453–497. ISSN: 0012-9682. https://doi.org/10.3982/ECTA5771. https://doi.org/10.3982/ECTA5771.

**Harville, David A.** 1997. *Matrix Algebra From a Statisticians Perspective.* ISBN: 038794978X. https://doi.org/10.1007/0-387-22677-x_20.

**Jondeau, Eric, and Michael Rockinger.** 2006. "Optimal portfolio allocation under higher moments." *European Financial Management* 12, no. 1 (January): 29–55. ISSN: 1354-7798. https://doi.org/10.1111/j.1354-7798.2006.00309.x. https://doi.org/10.1111/j.1354-7798.2006.00309.x.

**Kessentini, Sameh, Mariem Tounsi, and Raoudha Zine.** 2020. "The Riesz probability distribution: Generation and EM algorithm." *Communications in Statistics - Simulation and Computation* 49, no. 8 (August): 2114–2133. ISSN: 0361-0918. https://doi.org/10.1080/03610918.2018.1513139. https://doi.org/10.1080/03610918.2018.1513139.

**Kollo, Tõnu, and Dietrich von Rosen.** 2005. *Advanced Multivariate Statistics with Matrices.* Edited by M. Hazewinkel. Springer. ISBN: 9781402034183.

**Lunde, Asger, Neil Shephard, and Kevin Sheppard.** 2016. "Econometric Analysis of Vast Covariance Matrices Using Composite Realized Kernels and Their Application to Portfolio Choice." *Journal of Business & Economic Statistics* 34, no. 4 (October): 504–518. ISSN: 0735-0015. https://doi.org/10.1080/07350015.2015.1064432. https://doi.org/10.1080/07350015.2015.1064432.

**Lütkepohl, Helmut.** 1989. "A note on the asymptotic distribution of impulse response functions of estimated var models with orthogonal residuals." *Journal of Econometrics* 42 (3): 371–376. ISSN: 03044076. https://doi.org/10.1016/0304-4076(89)90059-6.

———. 2005. *New introduction to multiple time series analysis,* 1–764. ISBN: 3540401725. https://doi.org/10.1007/978-3-540-27752-1.

**Maaß, Hans.** 1971. *Siegel's Modular Forms and Dirichlet Series.* 1st ed. Springer. https://doi.org/10.1007/BFb0058625. https://doi.org/10.1007/BFb0058625.

**Magnus, Jan R.** 2010. "On the concept of matrix derivative." *Journal of Multivariate Analysis* 101 (9): 2200–2206. ISSN: 0047259X. https://doi.org/10.1016/j.jmva.2010.05.005. http://dx.doi.org/10.1016/j.jmva.2010.05.005.

**Magnus, Jan R., and Heinz Neudecker.** 1980. "The Elimination Matrix: Some Lemmas and Applications." *SIAM Journal on Algebraic Discrete Methods* 1 (4): 422–449. ISSN: 0196-5212. https://doi.org/10.1137/0601049.

———. 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons, Ltd.

———. 2019. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons, Ltd.

**McAleer, Michael, and Marcelo C Medeiros.** 2008. "Realized Volatility: A Review." *Econometric Reviews* 27, nos. 1-3 (February): 10–45. ISSN: 0747-4938. https://doi.org/10.1080/07474930701853509. https://doi.org/10.1080/07474930701853509.

**Muirhead, Robb J.** 1982. *Aspects of Multivariate Statistical Theory.* 673. Wiley, Ney Jersey.

**Murray, Ian.** 2016. "Differentiation of the Cholesky Decomposition." https://doi.org/10.48550/arXiv.1602.07527.

**Opschoor, Anne, Pawel Janus, André Lucas, and Dick Van Dijk.** 2018. "New HEAVY models for fat-tailed realized covariances and returns." *Journal of Business & Economic Statistics* 36, no. 4 (October): 643–657. ISSN: 0735-0015. https://doi.org/10.1080/07350015.2016.1245622. https://doi.org/10.1080/07350015.2016.1245622.

**Sutradhar, Brajendra C., and Mir M. Ali.** 1989. "A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model." *Journal of Multivariate Analysis* 29 (1): 155–162. ISSN: 10957243. https://doi.org/10.1016/0047-259X(89)90082-1.

**Thom, Herbert C. S.** 1958. "A Note on the Gamma Distribution." *Monthly Weather Review* 86 (4): 117–122.

**Yu, Philip L. H., Wai Keung Li, and F. C. Ng.** 2017. "The generalized conditional autoregressive Wishart model for multivariate realized volatility." *Journal of Business & Economic Statistics* 35, no. 4 (October): 513–527. ISSN: 0735-0015. https://doi.org/10.1080/07350015.2015.1096788. https://doi.org/10.1080/07350015.2015.1096788.

**Zhang, Lan, Per A Mykland, and Yacine Aït-Sahalia.** 2005. "A Tale of Two Time Scales." *Journal of the American Statistical Association* 100, no. 472 (December): 1394–1411. ISSN: 0162-1459. https://doi.org/10.1198/016214505000000169. https://doi.org/10.1198/016214505000000169.

**Zhou, Jiayuan, Feiyu Jiang, Ke Zhu, and Wai Keung Li.** 2019. *Time series models for realized covariance matrices based on the matrix-F distribution.* 17306818. ISBN: 2018111590. arXiv: 1903.12077. http://arxiv.org/abs/1903.12077.

# 7 Appendix

## 7.1 Tables and Figures

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $(p+1)\sum_{t=1}^{T}\log\lvert\mathbf{R}_t\rvert =$ | 47651 + | 42109 + | -64952 + | -84267 + | -59142 + | |
| Wishart | 157232 | 154940 | 458512 | 380271 | 1271907 | |
| Riesz | 132687 | 127106 | 339025 | 286922 | 653658 | |
| iWishart | 107098 | 105446 | 258251 | 200440 | 179353 | |
| iRiesz | 98512 | 93938 | 205245 | 155837 | -61247 | |
| tWishart | 77692 | 71426 | 129163 | 112881 | -26572 | |
| tRiesz | 61280 | 59340 | 73004 | 65053 | -354255 | |
| itWishart | 72666 | 64596 | 83708 | 44256 | -616862 | |
| itRiesz | 58673 | 57112 | 41222 | 8748 | -748501 | |
| F | 105093 | 103327 | 242505 | 187513 | 66481 | |
| FRiesz | 68965 | 60573 | 87322 | 50565 | -572730 | |
| iFRiesz | 78248 | 70101 | 121237 | 78658 | -460259 | |

Table 13: BIC values for the estimated static distributions and different datasets. The background shades are to be read column-wise, with the lowest BIC value being shaded white and the highest one being shaded black, with a linear scaling in between.

| # Assets: | 5 | 5 | 10 | 10 | 25 | 25 |
|---|---|---|---|---|---|---|
| $(p+1)\sum_{t=1}^{T} \log \lvert \mathbf{R}_t \rvert =$ | 47651 + | 42109 + | -64952 + | -84267 + | -59142 + | |
| Wishart | 59306 | 51832 | 69288 | 77329 | -278990 | |
| Riesz | 49689 | 45892 | 37701 | 42390 | -487121 | |
| iWishart | 21654 | 21370 | -68914 | -59861 | -1213602 | |
| iRiesz | 17963 | 18755 | -77355 | -70893 | -1261711 | |
| tWishart | 31420 | 28532 | -7543 | 4684 | -543867 | |
| tRiesz | 27082 | 24783 | -27813 | -17530 | -701239 | |
| itWishart | 13594 | 12536 | -98221 | -87678 | -1316341 | |
| itRiesz | 10641 | 10188 | -106109 | -97283 | -1357356 | |
| F | 21029 | 21048 | -69257 | -60700 | -1219373 | |
| FRiesz | 9904 | 9379 | -103977 | -99473 | -1363270 | |
| iFRiesz | 13539 | 13401 | -91562 | -89615 | -1323993 | |

Table 14: BIC values for the estimated GAS models and different datasets. The background shades are to be read column-wise, with the lowest BIC value being shaded white and the highest one being shaded black, with a linear scaling in between. NaN values are caused by prohibitively long computing times.

## 7.2 Proofs

Throughout the paper we exclusively consider real numbers. $p$ is the cross-sectional dimension, i.e the number of assets, which we index by $i = 1, \ldots, p$. If not otherwise specified, matrices are $p \times p$ and vectors are $p \times 1$.

$\mathbf{R}$ is a symmetric positive definite random matrix or a realization thereof.

$\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are symmetric positive (semi-)definite matrices.

$\mathbf{C}$ denotes the lower Cholesky factor of $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top$.

$\mathbf{n}$ and $\boldsymbol{\nu}$ are column vectors.

$\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ generic matrices of any dimension.

$\otimes$ denotes the Kronecker product.

$\odot$ denotes element-wise multiplication.

**Proof of Theorem 3.1**

*Proof.* Note that $\mathcal{I}$ of all Wishart-based have the form

$$\mathcal{I} = \mathbf{G}^\top \left( \alpha_{\boldsymbol{\theta}} \left( \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) + c_{\boldsymbol{\theta}} \operatorname{vec}\left(\boldsymbol{\Sigma}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top \right) \mathbf{G}, \tag{24}$$

where the scalars $\alpha_{\boldsymbol{\theta}}$ and $c_{\boldsymbol{\theta}}$ only depend on the degree of freedom parameter(s) of the respective distribution. Thus, using Theorem 7.7 we have

$$\begin{aligned}
\mathcal{I}^{-1} &= \alpha_{\boldsymbol{\theta}} \left( \mathbf{G}^\top \left[ \left( \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) + \alpha_{\boldsymbol{\theta}}^{-1} c_{\boldsymbol{\theta}} \operatorname{vec}\left(\boldsymbol{\Sigma}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top \right] \mathbf{G} \right)^{-1} &(25)\\
&= \alpha_{\boldsymbol{\theta}} \mathbf{G}^+ \left( \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} + \frac{\alpha_{\boldsymbol{\theta}}^{-1} c_{\boldsymbol{\theta}}}{1 + \alpha_{\boldsymbol{\theta}}^{-1} c_{\boldsymbol{\theta}} p} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top \right) \mathbf{G}^+ &(26)\\
&= \mathbf{G}^+ \left( \alpha_{\boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} + \frac{\alpha_{\boldsymbol{\theta}} c_{\boldsymbol{\theta}}}{\alpha_{\boldsymbol{\theta}} + c_{\boldsymbol{\theta}} p} \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top \right) \mathbf{G}^+. &(27)
\end{aligned}$$

Define $\beta_{\boldsymbol{\theta}} = \frac{\alpha_{\boldsymbol{\theta}} c_{\boldsymbol{\theta}}}{\alpha_{\boldsymbol{\theta}} + c_{\boldsymbol{\theta}} p}$,

$$\begin{aligned}
&\mathbf{G}^+ \left( \alpha_{\boldsymbol{\theta}} \left( \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) + \beta_{\boldsymbol{\theta}} \operatorname{vec}\left(\boldsymbol{\Sigma}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top \right) \left(\mathbf{G}^+\right)^\top \mathbf{G}^\top \operatorname{vec}\left(\triangle\right) &(28)\\
&\overset{(73)}{=} \mathbf{G}^+ \left( \alpha_{\boldsymbol{\theta}} \left( \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) + \beta_{\boldsymbol{\theta}} \operatorname{vec}\left(\boldsymbol{\Sigma}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top \right) \operatorname{vec}\left(\triangle\right) &(29)\\
&= \alpha_{\boldsymbol{\theta}} \mathbf{G}^+ \left( \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \right) \operatorname{vec}\left(\triangle\right) + \beta_{\boldsymbol{\theta}} \mathbf{G}^+ \operatorname{vec}\left(\boldsymbol{\Sigma}\right) \operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top \operatorname{vec}\left(\triangle\right) &(30)\\
&= \alpha_{\boldsymbol{\theta}} \mathbf{G}^+ \operatorname{vec}\left(\boldsymbol{\Sigma} \triangle \boldsymbol{\Sigma}\right) + \beta_{\boldsymbol{\theta}} \operatorname{tr}\left(\boldsymbol{\Sigma} \triangle\right) \mathbf{G}^+ \operatorname{vec}\left(\boldsymbol{\Sigma}\right) &(31)\\
&= \frac{\alpha_{\boldsymbol{\theta}}}{2} \operatorname{vech}\left( \boldsymbol{\Sigma}(\triangle + \triangle^\top)\boldsymbol{\Sigma} \right) + \beta_{\boldsymbol{\theta}} \operatorname{tr}\left(\boldsymbol{\Sigma} \triangle\right) \operatorname{vech}\left(\boldsymbol{\Sigma}\right) &(32)\\
&= \operatorname{vech}\left( \frac{\alpha_{\boldsymbol{\theta}}}{2} \boldsymbol{\Sigma}(\triangle + \triangle^\top)\boldsymbol{\Sigma} + \beta_{\boldsymbol{\theta}} \operatorname{tr}\left(\boldsymbol{\Sigma} \triangle\right) \boldsymbol{\Sigma} \right). &(33)
\end{aligned}$$

Now simply apply the *ivech* operator. $\square$

## 7.3 Special Functions

**Definition 7.1.** *Let* $\mathbf{X}$ *be a real* $p \times p$ *matrix and let* $\mathbf{X}_{[i]}$ *denote the square submatrix created by taking the first i rows and columns of* $\mathbf{X}$*. Then the generalized power function,* *denoted by* $|\mathbf{X}|_{\mathbf{n}}$ *is defined as*

$$|\mathbf{X}|_{\mathbf{n}} = |\mathbf{X}_{[1]}|^{n_1-n_2}|\mathbf{X}_{[2]}|^{n_2-n_3}\dots|\mathbf{X}_{[p-1]}|^{n_{p-1}-n_p}|\mathbf{X}|^{n_p}. \tag{34}$$

*Note: The determinant with subscript notation was introduced by Blasques et al. 2021 to make immediately visible the close relation to the determinant raised to the power* $n$*, as it is easily seen that for* $n_1 = n_2 = \dots = n_p = n$ *we have* $|\mathbf{X}|_{\mathbf{n}} = |\mathbf{X}|^n$*. They name* $|\mathbf{X}|_{\mathbf{n}}$ *power weighted determinant. It is also known as highest weight vector.*

**Definition 7.2.** *Let* $\mathbf{X}$ *be a* $p \times p$ *matrix, then the function* $\Phi(\mathbf{X})$ *takes the lower-triangular part of a matrix and halves its diagonal,*

$$\Phi_{ij}(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij} \text{ for } i > j, \\ \frac{1}{2}\mathbf{X}_{ii} \text{ for } i = j \text{ and} \\ 0 \text{ for } i < j. \end{cases} \tag{35}$$

**Lemma 7.1.** *Let* $\mathbf{\Sigma} = \mathbf{TDT}^\top$ *be the unique decomposition into lower triangular square matrix with ones on the main diagonal,* $\mathbf{T}$ *and diagonal matrix with positive entries on the diagonal* $\mathbf{D}$*. Then we can rewrite*

$$|\mathbf{\Sigma}|_{\mathbf{n}} = \prod_{i=1}^{p}\mathbf{D}_{ii}^{n_i} = \prod_{i=1}^{p}\mathbf{C}_{ii}^{2n_i}. \tag{36}$$

*Proof.* The equivalence between the two different representation is proofed in Maaß 1971, pp. 69-70. This proof is closely based on it. If

$$\mathbf{\Sigma} = \mathbf{TDT}^\top = \mathbf{CC}^\top, \tag{37}$$

then

$$\mathbf{\Sigma}_{[j]} = \mathbf{C}_{[j]}\mathbf{C}_{[j]}^\top = \mathbf{T}_{[j]}\mathbf{D}_{[j]}\mathbf{T}_{[j]}^\top \tag{38}$$

. So

$$|\mathbf{\Sigma}_{[j]}| = \prod_{i=1}^{j} \mathbf{D}_{ii} \tag{39}$$

and thus

$$|\mathbf{\Sigma}_{[1]}| = \mathbf{D}_{11} \text{ and for } j > 1 \text{ we have } |\mathbf{\Sigma}_{[j]}|/|\mathbf{\Sigma}_{[j-1]}| = \mathbf{D}_{jj}. \tag{40}$$

Finally

$$\prod_{i=1}^{p} \mathbf{D}_{ii}^{s_i} = |\mathbf{\Sigma}_{[1]}|^{s_1} \prod_{i=2}^{p} (|\mathbf{\Sigma}_{[i]}|/|\mathbf{\Sigma}_{[i-1]}|)^{s_i} = |\mathbf{\Sigma}_{[1]}|^{s_1 - s_2} |\mathbf{\Sigma}_{[2]}|^{s_2 - s_3} \dots |\mathbf{\Sigma}_{[p]}|^{s_p}. \tag{41}$$

□

**Lemma 7.2.**

$$\left|\mathbf{C}\mathrm{dg}(\mathbf{n})\mathbf{C}^\top\right|_{\boldsymbol{\nu}} = \prod_{i=1}^{p} n_i^{\nu_i} \, |\mathbf{\Sigma}|_{\boldsymbol{\nu}}\,. \tag{42}$$

*Proof.*

$$\begin{aligned}
\left|\mathbf{C}\mathrm{dg}(\mathbf{n})\mathbf{C}^\top\right|_{\boldsymbol{\nu}} &= \prod_{i=1}^{p} (\mathbf{C}_{ii}\sqrt{n_i})^{2\nu_i} &(43)\\
&= \prod_{i=1}^{p} \mathbf{C}_{ii}^{2\nu_i} \prod_{i=1}^{p} n_i^{\nu_i} &(44)\\
&= \prod_{i=1}^{p} n_i^{\nu_i} \left|\mathbf{C}\mathbf{C}^\top\right|_{\boldsymbol{\nu}}. &(45)
\end{aligned}$$

□

**Lemma 7.3.**

$$\frac{\partial \log |\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathbf{n}^\top} = \frac{\partial \log(\prod_{i=1}^{p} \mathbf{C}_{ii}^{2n_i})}{\partial \mathbf{n}^\top} = \frac{\partial \sum_{i=1}^{p} 2n_i \log(\mathbf{C}_{ii})}{\partial \mathbf{n}^\top} = 2 \log \mathrm{vecd}(\mathbf{C})^\top \tag{46}$$

**Lemma 7.4.** *For* $\mathbf{n}$ *with* $n_i > i - 1$ *we have,*

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = 2^{p\bar{n}/2}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)|\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}} \tag{47}$$

*and for* $n_i < i - p$ *we have,*

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}^{-1}|_{\frac{\mathbf{n}+p+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = \frac{1}{2^{p\bar{n}/2}}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right)|\mathbf{B}|_{\frac{\mathbf{n}}{2}}\,. \tag{48}$$

*Proof.* The proofs can be found in Faraut and Korányi 1994 Chapter VII.[28] Throughout, according to their table on p. 97, for the cone of symmetric positive definite matrices we have the dimension $n = p(p+1)/2$, the rank $r = p$ and $d = 1$.[29] Furthermore, throughout their book they use the Euclidean measure on a Euclidean space, which translated into our notation is $dx = \prod_{i=1}^{p} a_{ii} 2^{p(p-1)/4} \prod_{i<j} a_{ij} = 2^{p(p-1)/4} d\mathbf{A}$. Their use of the Euclidean measure leads to a slightly different multivariate gamma function. In particular from their Theorem VII.1.1.

$$\Gamma_\Omega(\mathbf{n}) = 2^{p(p-1)/4}\Gamma_p(\mathbf{n}), \tag{49}$$

with $\Gamma_p(\mathbf{n})$ as in (35.3.5) of the NIST Digital Library of Mathematical Functions.

Their Proposition VII.1.2., with $x = \mathbf{A}$, $y = \frac{1}{2}\mathbf{B}$ and $\mathbf{s} = \frac{\mathbf{n}}{2}$ translates to

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) 2^{p(p-1)/4} d\mathbf{A} = 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)|2\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}} \tag{50}$$

$$= 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)2^{p\bar{n}/2}|\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}}. \tag{51}$$

Their last equation on page 129 together with Proposition VII.1.5 (ii) and $x = \mathbf{A}$,

---

28. Further references are Díaz-García 2014, Maaß 1971 p. 76, Gupta and Nagar 2000, Theorem 1.4.7., which is based on Olkin, I. (1959). A class of integral identities with matrix argument. Duke Mathematical Journal, 26(2), 207–213. doi:10.1215/s0012-7094-59-02621-3, which in turn is based on the generalized Ingham formula in Bellman, R. (1956) (doi:10.1215/s0012-7094-56-02356-0).

29. For the notation see their Example 2 on p. 8 and p. 9.

$y = \frac{1}{2}\mathbf{B}$ and $\mathbf{s} = \frac{\mathbf{n}}{2}$ translates to

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}^{-1}|_{\frac{\mathbf{n}+p+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) 2^{p(p-1)/4} d\mathbf{A} = 2^{p(p-1)/4}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right)\left|\frac{1}{2}\mathbf{B}\right|_{\frac{\mathbf{n}}{2}} \tag{52}$$

$$= 2^{p(p-1)/4}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right)\frac{1}{2^{p\bar{n}/2}}|\mathbf{B}|_{\frac{\mathbf{n}}{2}}. \tag{53}$$

□

**Lemma 7.5.** *Let the upper generalized multivariate gamma function be defined as in Blasques et al. 2021 and denote a vector with its elements in reverse order by a superscript left arrow, e.g.* $\overleftarrow{\mathbf{n}} = (n_p, n_{p-1}, \ldots, n_1)^\top$ *, then*

$$\Gamma_p\left(\overleftarrow{\mathbf{n}}\right) = \bar{\Gamma}_U(\mathbf{n}). \tag{54}$$

*Proof.* ToDO. See Faraut and Korányi 1994. □

**Lemma 7.6.**

$$\frac{\partial \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n}^\top} = \frac{\partial \log\left(\pi^{p(p-1)/4}\prod_{j=1}^p \Gamma\left(n_j - \frac{1}{2}(j-1)\right)\right)}{\partial \mathbf{n}^\top} \tag{55}$$

$$= \frac{\sum_{j=1}^p \partial \log\left(\Gamma\left(n_j - \frac{1}{2}(j-1)\right)\right)}{\partial \mathbf{n}^\top} \tag{56}$$

$$= \left[\psi(n_1), \psi\left(n_2 - \frac{1}{2}\right), \ldots, \psi\left(n_p - \frac{1}{2}(p-1)\right)\right] \tag{57}$$

$$= \boldsymbol{\psi}(\mathbf{n})^\top, \tag{58}$$

*with* $\boldsymbol{\psi}_i(\mathbf{n}) = \psi\left(n_i - \frac{1}{2}(i-1)\right)$ *and*

$$\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n}\partial \mathbf{n}^\top} = \operatorname{dg}\left(\left[\psi'(n_1), \psi'\left(n_2 - \frac{1}{2}\right), \ldots, \psi'\left(n_p - \frac{1}{2}(p-1)\right)\right]\right) \tag{59}$$

$$= \operatorname{dg}\left(\boldsymbol{\psi}'(\mathbf{n})\right), \tag{60}$$

*with* $\boldsymbol{\psi}'_i(\mathbf{n}) = \psi'\left(n_i - \frac{1}{2}(i-1)\right)$. *Using equation (5.2.2) of NIST Digital Library of Mathematical Functions.*

## 7.4 Matrix Relations

For matrices $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ with appropriate dimensions we have (Magnus and Neudecker 2019, p.12, p. 35)

$$\operatorname{vec}(\mathbf{XYZ}) = \left(\mathbf{Z}^\top \otimes \mathbf{X}\right) \operatorname{vec}(\mathbf{Y}), \tag{61}$$

$$\operatorname{tr}(\mathbf{XYZ}) = \operatorname{tr}(\mathbf{YZX}) = \operatorname{tr}(\mathbf{ZXY}), \tag{62}$$

$$\operatorname{tr}(\mathbf{X}^\top \mathbf{Y}) = \operatorname{vec}(\mathbf{X})^\top \operatorname{vec}(\mathbf{Y}) \text{ and} \tag{63}$$

$$\operatorname{tr}(\mathbf{WXYZ}) = \operatorname{vec}(\mathbf{W})^\top \operatorname{vec}(\mathbf{XYZ}) = \operatorname{vec}(\mathbf{W})^\top (\mathbf{X} \otimes \mathbf{Z}) \operatorname{vec}(\mathbf{Y}), \tag{64}$$

where for the last equality we used (63) and (61).

### 7.4.1 Duplication, Elimination and Commutation Matrices

As a reference see Lütkepohl 2005, A.12.2. $\mathbf{G}_p$ denotes the *duplication matrix* defined by

$$\operatorname{vec}(\mathbf{X}) = \mathbf{G}_p \operatorname{vech}(\mathbf{X}), \tag{65}$$

where $\mathbf{X}$ is an arbitrary **symmetric** $p \times p$ matrix.

For symmetric $\mathbf{X}$ the *duplication matrix* $\mathbf{G}$ is unique, however the so called *elimination matrix*, which converts $\operatorname{vec}(\mathbf{X})$ to $\operatorname{vech}(\mathbf{X})$ is not unique (since for every lower-diagonal element of $\mathbf{X}$ we can take a fraction $c$ of the corresponding upper- and a fraction $1-c$ of the lower-diagonal element of $\mathbf{X}$). One possible choice is the Moore-Penrose inverse of $\mathbf{G}_p$,

$$\mathbf{G}_p^+ = \left(\mathbf{G}_p^\top \mathbf{G}_p\right)^{-1} \mathbf{G}_p^\top, \tag{66}$$

for which obviously

$$\mathbf{G}_p^+ \operatorname{vec}(\mathbf{X}) = \mathbf{G}_p^+ \mathbf{G}_p \operatorname{vech}(\mathbf{X}) = \operatorname{vech}(\mathbf{X}). \tag{67}$$

Another possible choice is the canonical elimination matrix $\mathbf{F}_p$ which sets the aforementioned fraction $c = 0$.

For **lower-triangular** $p \times p$ matrix $\mathbf{Y}$ Magnus and Neudecker 1980 note (Lemma 3.3 (i)) the unique *elimination* and *duplication matrices* are given by

$$\operatorname{vec}(\mathbf{Y}) = \mathbf{F}_p^\top \operatorname{vech}(\mathbf{Y}) \tag{68}$$

and

$$\operatorname{vech}(\mathbf{Y}) = \mathbf{F}_p \operatorname{vec}(\mathbf{Y}). \tag{69}$$

$\mathbf{K}_{pq}$ denotes the *commutation matrix* defined by

$$\operatorname{vec}\left(\mathbf{Z}^\top\right) = \mathbf{K}_{pq} \operatorname{vec}(\mathbf{Z}), \tag{70}$$

for arbitrary $p \times q$ matrix $\mathbf{Z}$. Note that the exact size and structure of $\mathbf{G}_p$, $\mathbf{F}_p$ and $\mathbf{K}_{pq}$ depends on the size of $\mathbf{X}$, but for better readability we choose to omit the size-indicating subscripts in the rest of this paper.

Magnus and Neudecker 2019 show (Theorem 3.12) that

$$(\mathbf{I} + \mathbf{K}) = 2\,\mathbf{G}\mathbf{G}^+. \tag{71}$$

Furthermore it holds that

$$\mathbf{G}\mathbf{G}^+ = \mathbf{G}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1}\mathbf{G}^\top = \left(\mathbf{G}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1}\mathbf{G}^\top\right)^\top = \left(\mathbf{G}\mathbf{G}^+\right)^\top \tag{72}$$

and

$$(\mathbf{G}^+)^\top\mathbf{G}^\top \operatorname{vec}(\mathbf{X}) = \left(\left(\mathbf{G}^\top\mathbf{G}\right)^{-1}\mathbf{G}^\top\right)^\top \mathbf{G}^\top\mathbf{G} \operatorname{vech}(\mathbf{X}) = \mathbf{G}\operatorname{vech}(\mathbf{X}) = \operatorname{vec}(\mathbf{X}). \tag{73}$$

For nonsingular matrix $\mathbf{X}$ it holds that (see Lütkepohl 2005, p. 664 or Magnus and

Neudecker 2019, Theorem 3.13)

$$\left(\mathbf{G}^{\top}(\mathbf{X}\otimes\mathbf{X})\,\mathbf{G}\right)^{-1}=\mathbf{G}^{+}\left(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1}\right)\left(\mathbf{G}^{+}\right)^{\top}. \tag{74}$$

**Lemma 7.7.** *For scalar $\alpha$ we have*

$$\left(\mathbf{G}^{\top}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\mathbf{G}\right)^{-1} \tag{75}$$

$$=\mathbf{G}^{+}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}+\frac{\alpha}{1+\alpha p}\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}\right)\mathbf{G}^{+}. \tag{76}$$

*Proof.*

$$\left(\mathbf{F}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\right)^{-1}=\left(\mathbf{G}^{+}\mathbf{G}\mathbf{F}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\mathbf{G}\right)^{-1} \tag{77}$$

$$=\left(\mathbf{G}^{+}\mathbf{G}\mathbf{F}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\mathbf{G}\right)^{-1} \tag{78}$$

$$\text{Magnus and Neudecker 1980 Lemma 4.4 (i):}=\left(\mathbf{G}^{+}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\mathbf{G}\right)^{-1} \tag{79}$$

$$=\left(\mathbf{G}^{\top}\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}+\alpha\mathrm{vec}\,(\boldsymbol{\Sigma})\,\mathrm{vec}\,(\boldsymbol{\Sigma})^{\top}\right)\mathbf{G}\right)^{-1}\mathbf{G}^{\top}\mathbf{G} \tag{80}$$

$$\text{Magnus and Neudecker 1980 Lemma 4.7 (iv):}=\mathbf{F}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}+\frac{\alpha}{1+\alpha p}\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}\right)\mathbf{G} \tag{81}$$

$$\text{Magnus and Neudecker 1980 Lemma 4.4 (i):}=\mathbf{G}^{+}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}+\frac{\alpha}{1+\alpha p}\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}\right)\mathbf{G}. \tag{82}$$

Then

$$
\begin{aligned}
&(80) = (82) && (83)\\
\Leftrightarrow & && (84)\\
&\left(\mathbf{G}^\top\left(\mathbf{\Sigma}\otimes\mathbf{\Sigma}+\alpha\mathrm{vec}\left(\mathbf{\Sigma}\right)\mathrm{vec}\left(\mathbf{\Sigma}\right)^\top\right)\mathbf{G}\right)^{-1} && (85)\\
&\qquad = \mathbf{G}^+\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}+\frac{\alpha}{1+\alpha p}\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right)\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right)^\top\right)\mathbf{G}^+. && (86)
\end{aligned}
$$

□

### 7.5 Matrix Derivatives

For properties of the **differential** “d” see Magnus and Neudecker 2019 pp. 163-169 and pp. 434-436. Some rules are

$$
\begin{aligned}
\mathrm{dvec}\left(\mathbf{X}\right) &= \mathrm{vec}\left(\mathrm{d}\mathbf{X}\right) && (87)\\
\mathrm{dvec}\left(\mathbf{X}\right) &= \mathbf{G}\mathrm{vech}\left(\mathrm{d}\mathbf{X}\right) && (88)\\
\mathrm{dvech}\left(\mathbf{X}\right) &= \mathbf{G}^+\mathrm{vec}\left(\mathrm{d}\mathbf{X}\right) && (89)\\
\mathrm{d}\mathbf{X}^\top &= \left(\mathrm{d}\mathbf{X}\right)^\top && (90)\\
\mathrm{dtr}\left(\mathbf{X}\right) &= \mathrm{tr}\left(\mathrm{d}\mathbf{X}\right) && (91)\\
\mathrm{d}\log\left|\mathbf{X}\right| &= \mathrm{tr}\left(\mathbf{X}^{-1}\mathrm{d}\mathbf{X}\right) && (92)\\
\mathrm{d}\mathbf{X}^{-1} &= -\mathbf{X}^{-1}\mathrm{d}\mathbf{X}\mathbf{X}^{-1} && (93)\\
\mathrm{d}\left(\mathbf{X}\mathbf{X}^\top\right) &= \mathrm{d}\mathbf{X}\mathbf{X}^\top+\mathbf{X}\mathrm{d}\mathbf{X}^\top && (94)
\end{aligned}
$$

To convert differentials to derivatives see Tables 9.2 and 10.1 in Magnus and Neudecker 2019.

There is a difference between gradient and derivative of a scalar valued function that takes multiple input variables. They are transposes of each other, where the gradient is a column vector. See p. 87 Magnus and Neudecker 1999.

For the definition of a matrix derivative we follow Magnus 2010. For an $m \times p$ matrix function $\mathbf{F} = (f_{st})$ of an $n \times q$ matrix of varibles $\mathbf{X} = (x_{ij})$, they define the $\alpha$-derivative as the $mp \times nq$ matrix

$$\frac{\partial \text{vec}\left(\mathbf{F}\left(\mathbf{X}\right)\right)}{\partial \text{vec}\left(\mathbf{X}\right)^{\top}}. \tag{95}$$

Note that

$$\frac{\partial \text{vech}\left(\bullet\right)}{\partial \bullet} = \frac{\partial \mathbf{G}^{+}\text{vec}\left(\bullet\right)}{\partial \bullet} = \mathbf{G}^{+}\frac{\partial \text{vec}\left(\bullet\right)}{\partial \bullet} \tag{96}$$

and

$$\frac{\partial \text{vec}\left(\bullet\right)}{\partial \bullet} = \frac{\partial \mathbf{G}\text{vech}\left(\bullet\right)}{\partial \bullet} = \mathbf{G}\frac{\partial \text{vech}\left(\bullet\right)}{\partial \bullet}. \tag{97}$$

**Lemma 7.8.** *(Magnus and Neudecker 1980, Lemma 3.8). Let $\mathbf{X}$ be a $p \times p$ matrix of variables. Then*

$$\frac{\partial \text{vec}\left(\mathbf{X}\right)}{\partial \text{vech}\left(\mathbf{X}\right)^{\top}} = \begin{cases} \mathbf{F}^{\top}, & \textit{for lower triangular } \mathbf{X}, \\ \mathbf{G}, & \textit{for symmetric } \mathbf{X}, \end{cases} \tag{98}$$

*Proof.* We include a proof for completeness.

$$\text{vec}\left(\text{d}\mathbf{X}\right) = \text{dvec}\left(\mathbf{X}\right) = \begin{cases} \text{d}\mathbf{G}\text{vech}\left(\mathbf{X}\right) = \mathbf{G}\text{dvech}\left(\mathbf{X}\right), & \text{for lower triangular } \mathbf{X}, \\ \text{d}\mathbf{F}^{\top}\text{vech}\left(\mathbf{X}\right) = \mathbf{F}^{\top}\text{dvech}\left(\mathbf{X}\right), & \text{for symmetric } \mathbf{X}, \end{cases} \tag{99}$$

using (65) and (68). $\square$

**Lemma 7.9.** *(Harville 1997, p. 371). Let $\mathbf{X}$ be a non-singular symmetric matrix of variables. Then*

$$\frac{\partial \text{vech}\left(\mathbf{X}^{-1}\right)}{\partial \text{vech}\left(\mathbf{X}\right)^{\top}} = -\mathbf{G}^{+}\left(\mathbf{X}^{-1} \otimes \mathbf{X}^{-1}\right)\mathbf{G}. \tag{100}$$

*Proof.* We include a proof for completeness.

$$\begin{aligned}
\mathrm{dvech}\left(\mathbf{X}^{-1}\right) &= \mathrm{d}\mathbf{G}^{+}\mathrm{vec}\left(\mathbf{X}^{-1}\right) && (101)\\
&= -\mathbf{G}^{+}\mathrm{vec}\left(\mathbf{X}^{-1}\mathrm{d}\mathbf{X}\mathbf{X}^{-1}\right) && (102)\\
&= -\mathbf{G}^{+}\left(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1}\right)\mathrm{dvec}\left(\mathbf{X}\right) && (103)\\
&= -\mathbf{G}^{+}\left(\mathbf{X}^{-1}\otimes\mathbf{X}^{-1}\right)\mathbf{G}\mathrm{dvech}\left(\mathbf{X}\right), && (104)
\end{aligned}$$

using (67), (61) and (65). □

#### 7.5.1 Derivative - Trace

**Lemma 7.10.** *Let $\boldsymbol{\Sigma}$ be a non-singular symmetric $p\times p$ matrix of variables, $\mathbf{C}$ its lower Cholesky factor and $\mathbf{X}$ and $\mathbf{Y}$ be $p\times p$ matrix of constants. Then*

$$\frac{\partial \mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{X}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = \mathrm{vec}\left(\mathbf{X}\right)^{\top}\mathbf{G}, \tag{105}$$

$$\frac{\partial \mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = -\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Sigma}^{-1}\right)^{\top}\mathbf{G}, \tag{106}$$

$$\frac{\partial \mathrm{tr}\left(\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathbf{Y}\right)}{\partial \mathrm{vech}\left(\mathbf{C}\right)^{\top}} = -2\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top}\right)^{\top}\mathbf{F}^{\top}, \tag{107}$$

*and*

$$\frac{\partial^{2} \mathrm{tr}\left(\boldsymbol{\Sigma}\mathbf{X}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = \mathbf{0}. \tag{108}$$

*Proof.* We have

$$
\begin{aligned}
\mathrm{dtr}\left(\mathbf{\Omega}\mathbf{X}\right) &= \mathrm{tr}\left(\mathbf{X}\mathrm{d}\mathbf{\Omega}\right) && (109)\\
&= \mathrm{vec}\left(\mathbf{X}\right)^{\top}\mathrm{vec}\left(\mathrm{d}\mathbf{\Omega}\right) && (110)\\
&= \mathrm{vec}\left(\mathbf{X}\right)^{\top}\mathbf{G}\mathrm{dvech}\left(\mathbf{\Omega}\right), && (111)
\end{aligned}
$$

using (63) and (65),

$$
\begin{aligned}
\mathrm{dtr}\left(\mathbf{\Sigma}^{-1}\mathbf{X}\right) &= -\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathrm{d}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{X}\right) && (112)\\
&= -\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Sigma}^{-1}\mathrm{d}\mathbf{\Sigma}\right) && (113)\\
&= -\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Sigma}^{-1}\right)^{\top}\mathrm{vec}\left(\mathrm{d}\mathbf{\Sigma}\right) && (114)\\
&= -\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\mathbf{X}\mathbf{\Sigma}^{-1}\right)^{\top}\mathbf{G}\mathrm{dvech}\left(\mathbf{\Sigma}\right), && (115)
\end{aligned}
$$

using (62), (63) and (65),

$$
\mathrm{d}^2\mathrm{tr}\left(\mathbf{\Sigma}\mathbf{X}\right) = \mathrm{dtr}\left(\mathbf{X}\mathrm{d}\mathbf{\Sigma}\right) = 0. \qquad (116)
$$

Furthermore,

$$
\begin{aligned}
\mathrm{dtr}\left(\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathbf{Y}\right) &= -\mathrm{tr}\left(\left(\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right)^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathbf{Y} + \mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\mathbf{Y}\right) && (117)\\
&= -\mathrm{tr}\left(\mathrm{d}\mathbf{C}^{\top}\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top} + \mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\right) && (118)\\
&= -2\mathrm{tr}\left(\mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\right) && (119)\\
&= -2\mathrm{vec}\left(\mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\right)^{\top}\mathrm{dvec}\left(\mathbf{C}\right) && (120)\\
&= -2\mathrm{vec}\left(\mathbf{C}^{-1}\mathbf{Y}\mathbf{C}^{-\top}\mathbf{X}\mathbf{C}^{-1}\right)^{\top}\mathbf{F}^{\top}\mathrm{dvech}\left(\mathbf{C}\right), && (121)
\end{aligned}
$$

□

### 7.5.2 Derivative - CYC'

**Lemma 7.11.** *Let $\mathbf{Y}$ be a diagonal matrix. Then*

$$\frac{\partial \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^+\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G}$$

$$= \mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}, \tag{122}$$

$$\frac{\partial \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^\top} = -\mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^+\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\mathbf{G} \tag{123}$$

$$= -\mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^\top\right)^{-1}, \tag{124}$$

$$\frac{\partial \mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^\top} = \mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^+\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\mathbf{G} \tag{125}$$

$$= \mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^\top\right)^{-1} \tag{126}$$

*and*

$$\frac{\partial \mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = -\mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^+\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G} \tag{127}$$

$$= -\mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}. \tag{128}$$

*30*

30. Note that in cases where two expressions are given those have been numerically checked to be the same. The longer versions make immediately obvious the nesting of the case where $\mathbf{Y} = c\mathbf{I}$. Magnus and Neudecker 1980 might offer tools for algebraic derivation of the equalities.

*Proof.* We have

$$\mathrm{dvec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right) = \mathrm{vec}\left(\mathrm{d}\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right) + \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathrm{d}\mathbf{C}^\top\right) \tag{129}$$

$$= \left(\mathbf{I} + \mathbf{K}_{pp}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right) \tag{130}$$

$$= \left(\mathbf{I} + \mathbf{K}_{pp}\right)\left(\mathbf{C}\mathbf{Y} \otimes \mathbf{I}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\right) \tag{131}$$

$$= 2\mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y} \otimes \mathbf{I}\right)\mathbf{F}^\top\mathrm{dvech}\left(\mathbf{C}\right), \tag{132}$$

$$\mathrm{dvec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right) = \mathrm{vec}\left(\mathrm{d}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right) + \mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathrm{d}\mathbf{C}^{-1}\right) \tag{133}$$

$$= -\mathbf{K}_{pp}\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) - \mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) \tag{134}$$

$$= -\left(\mathbf{I} + \mathbf{K}_{pp}\right)\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) \tag{135}$$

$$= -\left(\mathbf{I} + \mathbf{K}_{pp}\right)\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\right) \tag{136}$$

$$= -2\mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top\mathrm{dvech}\left(\mathbf{C}\right), \tag{137}$$

where we used (70), (61) and (71), such that

$$\frac{\partial\mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right)}{\partial\mathrm{vech}\left(\mathbf{C}\right)^\top} = 2\mathbf{G}\mathbf{G}^+\left(\mathbf{C}\mathbf{Y} \otimes \mathbf{I}\right)\mathbf{F}^\top \tag{138}$$

and

$$\frac{\partial\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial\mathrm{vech}\left(\mathbf{C}\right)^\top} = -2\mathbf{G}\mathbf{G}^+\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)\mathbf{F}^\top. \tag{139}$$

Furthermore, according to Lemma 1 in Lütkepohl 1989 we have

$$\frac{\partial\mathrm{vech}\left(\mathbf{C}\right)}{\partial\mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \frac{1}{2}\left(\mathbf{G}^+\left(\mathbf{C} \otimes \mathbf{I}\right)\mathbf{F}^\top\right)^{-1}, \tag{140}$$

for which we include the following proof for completeness,

$$
\begin{aligned}
\mathrm{dvech}\left(\mathbf{\Sigma}\right) &= \mathbf{G}^{+}\mathrm{dvec}\left(\mathbf{C}\mathbf{C}^{\top}\right) && (141)\\
&= \mathbf{G}^{+}\mathrm{vec}\left(\mathrm{d}\mathbf{C}\mathbf{C}^{\top}\right) + \mathbf{G}^{+}\mathrm{vec}\left(\mathbf{C}\mathrm{d}\mathbf{C}^{\top}\right) && (142)\\
&= \mathbf{G}^{+}\left(\mathbf{I} + \mathbf{K}_{pp}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\mathbf{C}^{\top}\right) && (143)\\
&= \mathbf{G}^{+}\mathbf{G}\mathbf{G}^{+}\mathrm{vec}\left(\mathrm{d}\mathbf{C}\mathbf{C}^{\top}\right) && (144)\\
&= 2\mathbf{G}^{+}\left(\mathbf{C} \otimes \mathbf{I}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\right) && (145)\\
&= 2\mathbf{G}^{+}\left(\mathbf{C} \otimes \mathbf{I}\right)\mathbf{F}^{\top}\mathrm{dvech}\left(\mathbf{C}\right). && (146)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\mathrm{dvech}\left(\mathbf{\Sigma}^{-1}\right) &= \mathbf{G}^{+}\mathrm{dvec}\left(\mathbf{C}^{-\top}\mathbf{C}^{-1}\right) && (147)\\
&= -\mathbf{G}^{+}\mathrm{vec}\left(\left[\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right]^{\top}\mathbf{C}^{\top}\right) - \mathbf{G}^{+}\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) && (148)\\
&= -\mathbf{G}^{+}\left(\mathbf{I} + \mathbf{K}_{pp}\right)\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) && (149)\\
&= -\mathbf{G}^{+}\mathbf{G}\mathbf{G}^{+}\mathrm{vec}\left(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\right) && (150)\\
&= -2\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{C}^{-1}\right)\mathrm{vec}\left(\mathrm{d}\mathbf{C}\right) && (151)\\
&= -2\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right)\mathbf{F}^{\top}\mathrm{dvech}\left(\mathbf{C}\right). && (152)
\end{aligned}
$$

such that

$$
\frac{\partial \mathrm{vech}\left(\mathbf{C}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}^{-1}\right)^{\top}} = -\frac{1}{2}\left(\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right)\mathbf{F}^{\top}\right)^{-1}. \qquad (153)
$$

Then the lemma follows by applying the chain rule,

$$
\begin{aligned}
\frac{\partial \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} &= \frac{\partial \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \mathrm{vech}\left(\mathbf{C}\right)^{\top}}\frac{\partial \mathrm{vech}\left(\mathbf{C}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}^{-1}\right)^{\top}}\frac{\partial \mathrm{vech}\left(\mathbf{\Sigma}^{-1}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} && (154)\\
&= \frac{\partial \mathrm{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \mathrm{vech}\left(\mathbf{C}\right)^{\top}}\frac{\partial \mathrm{vech}\left(\mathbf{C}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}}, && (155)
\end{aligned}
$$

$$\frac{\partial \text{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}} = \frac{\partial \text{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} \frac{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}} \tag{156}$$

$$= \frac{\partial \text{vec}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^{\top}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}}, \tag{157}$$

$$\frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}} = \frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} \frac{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}} \tag{158}$$

$$= \frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}}, \tag{159}$$

$$\frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = \frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}} \frac{\partial \text{vech}\left(\boldsymbol{\Sigma}^{-1}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} \tag{160}$$

$$= \frac{\partial \text{vec}\left(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\right)}{\partial \text{vech}\left(\mathbf{C}\right)^{\top}} \frac{\partial \text{vech}\left(\mathbf{C}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}}. \tag{161}$$

$\square$

### 7.5.3 Derivative - Lower Power Weighted Determinant

**Lemma 7.12.**

$$\frac{\partial \log |\boldsymbol{\Sigma}|_{\mathbf{n}}}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = \text{vec}\left(\mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)\mathbf{C}^{-1}\right)^{\top}\mathbf{G} \tag{162}$$

*and*

$$\frac{\partial \log |\boldsymbol{\Sigma}|_{\mathbf{n}}}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)\text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = -\mathbf{G}^{\top}\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{\top}\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^{\top}\right)^{-1}$$

$$\times\,\mathbf{G}^{\top}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G} \tag{163}$$

$$= -\mathbf{G}^{\top}\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1} \tag{164}$$

*Proof.* Decompose $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{D}\mathbf{T}^{\top}$, where $\mathbf{T}$ is a lower triangular matrix with diagonal elements being 1 and $\mathbf{D}$ is a diagonal matrix with positive diagonal elements, such that

$\mathbf{C} = \mathbf{T}\mathbf{D}^{\frac{1}{2}}$ is the lower Cholesky factor. Note that

$$\frac{\partial \sum_{i=1}^{p} n_i \log\left(\mathbf{D}_{ii}\right)}{\partial \mathbf{D}} = \begin{bmatrix} \frac{\mathbf{n}_1}{\mathbf{D}_{11}} & & & \\ & \frac{\mathbf{n}_2}{\mathbf{D}_{22}} & & \\ & & \ddots & \\ & & & \frac{\mathbf{n}_p}{\mathbf{D}_{pp}} \end{bmatrix} = \mathbf{D}^{-\frac{1}{2}} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{D}^{-\frac{1}{2}}, \tag{165}$$

such that

$$\frac{\partial \sum_{i=1}^{p} n_i \log\left(\mathbf{D}_{ii}\right)}{\partial \mathrm{vec}\left(\mathbf{D}\right)^{\top}} = \mathrm{vec}\left(\mathbf{D}^{-\frac{1}{2}} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{D}^{-\frac{1}{2}}\right)^{\top} \tag{166}$$

and

$$\mathrm{dvec}\left(\mathbf{\Sigma}\right) = \mathrm{dvec}\left(\mathbf{T}\mathbf{D}\mathbf{T}^{\top}\right) = \left(\mathbf{T} \otimes \mathbf{T}\right) \mathrm{dvec}\left(\mathbf{D}\right), \tag{167}$$

such that

$$\frac{\partial \mathrm{vec}\left(\mathbf{D}\right)}{\partial \mathrm{vec}\left(\mathbf{\Sigma}\right)^{\top}} = \left(\mathbf{T} \otimes \mathbf{T}\right)^{-1}. \tag{168}$$

Then

$$\begin{aligned}
\frac{\partial \log\left|\mathbf{\Sigma}\right|_{\mathbf{n}}}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} &= \frac{\partial \sum_{i=1}^{p} n_i \log\left(\mathbf{D}_{ii}\right)}{\partial \mathrm{vec}\left(\mathbf{D}\right)^{\top}} \frac{\partial \mathrm{vec}\left(\mathbf{D}\right)}{\partial \mathrm{vec}\left(\mathbf{\Sigma}\right)^{\top}} \frac{\partial \mathrm{vec}\left(\mathbf{\Sigma}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} && (169)\\
&= \mathrm{vec}\left(\mathbf{D}^{-\frac{1}{2}} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{D}^{-\frac{1}{2}}\right)^{\top} \left(\mathbf{T} \otimes \mathbf{T}\right)^{-1} \mathbf{G} && (170)\\
&= \mathrm{vec}\left(\mathbf{D}^{-\frac{1}{2}} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{D}^{-\frac{1}{2}}\right)^{\top} \left(\mathbf{T}^{-\top} \otimes \mathbf{T}^{-\top}\right)^{\top} \mathbf{G} && (171)\\
&= \mathrm{vec}\left(\mathbf{T}^{-\top}\mathbf{D}^{-\frac{1}{2}} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{D}^{-\frac{1}{2}}\mathbf{T}^{-1}\right)^{\top} \mathbf{G} && (172)\\
&= \mathrm{vec}\left(\mathbf{C}^{-\top} \mathrm{dg}\left(\mathbf{n}\right) \mathbf{C}^{-1}\right)^{\top} \mathbf{G}, && (173)
\end{aligned}$$

where we used (61).

Now application of Lemma (7.11) and using Theorem 3.13 (c) of Magnus and Neudecker

2019 on

$$\frac{\partial \log |\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}\,(\mathbf{\Sigma})\, \mathrm{vech}\,(\mathbf{\Sigma})^{\top}} = \mathbf{G}^{\top} \frac{\partial \mathrm{vec}\left(\mathbf{C}^{-\top} \mathrm{dg}\,(\mathbf{n})\, \mathbf{C}^{-1}\right)}{\partial \mathrm{vech}\,(\mathbf{\Sigma})^{\top}} \tag{174}$$

gives

$$\begin{aligned}\frac{\partial \log |\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}\,(\mathbf{\Sigma})\, \mathrm{vech}\,(\mathbf{\Sigma})^{\top}} = -\mathbf{G}^{\top} &\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top} \mathrm{dg}(\mathbf{n}) \mathbf{C}^{-1}\right) \mathbf{F}^{\top} \\ &\times \left(\mathbf{G}^{+} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \mathbf{G}^{+} \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G},\end{aligned}$$

with

$$\begin{aligned}&\left(\mathbf{G}^{+} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \mathbf{G}^{+} \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G} \\ &= \left(\mathbf{G}^{\top} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \mathbf{G}^{\top}\mathbf{G}\mathbf{G}^{+} \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G} \\ &= \left(\mathbf{G}^{\top} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \mathbf{G}^{\top} \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G} \\ &= \left(\mathbf{G}^{\top} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \left(\mathbf{G}^{+} \left(\mathbf{\Sigma} \otimes \mathbf{\Sigma}\right) \left(\mathbf{G}^{+}\right)^{\top}\right)^{-1} \\ &= \left(\mathbf{G}^{+} \left(\mathbf{\Sigma} \otimes \mathbf{\Sigma}\right) \left(\mathbf{G}^{+}\right)^{\top} \mathbf{G}^{\top} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \\ &= \left(\mathbf{G}^{+} \left(\mathbf{\Sigma} \otimes \mathbf{\Sigma}\right) \left(\mathbf{G}\mathbf{G}^{+}\right)^{\top} \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \\ &= \left(\mathbf{G}^{+}\mathbf{G}\mathbf{G}^{+} \left(\mathbf{\Sigma} \otimes \mathbf{\Sigma}\right) \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \\ &= \left(\mathbf{G}^{+} \left(\mathbf{\Sigma} \otimes \mathbf{\Sigma}\right) \left(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{F}^{\top}\right)^{-1} \\ &= \left(\mathbf{G}^{+} \left(\mathbf{C} \otimes \mathbf{I}\right) \mathbf{F}^{\top}\right)^{-1}.\end{aligned}$$

□

**Lemma 7.13.** *Let* $\mathbf{\Omega} = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^{\top}$*. Then*

$$\left(\frac{\partial \mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}}\right)^{\top} \frac{\partial \log|\mathbf{\Omega}|_{\mathbf{n}}}{\partial \mathrm{vech}\left(\mathbf{\Omega}^{-1}\right) \mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)^{\top}} \frac{\partial \mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} \tag{175}$$

$$= \left(-\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}^{-1}\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\right)^{\top} \tag{176}$$

$$\times\, \mathbf{G}^{\top}\left(\mathbf{C}_{\mathbf{\Omega}}\,\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}^{-\top}\otimes\mathbf{\Omega}^{-1}\right)\mathbf{F}^{\top}\right)^{-1} \tag{177}$$

$$\times\left(-\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}^{-1}\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\right) \tag{178}$$

$$= \left(-\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}^{-1}\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\right)^{\top} \tag{179}$$

$$\times\, \mathbf{G}^{\top}\left(\mathbf{C}\,\mathrm{dg}(\mathbf{n})^{1/2}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}^{-\top}\otimes\mathbf{\Omega}^{-1}\right)\mathbf{F}^{\top}\right)^{-1} \tag{180}$$

$$\times\left(-\mathbf{G}^{+}\left(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathbf{Y}^{-1}\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\right) \tag{181}$$

$$\tag{182}$$

Another important related relation rewriting the derivative of the Cholesky factor w.r.t. its "square matrix" (see equation (140)) derived in Murray 2016 based on (https://mathoverflow.net/u pav) is given in the following Lemma.

**Lemma 7.14.**

$$\left(\frac{\partial \mathrm{vech}\left(\mathbf{C}\mathbf{C}^{\top}\right)}{\partial \mathrm{vech}\left(\mathbf{C}\right)^{\top}}\right)^{-1} \overset{\textit{Lütkepohl 1989}}{=} \frac{1}{2}\left(\mathbf{G}^{+}\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1} \tag{183}$$

$$\overset{\textit{Murray 2016}}{=} \mathbf{F}\left(\mathbf{I}\otimes\mathbf{C}\right)\mathbf{Z}\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{184}$$

$$\overset{\textit{comment on mo}}{=} \frac{1}{2}\mathbf{F}\left(\mathbf{I}\otimes\mathbf{C}\right)\left(\mathbf{G}\mathbf{F}\right)^{\top}\left(\mathbf{G}\mathbf{F}\right)\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{185}$$

$$\overset{\textit{not proved yet}}{=} \frac{1}{2}\mathbf{F}\left(\mathbf{I}\otimes\mathbf{C}\right)\left(\mathbf{G}\mathbf{F}\right)^{\top}\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G}, \tag{186}$$

*where* $\mathbf{Z} = \frac{1}{2}(\mathbf{G}\mathbf{F})^{\top}(\mathbf{G}\mathbf{F})$ *(see comments in mathoverflow post) is a diagonal matrix defined such that for any square matrix* $\mathbf{A}$*,* $\mathbf{Z}\mathrm{vec}(\mathbf{A}) = \mathrm{vec}(\Phi(\mathbf{A}))$*, where* $\Phi(\mathbf{A}) = \mathrm{tril}\left(\mathbf{A} - \frac{1}{2}\mathbf{I}\odot\mathbf{A}\right)$ *returns* $\mathbf{A}$ *with its upper triangular part set to 0's and its diagonals halved.*

The next important one is

**Lemma 7.15.**

$$\mathbf{F}^\top \mathbf{F}\mathrm{vec}\left(\mathbf{A}\right) = \mathrm{vec}\left(\mathrm{tril}\left(\mathbf{A}\right)\right) \quad \textit{and} \tag{187}$$

$$\frac{1}{2}(\mathbf{GF})^\top (\mathbf{GF})\mathrm{vec}\left(\mathbf{A}\right) = \mathrm{vec}\left(\Phi\left(\mathbf{A}\right)\right), \tag{188}$$

*Proof.* The first one is obvious, since $\mathbf{F}$ eliminates those elements from vec($\mathbf{A}$) which are on $\mathbf{A}$'s upper triangular part and then $\mathbf{F}^\top$ we know from equation (68) is the matrix which maps $\mathrm{vech}_\triangle$ to vec.

The second one we know from the aforementioned comment on mathoverflow. For symmetric $\mathbf{A}$ it obviously reduces to $\frac{1}{2}(\mathbf{GF})^\top \mathrm{vec}\left(\mathbf{A}\right)$. □

Finally this helps us with rewriting the scores in a format that is quick to evaluate, since it avoids Kronecker product products or inversions. To see this, consider

**Lemma 7.16.** *Let* $\mathbf{X}$ *be a matrix,* $\mathbf{Y}$ *a diagonal matrix,* $\mathbf{C}$ *a lower triangular,* $\mathbf{G}$, $\mathbf{G}^+$ *and* $\mathbf{F}$ *the duplication, its Moore-Penrose inverse and the canonical elimination matrix, as defined above. Then*

$$\mathrm{vec}\left(\mathbf{X}\right)^\top \mathbf{G}\frac{\partial \mathrm{vech}\left(\mathbf{CYC}^\top\right)}{\partial \mathrm{vech}\left(\mathbf{CC}^\top\right)} = 2\,\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top \mathrm{tril}(\mathbf{XCY})\right)\mathbf{C}^{-1}\right)^\top \mathbf{G}. \tag{189}$$

*Proof.*

$$\operatorname{vec}(\mathbf{X})^\top \mathbf{G}\frac{\partial \operatorname{vech}\left(\mathbf{C}\mathbf{Y}\mathbf{C}^\top\right)}{\partial \operatorname{vech}\left(\mathbf{C}\mathbf{C}^\top\right)} \tag{190}$$

$$\overset{(122)}{=} \operatorname{vec}(\mathbf{X})^\top \mathbf{G}\mathbf{G}^+ (\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\,\mathbf{F}^\top \left(\mathbf{G}^+ (\mathbf{C}\otimes\mathbf{I})\,\mathbf{F}^\top\right)^{-1} \tag{191}$$

$$\overset{(73)}{=} \operatorname{vec}(\mathbf{X})^\top (\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\,\mathbf{F}^\top \left(\mathbf{G}^+ (\mathbf{C}\otimes\mathbf{I})\,\mathbf{F}^\top\right)^{-1} \tag{192}$$

$$= \operatorname{vec}(\mathbf{X}\mathbf{C}\mathbf{Y})^\top \mathbf{F}^\top \left(\mathbf{G}^+ (\mathbf{C}\otimes\mathbf{I})\,\mathbf{F}^\top\right)^{-1} \tag{193}$$

$$= \operatorname{vec}(\mathbf{X}\mathbf{C}\mathbf{Y})^\top \mathbf{F}^\top\mathbf{F}\,(\mathbf{I}\otimes\mathbf{C})\,(\mathbf{G}\mathbf{F})^\top(\mathbf{G}\mathbf{F})\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{194}$$

$$\overset{(187)}{=} \operatorname{vec}(\operatorname{tril}(\mathbf{X}\mathbf{C})\mathbf{Y})^\top (\mathbf{I}\otimes\mathbf{C})\,(\mathbf{G}\mathbf{F})^\top(\mathbf{G}\mathbf{F})\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{195}$$

$$= \operatorname{vec}\left(\mathbf{C}^\top\operatorname{tril}(\mathbf{X}\mathbf{C})\mathbf{Y}\right)^\top (\mathbf{G}\mathbf{F})^\top(\mathbf{G}\mathbf{F})\left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{196}$$

$$\overset{(188)}{=} 2\operatorname{vec}\left(\Phi\left(\mathbf{C}^\top\operatorname{tril}(\mathbf{X}\mathbf{C})\mathbf{Y}\right)\right)^\top \left(\mathbf{C}^{-1}\otimes\mathbf{C}^{-1}\right)\mathbf{G} \tag{197}$$

$$= 2\operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}(\mathbf{X}\mathbf{C})\mathbf{Y}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{198}$$

$$= 2\operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}(\mathbf{X}\mathbf{C}\mathbf{Y})\right)\mathbf{C}^{-1}\right)^\top \mathbf{G}. \tag{199}$$

Note that

$$\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{tril}(\mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n}))\right)\mathbf{C}^{-1} = \mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\operatorname{dg}(\mathbf{C}^{-\top})\operatorname{dg}(\mathbf{n})\right)\mathbf{C}^{-1} \tag{200}$$

$$= \mathbf{C}^{-\top}\operatorname{dg}\left(\mathbf{C}^\top\right)\operatorname{dg}(\mathbf{C}^{-\top})\operatorname{dg}(\mathbf{n})\mathbf{C}^{-1} \tag{201}$$

$$= \mathbf{C}^{-\top}\operatorname{dg}(\mathbf{n})\mathbf{C}^{-1}. \tag{202}$$

□

## 7.6 Distributions

In the distribution-specific subsections below we omit subscripts $i$ for indication of the respective distribution. For all distributions, $\mathbf{R}$ denotes the symmetric positive definite random matrix and

$$\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top := \mathbb{E}\left[\mathbf{R}\right] \tag{203}$$

denotes the respective expected value, where $\mathbf{C}$ denotes the lower Cholesky factor of $\mathbf{\Sigma}$. As described in section 2.1 EXPLAIN RESTRICTIONS ON DOFS AND DISTRIBUTIONS SPECIFIC DEFINITION OF OMEGA HERE. For the non-inverse Riesz type $I$ named distributions and for the inverse Riesz type $II$ named distributions we have to pre- and post multiply the random elements of the stochastic representations with the lower Cholesky factor of $\mathbf{\Sigma}$, whereas for the non-inverse Riesz type $II$ named and for the inverse Riesz type $I$ named distributions one has to use the upper Cholesky factor. To stay consistent and since we want to work only with one type of Cholesky decomposition, we choose type $I$ distibutions for the non-inverted Riesz named and type $II$ distributions for the inverted ones.

For application in the GAS framework it is instructive to standardize all distributions (rewrite the pdfs in terms of $\mathbf{\Sigma}$ rather than their usual parameter matrix $\mathbf{\Omega}$) and derive the scores, Fisher information matrices and covariance matrices of the *standardized distributions*.

In the following we note the stochastic representation, pdf and expected value of the non-standardized distributions, then derive their standardized pdfs, covariance matrices, scores and fisher information matrices. All distributions we consider in this paper are matrix-variate distributions which have exactly one real symmetric positive definite parameter matrix (which we always denote by $\mathbf{\Omega}$) and one or more real scalar degree of freedom parameter (stacked in the vector $\boldsymbol{\theta}$) and whos expected value $\mathbf{\Sigma}$ is also a symmetric positive definite matrix. We denote the score with respect to $\mathbf{\Sigma}$ as

$$\nabla = \left( \frac{\partial \log \mathcal{L}}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} \right)^\top , \tag{204}$$

the Fisher information matrix with respect to $\mathbf{\Sigma}$ as

$$\mathcal{I} = \mathbb{E}\left[ \frac{\partial \log \mathcal{L}}{\partial \text{vech}\left(\mathbf{\Sigma}\right)} \left( \frac{\partial \log \mathcal{L}}{\partial \text{vech}\left(\mathbf{\Sigma}\right)} \right)^\top \right] = -\mathbb{E}\left[ \frac{\partial^2 \log \mathcal{L}}{\partial \text{vech}\left(\mathbf{\Sigma}\right) \partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} \right], \tag{205}$$

and our two versions of the standardized score as

$$\mathbf{S} = \text{ivech}\left(\mathcal{I}^{-1}\nabla\right) \tag{206}$$

and

$$\dot{\mathbf{S}} = \text{ivech}\left(\mathbf{Cov}\left(\text{vech}(\mathbf{R})\right)\nabla\right). \tag{207}$$

**Theorem 7.1.** *Let $p^I$ denote a type-I distribution and $p^{II}$ denote the corresponding a type-II distribution. Then*

$$p^I(\boldsymbol{\Omega}, \boldsymbol{\theta}) = p^{II}(\overleftarrow{\boldsymbol{\Omega}}, \overleftarrow{\boldsymbol{\theta}}) \tag{208}$$

*and*

$$p^I(\boldsymbol{\Sigma}, \boldsymbol{\theta}) = p^{II}(\overleftarrow{\boldsymbol{\Sigma}}, \overleftarrow{\boldsymbol{\theta}}), \tag{209}$$

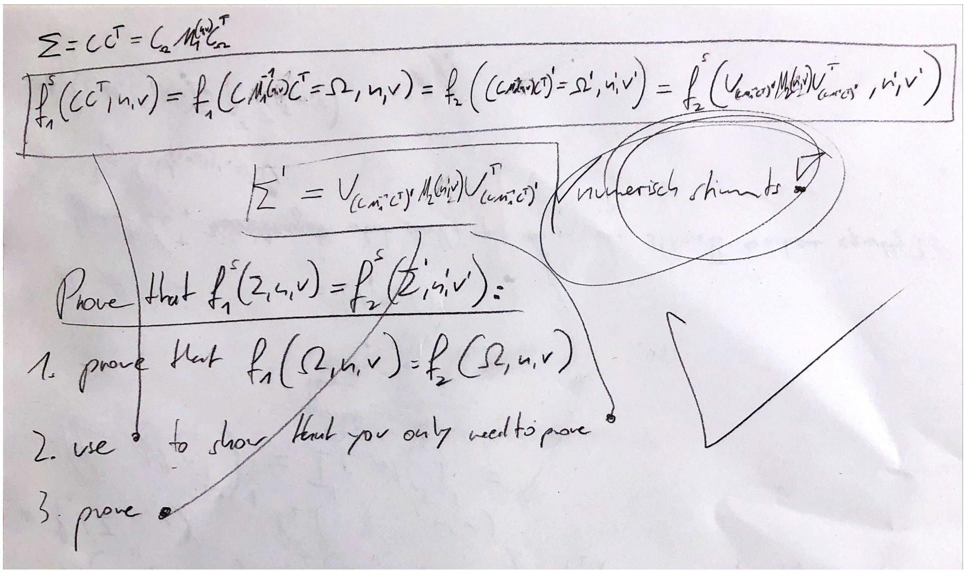
*where,*

$$\begin{aligned}
&\textit{if } \boldsymbol{\theta} = \mathbf{n}, \textit{ then } \overleftarrow{\boldsymbol{\theta}} = \overleftarrow{\boldsymbol{\nu}}, \\
&\textit{if } \boldsymbol{\theta} = \boldsymbol{\nu}, \textit{ then } \overleftarrow{\boldsymbol{\theta}} = \overleftarrow{\mathbf{n}} \textit{ and} \\
&\textit{if } \boldsymbol{\theta} = (\mathbf{n}^\top, \boldsymbol{\nu}^\top)^\top, \textit{ then } \overleftarrow{\boldsymbol{\theta}} = (\overleftarrow{\boldsymbol{\nu}}^\top, \overleftarrow{\mathbf{n}}^\top)^\top.
\end{aligned}$$

**Theorem 7.2.** *Let $p_I^s$ denote a standardized type-I distribution and $p_{II}^s$ denote the corresponding standardized type-II distribution. Then*

$$p^I(\boldsymbol{\Sigma}, \boldsymbol{\theta}) = p^{II}(\overleftarrow{\boldsymbol{\Sigma}}, \overleftarrow{\boldsymbol{\theta}}), \tag{210}$$

*with $\overleftarrow{\boldsymbol{\theta}} = (\overleftarrow{\mathbf{n}}, [\overleftarrow{\boldsymbol{\nu}}])$.*

*Proof.* 


∎

**Theorem 7.3 (Expectation of $\underline{\mathbf{B}}\underline{\mathbf{B}}^\top$).** *Let $(\underline{\mathbf{B}}_{ij})_{1\le j\le i\le p}$ be independent with $\underline{\mathbf{B}}_{ii} \sim \chi^2_{n_i-i+1}$, $n_i \geq i-1$, and $\underline{\mathbf{B}}_{ij} \sim \mathcal{N}(0,1)$, i.e.*

$$\underline{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{n_1-1+1}} & 0 & \dots & 0 \\ \mathcal{N}(0,1) & \ddots & 0 & \vdots \\ \vdots & \mathcal{N}(0,1) & \ddots & 0 \\ \mathcal{N}(0,1) & \dots & \mathcal{N}(0,1) & \sqrt{\chi^2_{n_p-p+1}} \end{bmatrix}. \tag{211}$$

*Then*

$$\mathbb{E}\left[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right] = \mathrm{dg}\left(\mathbf{n}\right), \tag{212}$$

*with $\mathbf{n} = (n_1, n_2, \dots, n_p)^\top$.*

*Proof.* We have

$$\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right)_{ij} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}\left(\underline{\mathbf{B}}^\top\right)_{kj} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}. \tag{213}$$

For the off-diagonal elements, i.e. $i \neq j$, we have

$$\mathbb{E}\left[\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\right)_{ij}\right] = \sum_{k=1}^{p} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}\right] = \sum_{k=1}^{p} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\right]\mathbb{E}\left[\underline{\mathbf{B}}_{jk}\right] = 0, \tag{214}$$

where we have used independence of the elements in $\underline{\mathbf{B}}$ and the fact that at least one of the elements in each summand above is a mean zero normal random variable. For the diagonal elements, i.e. $i = j$, we have

$$\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\right)_{ii} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}^{2} = \sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^{2}, \tag{215}$$

which is the sum of a $\chi^2_{n_i-i+1}$ and $(i-1)$ independent $\mathcal{N}(0,1)^2$ random variables, which implies that

$$\sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^{2} \sim \chi^2_{n_i} \tag{216}$$

with expectation $n_i$. Thus

$$\mathbb{E}\left[\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\right)_{ii}\right] = n_i. \tag{217}$$

$\square$

**Theorem 7.4 (Expectation of $\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}$).** *Let $(\bar{\mathbf{B}}_{ij})_{1\le i\le j\le p}$ be independent with $\bar{\mathbf{B}}_{ii} \sim \chi^2_{\nu_i-p+i}$, $\nu_i \geq p-i$, and $\bar{\mathbf{B}}_{ij} \sim \mathcal{N}(0,1)$, i.e.*

$$\bar{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{\nu_1-p+1}} & \mathcal{N}(0,1) & \dots & \mathcal{N}(0,1) \\ 0 & \ddots & \mathcal{N}(0,1) & \vdots \\ \vdots & 0 & \ddots & \mathcal{N}(0,1) \\ 0 & \dots & 0 & \sqrt{\chi^2_{\nu_p-p+p}}, \end{bmatrix}, \tag{218}$$

*Then*

$$\mathbb{E}\left[\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}\right)^{-1}\right] = \mathrm{dg}\left(\mathbf{m}\right) \tag{219}$$

*where the entries of the vector* $\mathbf{m} = (m_1, \ldots, m_p)^{\top}$ *are given by*

$$m_i = \begin{cases} \frac{1}{n_i - p - 1}, & \text{for } i = 1 \\ \frac{1}{n_i - p + i - 2}\left(1 + \sum_{j=1}^{j-1} \mathbf{m}_j\right) & \text{for } i > 1. \end{cases} \tag{220}$$

*Proof.* ToDo □

### 7.6.1 Riesz

**Theorem 7.5.** *Let*

$$\mathbf{R} = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^{\top}, \tag{221}$$

*then* $\mathbf{R}$ *is said to follow a standardized Riesz distribution of type Idenoted by* $\mathbf{R} \sim \mathcal{R}^{I}(\mathbf{\Sigma}, \mathbf{n})$. *The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma}, \mathbf{n}) = \frac{\prod_{i=1}^{p} n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}} \frac{1}{\Gamma_p\left(\mathbf{n}/2\right)} \left|\mathbf{R}\right|^{-\frac{p+1}{2}} \left|\mathbf{Z}\right|_{\frac{\mathbf{n}}{2}} \mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{Z}\right). \tag{222}$$

*The expected value* $\mathbb{E}\left[\mathbf{R}\right]$ *obtains as*

$$\mathbb{E}\left[\mathbf{R}\right] = \mathbf{\Sigma}, \tag{223}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top} - \mathbf{C}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\right)\right)\mathbf{C}^{-1}\right) \tag{224}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = \frac{1}{2}\mathbf{G}^\top\mathbf{G}\left(\mathbf{F}(\mathbf{C}^\top \otimes \mathbf{I})\mathbf{G}\right)^{-1}\mathbf{F}\left(\mathbf{C}^{-1} \otimes \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{G}. \tag{225}$$

*Proof.* For the probability density function of $\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top$ see Blasques et al. 2021, theorem 4, with $\mathbf{\Sigma}_{Blasques} = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top$, i.e. $\mathbf{C}_{Blasques} = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}$.[31]. Using Lemma (7.1) of this paper and Lemma 3 (v) of Blasques et al. 2021 we have

$$p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n}) = \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p(\mathbf{n}/2)}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(\left|\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right|_{\frac{\mathbf{n}}{2}}\right)^{-1}\mathrm{etr}\left(-\frac{1}{2}\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\mathbf{R}\right) \tag{226}$$

$$= \frac{\prod_{i=1}^p n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}}\frac{1}{\Gamma_p(\mathbf{n}/2)}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}|\mathbf{\Sigma}|_{-\frac{\mathbf{n}}{2}}\mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}\right) \tag{227}$$

$$= \frac{\prod_{i=1}^p n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}}\frac{1}{\Gamma_p(\mathbf{n}/2)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}\mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}(\mathbf{n})\mathbf{Z}\right). \tag{228}$$

For the expected value simply apply thereom 7.3,

$$\mathbb{E}\left[\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top\right] = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\mathbb{E}\left[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right]\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top \tag{229}$$

$$= \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\mathrm{dg}(\mathbf{n})\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top = \mathbf{\Sigma}. \tag{230}$$

Now, define for better readability $\mathbf{\Omega} = \mathbf{\Sigma}_{Blasques}$. For the score w.r.t. $\mathbf{\Sigma}$ start from equation (226), such that

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} = \frac{\log p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n})}{\partial \mathrm{vech}(\mathbf{\Omega})^\top}\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \tag{231}$$

$$= \left(\frac{\partial \log|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\mathbf{\Omega})^\top} - \frac{1}{2}\frac{\partial \mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\partial \mathrm{vech}(\mathbf{\Omega})^\top}\right)\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \tag{232}$$

$$(7.10),(7.12) = \frac{1}{2}\mathrm{vec}\left(-\mathbf{C}_{\mathbf{\Omega}}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\mathbf{\Omega}}^{-1} + \mathbf{\Omega}^{-1}\mathbf{R}\mathbf{\Omega}^{-1}\right)^\top\mathbf{G}\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \tag{233}$$

$$(199) = \mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\mathrm{tril}\left(\mathbf{\Omega}^{-1}\mathbf{R}\mathbf{\Omega}^{-1}\mathbf{C} - \mathbf{C}_{\mathbf{\Omega}}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\mathbf{\Omega}}^{-1}\mathbf{C}\right)\mathrm{dg}(\mathbf{n})^{-1}\right)\mathbf{C}^{-1}\right)^\top\mathbf{G}. \tag{234}$$

31. See also Kessentini, Tounsi, and Zine 2020

Now, using the definition of $\boldsymbol{\Omega}$ and $\mathbf{C}_{\boldsymbol{\Omega}} = \mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}$ this reduces to

$$\operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\operatorname{tril}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1}\mathbf{C} - \mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{C}\right)\mathrm{dg}(\mathbf{n})^{-1}\right)\mathbf{C}^{-1}\right)^{\top}\mathbf{G} \tag{235}$$

$$= \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\operatorname{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n}) - \mathbf{C}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)^{2}\right)\mathrm{dg}(\mathbf{n})^{-1}\right)\mathbf{C}^{-1}\right)^{\top}\mathbf{G} \tag{236}$$

$$= \operatorname{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\operatorname{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top} - \mathbf{C}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\right)\right)\mathbf{C}^{-1}\right)^{\top}\mathbf{G}. \tag{237}$$

For the Fisher information matrix with respect to $\boldsymbol{\Sigma}$ we first derive the Fisher information matrix with respect to $\boldsymbol{\Omega}$. Using Lemma **??**,

$$\mathcal{I}_{\boldsymbol{\Omega}} = -\left(\frac{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}}\right)^{\top}\mathbb{E}\left[\frac{\partial^{2}\log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n})}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\right]\frac{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}} \tag{238}$$

where the expectation, using Lemma 7.10 and Lemma 7.12, boils down to

$$\mathbb{E}\left[\frac{\partial^{2}\log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n})}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\right] \tag{239}$$

$$= \mathbb{E}\left[\frac{\partial^{2}\log|\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} - \frac{\partial^{2}\frac{\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{2}}{\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial\operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\right] \tag{240}$$

$$= -\frac{1}{2}\mathbf{G}^{\top}\left(\mathbf{C}_{\boldsymbol{\Omega}}\,\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{\top}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\mathbf{G}^{\top}\left(\boldsymbol{\Omega}\otimes\boldsymbol{\Omega}\right)\mathbf{G}. \tag{241}$$

Then, using Lemma (7.9), we have

$$\mathcal{I}_{\mathbf{\Omega}} = \mathbf{G}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right) \left(\mathbf{G}^+\right)^\top \tag{242}$$

$$\times \frac{1}{2}\mathbf{G}^\top \left(\mathbf{C}_{\mathbf{\Omega}}\, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^\top \left(\mathbf{C}_{\mathbf{\Omega}} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1} \mathbf{G}^\top \left(\mathbf{\Omega} \otimes \mathbf{\Omega}\right) \mathbf{G} \tag{243}$$

$$\times \mathbf{G}^+ \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right) \mathbf{G} \tag{244}$$

$$= \frac{1}{2}\mathbf{G}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right) \left(\mathbf{G}\mathbf{G}^+\right)^\top \left(\mathbf{C}_{\mathbf{\Omega}}\, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\mathbf{C}_{\mathbf{\Omega}} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1} \tag{245}$$

$$= \frac{1}{2}\mathbf{G}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right) \mathbf{G}\mathbf{G}^+ \left(\mathbf{C}_{\mathbf{\Omega}}\, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\mathbf{C}_{\mathbf{\Omega}} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1} \tag{246}$$

$$= \frac{1}{2}\mathbf{G}^\top \left(\mathbf{\Omega}^{-1} \otimes \mathbf{\Omega}^{-1}\right) \left(\mathbf{C}_{\mathbf{\Omega}}\, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\mathbf{C}_{\mathbf{\Omega}} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1} \tag{247}$$

$$= \frac{1}{2}\mathbf{G}^\top \left(\mathbf{C}_{\mathbf{\Omega}}^{-\top}\, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{\Omega}^{-1}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\mathbf{C}_{\mathbf{\Omega}} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1}. \tag{248}$$

Using again Lemma **??**

$$\mathcal{I} = \left(\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top}\right)^\top \mathcal{I}_{\mathbf{\Omega}} \frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top}, \tag{249}$$

rewrite in terms of $\mathbf{C}$ and $\mathbf{\Sigma}$, on the right multiplication you can use equation (**??**) on the left hand side standard arguments to arrive at

$$\frac{1}{2}\mathbf{G}^\top\mathbf{G}\left(\mathbf{F}(\mathbf{C}^\top \otimes \mathbf{I})\mathbf{G}\right)^{-1} \mathbf{F}\left(\mathbf{C}^{-1} \otimes \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{G}. \tag{250}$$

Later on, for the derivation of the Fisher information matrix of the t-Riesz distribution we will need the covariance matrix of vech($\mathbf{R}$), which can be derived as follows. First, see §2.1.2 of Kollo and Rosen 2005 for the characteristic function of a patterned (in our case symmetric) matrix variate distribution. Note there are two approaches here. Either we ignore symmetry and get the characteristic function of vec($\mathbf{R}$) or we take it into account by getting the characteristic function of e.g. vech ($\mathbf{R}$). In consistency with the rest of this paper we take symmetry into account. Díaz-García 2013 and Gribisch and Hartkopf 2022 don't. Gupta and Nagar 2000 and Kollo and Rosen 2005 do. The characteristic function of $\mathrm{vech}\left(\mathbf{A}\right)$ where $\mathbf{A} = \underline{\mathbf{B}}\underline{\mathbf{B}}^\top$ is given by

$$
\begin{aligned}
\phi(\mathbf{Z}) &= \mathbb{E}\left[e^{i\,\mathrm{vech}(\mathbf{Z})^\top \mathrm{vech}(\mathbf{A})}\right] && (251)\\
&= \mathbb{E}\left[\mathrm{etr}\left(i\frac{1}{2}\left(\mathbf{Z}+\mathbf{Z}\right)\mathbf{A}\right)\right] \text{ p. 244 Kollo and Rosen 2005} && (252)\\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}}\,\mathrm{etr}\left(i\frac{1}{2}\left(\mathbf{Z}+\mathbf{Z}\right)\mathbf{A}\right)\mathrm{etr}\left(-\frac{1}{2}\mathbf{I}\mathbf{A}\right) && (253)\\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)\mathbf{A}\right) && (254)\\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\Gamma_p\left(\mathbf{n}/2\right)\left|\frac{1}{2}\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}} && (255)\\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}}2^{p\bar{\mathbf{n}}/2}\left|\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}} && (256)\\
&= \left|\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}}, && (257)
\end{aligned}
$$

where $\mathbf{Z}$ is a diagonal matrix with elements $\mathrm{dg}(\mathbf{Z})$ and where we used Lemma (7.4). See also Díaz-García 2013, Lemma 1.

Denote $\mathbf{\Xi} = \mathbf{I} - i\left(\mathbf{Z}+\mathbf{Z}\right)$, then

$$
\frac{\partial \mathrm{vech}\left(\mathbf{\Xi}\right)}{\partial \mathrm{vech}\left(\mathbf{Z}\right)^\top} = -i2\left(\mathbf{G}^\top\mathbf{G}\right)^{-1} = -i2\left(\mathbf{G}^\top\mathbf{G}\right)^{-\top} \qquad (258)
$$

*Commented out is an alternative way using the characteristic function to derive the expectation.*

Then

$$
\begin{aligned}
&\frac{\partial^2 \phi(\mathbf{Z})}{\partial \mathrm{vech}\left(\mathbf{Z}\right) \partial \mathrm{vech}\left(\mathbf{Z}\right)^\top} = i \frac{\partial}{\partial \mathrm{vech}\left(\mathbf{Z}\right)} \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)^\top \\
&= i \frac{\partial \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}\left(\mathbf{Z}\right)} \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)^\top + i \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \frac{\partial \mathrm{vec}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)}{\partial \mathrm{vech}\left(\mathbf{Z}\right)} \left(\mathbf{G}^+\right)^\top \\
&= i^2 \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right) \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)^\top \\
&\quad - i \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \left( \frac{\partial \mathrm{vec}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Xi}\right)^\top} \frac{\partial \mathrm{vech}\left(\boldsymbol{\Xi}\right)}{\partial \mathrm{vech}\left(\mathbf{Z}\right)^\top} \right)^\top \left(\mathbf{G}^+\right)^\top \\
&= i^2 \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right) \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)^\top + i^2 2 \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \\
&\quad \times \left( \mathbf{G}\mathbf{G}^+ \left(\dot{\mathbf{C}}_\xi \, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\dot{\mathbf{C}}_\xi{}^{-\top} \otimes \boldsymbol{\Xi}\right) \mathbf{F}^\top\right)^{-1} \left(\mathbf{G}^\top \mathbf{G}\right)^{-\top} \right)^\top \left(\mathbf{G}^+\right)^\top \\
&= i^2 \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right) \mathrm{vech}\left(\dot{\mathbf{C}}_\xi \; \mathrm{dg}\left(\mathbf{n}\right) \dot{\mathbf{C}}_\xi{}^\top\right)^\top + i^2 2 \left|\boldsymbol{\Xi}^{-1}\right|_{\frac{\mathbf{n}}{2}} \\
&\quad \times \left(\mathbf{G}^\top \mathbf{G}\right)^{-1} \left( \mathbf{G}\mathbf{G}^+ \left(\dot{\mathbf{C}}_\xi \, \mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\dot{\mathbf{C}}_\xi{}^{-\top} \otimes \boldsymbol{\Xi}\right) \mathbf{F}^\top\right)^{-1} \right)^\top \left(\mathbf{G}^+\right)^\top,
\end{aligned}
$$

where $\dot{\mathbf{C}}_\xi$ is the lower Cholesky factor of $\boldsymbol{\Xi}^{-1}$, such that

$$
\begin{aligned}
&\mathbb{E}\left[\mathrm{vech}\left(\mathbf{A}\right) \mathrm{vech}\left(\mathbf{A}\right)^\top\right] = \left.\frac{\partial^2 \phi(\mathbf{Z})}{i^2 \partial \mathrm{vech}\left(\mathbf{Z}\right) \partial \mathrm{vech}\left(\mathbf{Z}\right)^\top}\right|_{\mathbf{Z}=\mathbf{0}} \\
&= \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right) \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right)^\top \\
&\quad + 2 \left(\mathbf{G}^\top \mathbf{G}\right)^{-1} \left( \mathbf{G}\mathbf{G}^+ \left(\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{F}^\top \left(\mathbf{G}^+ \left(\mathbf{I} \otimes \mathbf{I}\right) \mathbf{F}^\top\right)^{-1} \right)^\top \left(\mathbf{G}^+\right)^\top \\
&= \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right) \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right)^\top + 2\mathbf{G}^+ \left( \left(\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \mathbf{G}\mathbf{G}^+ \mathbf{F}^\top \left(\mathbf{G}^+ \mathbf{F}^\top\right)^{-1} \right)^\top \left(\mathbf{G}^+\right)^\top \\
&= \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right) \mathrm{vech}\left(\mathrm{dg}\left(\mathbf{n}\right)\right)^\top + 2\mathbf{G}^+ \left(\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \left(\mathbf{G}^+\right)^\top .
\end{aligned}
$$

and consequently

$$
\mathbf{Cov}\left(\mathrm{vech}\left(\mathbf{A}\right)\right) = 2\mathbf{G}^+ \left(\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I}\right) \left(\mathbf{G}^+\right)^\top . \qquad (259)
$$

Finally

$$
\begin{aligned}
\mathbf{Cov}\left(\operatorname{vech}\left(\mathbf{R}\right)\right) &= \mathbf{Cov}\left(\operatorname{vech}\left(\mathbf{C}_{\boldsymbol{\Omega}}\mathbf{A}\mathbf{C}_{\boldsymbol{\Omega}}^{\top}\right)\right) \\
&= \mathbf{G}^{+}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\mathbf{G}\mathbf{Cov}\left(\left(\operatorname{vech}\left(\mathbf{A}\right)\right)\right)\left(\mathbf{G}^{+}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\mathbf{G}\right)^{\top} \\
&= 2\mathbf{G}^{+}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\mathbf{G}\mathbf{G}^{+}\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\left(\mathbf{G}^{+}\right)^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\mathbf{G}\right)^{\top} \\
&= 2\mathbf{G}^{+}\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^{\top}\otimes\mathbf{C}_{\boldsymbol{\Omega}}^{\top}\right)\left(\mathbf{G}^{+}\right)^{\top} \\
&= 2\mathbf{G}^{+}\left(\boldsymbol{\Sigma}\otimes\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^{\top}\right)\left(\mathbf{G}^{+}\right)^{\top}.
\end{aligned}
$$

*Commented out is the proof that the scaling of the score by the inverse Fisherinfo yields standard GARCH dynamics in the standard GAS model using the Riesz distribution.*

□

### 7.6.2 Wishart Distribution

**Theorem 7.6.** *Let*

$$\mathbf{R} = \frac{1}{n}\mathbf{C}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top\mathbf{C}^\top, \tag{260}$$

*then* $\mathbf{R}$ *is said to follow a standardized Wishart distribution denoted by* $\mathbf{R} \sim \mathcal{W}(\mathbf{\Sigma}, \mathbf{n})$. *The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma}, \mathbf{n}) = \frac{n^{np/2}}{2^{np/2}} \frac{1}{\Gamma_p(n/2)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{\frac{n}{2}} \operatorname{etr}\left(-\frac{1}{2}n\mathbf{Z}\right). \tag{261}$$

*The expected value* $\mathbb{E}[\mathbf{R}]$ *obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma}, \tag{262}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \frac{1}{2}\mathbf{G}^\top \operatorname{vec}\left(n\,\mathbf{\Sigma}^{-1}\mathbf{R}\mathbf{\Sigma}^{-1} - n\,\mathbf{\Sigma}^{-1}\right) \tag{263}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = \frac{n}{2}\mathbf{G}^\top \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G}. \tag{264}$$

*Proof.* Remember that the Wishart distribution is just a special case of the Riesz distribution, with all degree of freedom parameters being equal,

$$n_1, \ldots, n_p = n. \tag{265}$$

Using this, the probability density function[32] is easily obtained from (222). The expected value remains the same as for the Riesz. For the score, start from (233), observe that

32. See also Muirhead 1982, Theorem 3.2.1.

for $n_1, \ldots, n_p = n$, $\mathbf{\Sigma} = n\mathbf{\Omega}$ to arrive at

$$\frac{1}{2}\mathbf{G}^\top \mathrm{vec}\left(-n^2\,\mathbf{\Sigma}^{-1} + n^2\,\mathbf{\Sigma}^{-1}\mathbf{R}\mathbf{\Sigma}^{-1}\right)\frac{1}{n} = \frac{1}{2}\mathbf{G}^\top \mathrm{vec}\left(n\,\mathbf{\Sigma}^{-1}\mathbf{R}\mathbf{\Sigma}^{-1} - n\,\mathbf{\Sigma}^{-1}\right). \tag{266}$$

For the Fisher information matrix start from (247) to arrive at

$$\begin{aligned}
\mathcal{I}_{\mathbf{\Omega}} &= \frac{n}{2}\mathbf{G}^\top\left(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1}\right)\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1} && (267)\\
&= \frac{n}{2}\mathbf{G}^\top\mathbf{G}\mathbf{G}^+\left(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1}\right)\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1} && (268)\\
&= \frac{n}{2}\mathbf{G}^\top\left(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1}\right)\mathbf{G}\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1} && (269)\\
&= \frac{n}{2}\mathbf{G}^\top\left(\mathbf{\Omega}^{-1}\otimes\mathbf{\Omega}^{-1}\right)\mathbf{G}. && (270)
\end{aligned}$$

Then again using $\mathbf{\Sigma} = n\mathbf{\Omega}$ we have

$$\begin{aligned}
\mathcal{I} &= \left(\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top}\right)^\top \mathcal{I}_{\mathbf{\Omega}}\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} && (271)\\
&= \frac{n}{2}\mathbf{G}^\top\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{G}. && (272)
\end{aligned}$$

*Commented out here are the covariance structure and the proof that when scaling the score with the inverse Fisher info we arrive at a standard GARCH structure from the standard GAS dynamics.*

□

### 7.6.3 Inverse Riesz

**Theorem 7.7.** *Let*

$$\mathbf{R} = \mathbf{C}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1/2}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1/2}\mathbf{C}^{\top}, \tag{273}$$

*then* $\mathbf{R}$ *is said to follow a standardized Riesz distribution of type II denoted by* $\mathbf{R} \sim \mathcal{R}^{II}(\mathbf{\Sigma}, \boldsymbol{\nu})$, *with* $\mathbf{m}^{i\mathcal{R}^{II}} = (m_1^{i\mathcal{R}^{II}}, \ldots, m_p^{i\mathcal{R}^{II}})^{\top}$,

$$m_i^{i\mathcal{R}^{II}} = \begin{cases} \frac{1}{\nu_i - p - 1}, & \textit{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(1 + \sum_{j=1}^{j-1} m_j^{i\mathcal{R}^{II}}\right) & \textit{for } i > 1. \end{cases} \tag{274}$$

*The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma}, \boldsymbol{\nu}) = \frac{\prod_{i=1}^{p}\left(m_i^{i\mathcal{R}^{II}}\right)^{-\nu_i/2}}{2^{p\bar{\nu}/2}}\frac{1}{\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\mathbf{Z}^{-1}\right). \tag{275}$$

*The expected value* $\mathbb{E}[\mathbf{R}]$ *obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma} \tag{276}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = -\mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathrm{tril}\left(\mathbf{R}^{-1}\mathbf{C}\,\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} - \mathbf{C}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\right)\right)\mathbf{C}^{-1}\right) \tag{277}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = -\left(\frac{\partial \mathrm{vech}\left(\mathbf{\Omega}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}}\right)^{\top}\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\mathbf{\Sigma}, \boldsymbol{\nu})}{\partial \mathrm{vech}\left(\mathbf{\Omega}\right)\partial \mathrm{vech}\left(\mathbf{\Omega}\right)^{\top}}\right]\frac{\partial \mathrm{vech}\left(\mathbf{\Omega}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} \tag{278}$$

*with*

$$
\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma},\boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right) \partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top}\right] \tag{279}
$$

$$
= -\frac{1}{2}\mathbf{G}^\top \left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2} \otimes \mathbf{C}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\right)\mathbf{F}^\top \left(\mathbf{G}^+\left(\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1/2} \otimes \mathbf{I}\right)\mathbf{F}^\top\right)^{-1}. \tag{280}
$$

*and*

$$
\frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \mathbf{G}^+\left(\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} \otimes \mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}. \tag{281}
$$

*Proof.* For the probability density function of $\mathbf{C}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1/2}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1/2}\mathbf{C}^\top$ see Blasques et al. 2021. According to their, theorem 4 (ii) and definition 6 (ii), $\mathbf{U}^{-\top}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathbf{U}^{-1}$, where $\mathbf{U}$ is the upper Cholesky factor of $\boldsymbol{\Sigma}_{Blasques}^{-1}$ and consequently $\mathbf{U}^{-\top} = \mathbf{C}_{Blasques}$, where $\mathbf{C}_{Blasques}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}_{Blasques}$, follows an inverse Riesz distribution of type II. Now set $\boldsymbol{\Sigma}_{Blasques} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top$ and use their theorem 7 to obtain

$$
f(\mathbf{R}|\boldsymbol{\Omega},\boldsymbol{\nu}) = \frac{1}{2^{p\bar{\boldsymbol{\nu}}/2}\,{}_U\Gamma_p\left(\boldsymbol{\nu}/2\right)}{}_U|\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}|_{-\frac{\boldsymbol{\nu}}{2}}\,{}_U|\mathbf{R}^{-1}|_{\frac{\boldsymbol{\nu}+p+1}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top\mathbf{R}^{-1}\right) \tag{282}
$$

$$
= \frac{1}{2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}|\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top\mathbf{R}^{-1}\right) \tag{283}
$$

$$
= \frac{\prod_{i=1}^p \left(m_i^{i\mathcal{R}^{II}}\right)^{-\nu_i/2}}{2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}|\boldsymbol{\Sigma}|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{Z}^{-1}\right) \tag{284}
$$

$$
= \frac{\prod_{i=1}^p \left(m_i^{i\mathcal{R}^{II}}\right)^{-\nu_i/2}}{2^{p\bar{\boldsymbol{\nu}}/2}}\frac{1}{\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}}\,\mathrm{etr}\left(-\frac{1}{2}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\mathbf{Z}^{-1}\right), \tag{285}
$$

where we used Lemma (7.1) of this paper and Lemma 3 (v) of Blasques et al. 2021. For the expected value see Blasques et al. 2021, Theorem 26, (iv). Now define for better readability $\boldsymbol{\Omega} = \boldsymbol{\Sigma}_{Blasques}$. For the score w.r.t. $\boldsymbol{\Sigma}$ start from (283),

$$\frac{\log p(\mathbf{R}|\boldsymbol{\Sigma},\boldsymbol{\nu})}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \frac{\log p(\mathbf{R}|\boldsymbol{\Sigma},\boldsymbol{\nu})}{\partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top}\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} \tag{286}$$

$$= \left(\frac{\partial \log |\boldsymbol{\Omega}|_{\frac{\nu}{2}}}{\partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top} - \frac{1}{2}\frac{\partial \text{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{\partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top}\right)\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} \tag{287}$$

$$(7.10),(7.12) = \frac{1}{2}\text{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\text{dg}\left(\boldsymbol{\nu}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1} - \mathbf{R}^{-1}\right)^\top \mathbf{G}\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} \tag{288}$$

$$(199) = \text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\text{dg}\left(\boldsymbol{\nu}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{C} - \mathbf{R}^{-1}\mathbf{C}\right)\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G}. \tag{289}$$

Now, using the definition of $\boldsymbol{\Omega}$ and $\mathbf{C}_{\boldsymbol{\Omega}} = \mathbf{C}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1/2}$ this reduces to

$$\text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\text{dg}\left(\boldsymbol{\nu}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{C} - \mathbf{R}^{-1}\mathbf{C}\right)\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{290}$$

$$= \text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\mathbf{C}^{-\top}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2}\text{dg}\left(\boldsymbol{\nu}\right)\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2} - \mathbf{R}^{-1}\mathbf{C}\right)\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{291}$$

$$= \text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\mathbf{C}^{-\top}\text{dg}\left(\boldsymbol{\nu}\right) - \mathbf{R}^{-1}\mathbf{C}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\right)\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{292}$$

$$= -\text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\mathbf{R}^{-1}\mathbf{C}\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} - \mathbf{C}^{-\top}\text{dg}(\boldsymbol{\nu})\right)\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{293}$$

For the Fisher information matrix with respect to $\boldsymbol{\Sigma}$ we use Lemma **??**,

$$\mathcal{I} = -\left(\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top}\right)^\top \mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma},\boldsymbol{\nu})}{\partial \text{vech}\left(\boldsymbol{\Omega}\right)\partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top}\right]\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} \tag{294}$$

where the expectation, using Lemma 7.10 and Lemma 7.12, boils down to

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma},\boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top}\right] && (295)\\
&= \mathbb{E}\left[\frac{\partial^2 \log |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} - \frac{\partial^2 \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{2}}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top}\right] && (296)\\
&= -\frac{1}{2}\mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\otimes\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1} && (297)\\
&= -\frac{1}{2}\mathbf{G}^\top\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2}\otimes\mathbf{C}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1/2}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}. && (298)
\end{aligned}
$$

and

$$
\frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \mathbf{G}^+\left(\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^+\left(\mathbf{C}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}. \tag{299}
$$

□

*Commented out are the inverse Riesz type I distribution with attempts to derive its expected value theoretically.*

### 7.6.4 Inverse Wishart

**Theorem 7.8.** *Let*

$$
\mathbf{R} = (n-p-1)\mathbf{C}\bar{\mathcal{B}}\bar{\mathcal{B}}^\top\mathbf{C}^\top, \tag{300}
$$

*then* $\mathbf{R}$ *is said to follow a standardized inverse Wishart distribution denoted by* $\mathbf{R} \sim i\mathcal{W}(\boldsymbol{\Sigma},\mathbf{n})$. *The probability density function of* $\mathbf{R}$ *is*

$$
p(\mathbf{R}|\boldsymbol{\Sigma},n) = \frac{(n-p-1)^{np/2}}{2^{np/2}}\frac{1}{\Gamma_p(n/2)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|^{-\frac{n}{2}}\,\mathrm{etr}\left(-\frac{1}{2}(n-p-1)\mathbf{Z}^{-1}\right). \tag{301}
$$

*The expected value* $\mathbb{E}[\mathbf{R}]$ *obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma}, \tag{302}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = -\frac{1}{2}\mathbf{G}^\top \mathrm{vec}\left((n-p-1)\mathbf{R}^{-1} - n\mathbf{\Sigma}^{-1}\right) \tag{303}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = -\frac{n}{2}\mathbf{G}^\top \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G}. \tag{304}$$

*Proof.* Remember that the Inverse Wishart distribution is just a special case of the Inverse Riesz distribution, with all degree of freedom parameters being equal,

$$n_1, \ldots, n_p = n. \tag{305}$$

Using this, the probability density function is easily obtained from (275).[33] The expected value remains the same as for the Inverse Riesz type $II$.[34] For the score, start from (288), observe that for $n_1, \ldots, n_p = n$, $\mathbf{\Omega} = (n-p-1)\mathbf{\Sigma}$ to arrive at

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma}, n)}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} = \frac{1}{2}\mathrm{vec}\left(\frac{n}{n-p-1}\mathbf{\Sigma}^{-1} - \mathbf{R}^{-1}\right)^\top \mathbf{G}\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \tag{306}$$

$$= \frac{1}{2}\mathrm{vec}\left(n\mathbf{\Sigma}^{-1} - (n-p-1)\mathbf{R}^{-1}\right)^\top \mathbf{G} \tag{307}$$

33. See also Kollo and Rosen 2005, Corollary 2.4.6.1.
34. See also Kollo and Rosen 2005, Corollary 2.4.14.

For the Fisher information matrix simply start from the equation above to obtain

$$\frac{\partial^2 \log p(\mathbf{R}|\mathbf{\Sigma}, n)}{\partial \text{vech}\left(\mathbf{\Sigma}\right) \partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} = \frac{n}{2} \frac{\partial \mathbf{G}^\top \text{vec}\left(\mathbf{\Sigma}^{-1}\right)}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} \tag{308}$$

$$= \frac{n}{2} \mathbf{G}^\top \mathbf{G} \frac{\partial \text{vech}\left(\mathbf{\Sigma}^{-1}\right)^\top}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} \tag{309}$$

$$\text{Lemma 7.9:} = -\frac{n}{2} \mathbf{G}^\top \mathbf{G} \mathbf{G}^+ \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G} \tag{310}$$

$$= -\frac{n}{2} \mathbf{G}^\top \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G}. \tag{311}$$

$$\tag{312}$$

*Commented out are the derivation of the dispersion matrix and the score scaled by the inverse Fisher Information Matrix and by the Dispersion Matrix (which do not reduce to simple GARCH dynamics as for the Riesz and Wishart).*

□

### 7.6.5 t-Riesz

**Theorem 7.9.** *Let*

$$\mathbf{R} = \frac{\nu-2}{\nu}\Gamma^{-1}_{\left(\frac{\nu}{2},\frac{2}{\nu}\right)}\mathbf{C}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top, \tag{313}$$

*then* $\mathbf{R}$ *is said to follow a standardized t-Riesz distribution of type I denoted by* $\mathbf{R} \sim t\mathcal{R}^I(\mathbf{\Sigma}, \mathbf{n})$. *The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n}) = \frac{\prod_{i=1}^p n_i^{n_i/2}}{(\nu-2)^{p\bar{\mathbf{n}}/2}}\frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}\left(1+\frac{1}{\nu-2}\mathrm{tr}\left(\mathrm{dg}(\mathbf{n})\mathbf{Z}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}. \tag{314}$$

*The expected value* $\mathbb{E}\left[\mathbf{R}\right]$ *obtains as*

$$\mathbb{E}\left[\mathbf{R}\right] = \mathbf{\Sigma} \tag{315}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \mathbf{G}^\top\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\mathrm{tril}\left(\frac{\nu+p\bar{\mathbf{n}}}{\nu-2+\mathrm{tr}(\mathrm{dg}(\mathbf{n})\mathbf{Z})}\mathbf{C}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{Z}-\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\right)\right)\mathbf{C}^{-1}\right) \tag{316}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = -\left(\frac{\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)}{\partial\mathrm{vech}\left(\mathbf{\Sigma}\right)^\top}\right)^\top\mathbb{E}\left[\frac{\partial^2\log p(\mathbf{R}|\mathbf{\Omega},\mathbf{n})}{\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)^\top}\right]\frac{\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)}{\partial\mathrm{vech}\left(\mathbf{\Sigma}\right)^\top}, \tag{317}$$

*with*

$$\begin{aligned}\mathbb{E}\left[\frac{\partial^2\log p(\mathbf{R}|\mathbf{\Omega},\mathbf{n})}{\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)\partial\mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)^\top}\right] &= -\frac{1}{2}\left(\frac{\nu-2}{\nu}\right)^2\mathbf{G}^\top\left(\mathbf{C}\,\mathrm{dg}(\mathbf{n})^{1/2}\otimes\mathbf{I}\right)\mathbf{F}^\top \\ &\times\left(\mathbf{G}^\top\left(\mathbf{C}\,\mathrm{dg}(\mathbf{n})^{-1/2}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^\top\left(\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\otimes\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)\mathbf{G} \\ &+\frac{1}{2}\left(\frac{\nu-2}{\nu}\right)^2\frac{1}{\nu+p\bar{\mathbf{n}}+2}\mathbf{G}^\top\left(2\left(\mathbf{\Sigma}\otimes\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)+\mathrm{vec}\left(\mathbf{\Sigma}\right)\mathrm{vec}\left(\mathbf{\Sigma}\right)^\top\right)\mathbf{G}\end{aligned} \tag{318}$$

*and*

$$\frac{\partial \text{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} = -\frac{\nu}{\nu-2}\mathbf{G}^{+}\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{F}^{\top}$$
$$\times\left(\mathbf{G}^{+}\left(\mathbf{C}^{-\top}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{F}^{\top}\right)^{-1}\mathbf{G}^{+}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G}. \tag{319}$$

*Proof.* Define $\boldsymbol{\Omega} = (\nu-2)/\nu\,\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^{\top}$, $\mathbf{A} = \mathbf{C}_{\boldsymbol{\Omega}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathbf{C}_{\boldsymbol{\Omega}}^{\top}$, which implies that $\mathbf{A} \sim \mathcal{R}^{I}(\boldsymbol{\Omega},\mathbf{n})$, and let $w \sim \Gamma_{\left(\frac{\nu}{2},\frac{2}{\nu}\right)}$. Then we can rewrite $\mathbf{R} = w^{-1}\mathbf{A}$ with probability density function

$$p_{t\mathcal{R}^{I}}(\mathbf{R}|\boldsymbol{\Omega},\mathbf{n},\nu) = \int_{0}^{\infty} p_{\Gamma}\left(w\middle|\frac{\nu}{2},\frac{2}{\nu}\right)p_{\mathcal{R}^{I}}(w\mathbf{R}|\boldsymbol{\Omega},\mathbf{n})J(w\mathbf{R}\to\mathbf{R})dw$$
$$= \int_{0}^{\infty}\frac{1}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{2}{\nu}\right)^{\frac{\nu}{2}}}w^{\frac{\nu}{2}-1}e^{-w\frac{\nu}{2}}\frac{|w\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}e^{-\frac{\text{tr}\left(\boldsymbol{\Omega}^{-1}w\mathbf{R}\right)}{2}}}{|\boldsymbol{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_{p}\left(\frac{\mathbf{n}}{2}\right)2^{\frac{p\bar{\mathbf{n}}}{2}}}w^{\frac{p(p+1)}{2}}dw$$
$$= \frac{|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\int_{0}^{\infty}w^{\frac{\nu}{2}-1-p(p+1)/2+p\bar{\mathbf{n}}/2+p(p+1)/2}\exp\left[-w\frac{\nu}{2}\left(1+\frac{\text{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)\right]dw}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{2}{\nu}\right)^{\frac{\nu}{2}}|\boldsymbol{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_{p}\left(\frac{\mathbf{n}}{2}\right)2^{\frac{p\bar{\mathbf{n}}}{2}}}$$
$$= \frac{|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\left(\frac{2}{\nu}\right)^{\frac{\nu}{2}}|\boldsymbol{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_{p}(\frac{\mathbf{n}}{2})2^{\frac{p\bar{\mathbf{n}}}{2}}}\int_{0}^{\infty}w^{(\nu+p\bar{\mathbf{n}})/2-1}\exp\left[-w\frac{\nu}{2}\left(1+\frac{\text{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right)\right]dw$$
$$= \frac{|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}}{\Gamma(\frac{\nu}{2})(\frac{2}{\nu})^{\frac{\nu}{2}}|\boldsymbol{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_{p}(\frac{\mathbf{n}}{2})2^{\frac{p\bar{\mathbf{n}}}{2}}}\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)\left[\frac{\nu}{2}\left(1+\frac{\text{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right)\right]^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}$$
$$= \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma(\frac{\nu}{2})\Gamma_{p}(\frac{\mathbf{n}}{2})\nu^{p\bar{\mathbf{n}}/2}}|\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(1+\frac{\text{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}},$$

where we used https://dlmf.nist.gov/5.9#i. Now substituting $\boldsymbol{\Omega}$ with its definition we get the standardized probability density function

$$
\begin{aligned}
&p_{t\mathcal{R}^I}(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\nu)\\
&\quad= \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma(\frac{\nu}{2})\Gamma_p(\frac{\mathbf{n}}{2})\nu^{p\bar{\mathbf{n}}/2}} \left|\frac{\nu-2}{\nu}\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right|_{-\frac{\mathbf{n}}{2}} |\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} \left(1+\frac{\mathrm{tr}\left(\left(\frac{\nu-2}{\nu}\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\right)}{\nu}\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}\\
&\quad= \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma(\frac{\nu}{2})\Gamma_p(\frac{\mathbf{n}}{2})\nu^{p\bar{\mathbf{n}}/2}} \left(\frac{\nu-2}{\nu}\right)^{-p\bar{\mathbf{n}}/2} \left|\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right|_{-\frac{\mathbf{n}}{2}} |\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} \left(1+\frac{\mathrm{tr}\left(\left(\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\right)}{\nu-2}\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}\\
&\quad= \frac{\prod_{i=1}^p n_i^{n_i/2}}{(\nu-2)^{p\bar{\mathbf{n}}/2}} \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|_{\frac{\mathbf{n}}{2}} \left(1+\frac{1}{\nu-2}\mathrm{tr}\left(\mathrm{dg}(\mathbf{n})\mathbf{Z}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}.
\end{aligned}
$$

For the expected value we have

$$
\begin{aligned}
\mathbb{E}[\mathbf{R}] &= \mathbb{E}\left[\frac{1}{w}\right]\mathbb{E}\left[\mathbf{A}\right] = \mathbb{E}\left[\frac{1}{w}\right]\mathbf{C}_{\boldsymbol{\Omega}}\,\mathbb{E}\left[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right]\,\mathbf{C}_{\boldsymbol{\Omega}}^\top && (320)\\
&= \frac{\nu}{\nu-2}\mathbf{C}_{\boldsymbol{\Omega}}\,\mathrm{dg}(\mathbf{n})\,\mathbf{C}_{\boldsymbol{\Omega}}^\top && (321)\\
&= \frac{\nu}{\nu-2}\frac{\nu-2}{\nu}\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1/2}\,\mathrm{dg}(\mathbf{n})\,\mathrm{dg}(\mathbf{n})^{-1/2}\mathbf{C}^\top && (322)\\
&= \boldsymbol{\Sigma}, && (323)
\end{aligned}
$$

where $\nu/(\nu-2)$ is the expectation of the inverse Gamma distribution.

For the score w.r.t. $\boldsymbol{\Sigma}$ consider again

$$
\frac{\log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\nu)}{\partial\mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \frac{\log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\nu)}{\partial\mathrm{vech}\left(\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^\top}\frac{\partial\mathrm{vech}\left(\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)}{\partial\mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top}, \qquad (324)
$$

where starting from equation (320) we have

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n},\nu)}{\partial \text{vech}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^\top}$$

$$= \frac{\partial \log|\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top|_{-\frac{\mathbf{n}}{2}}}{\partial \text{vech}(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top)^\top} - \frac{\nu+p\bar{\mathbf{n}}}{2}\frac{\partial \log\left(1+\frac{\text{tr}\left(\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\right)}{\nu-2}\right)}{\partial \text{vech}(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top)^\top}$$

$$= -\frac{1}{2}\text{vec}\left(\mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)^{1/2}\text{dg}\left(\mathbf{n}\right)\text{dg}\left(\mathbf{n}\right)^{1/2}\mathbf{C}^{-1}\right)^\top\mathbf{G}$$

$$-\frac{\nu+p\bar{\mathbf{n}}}{2}\frac{1}{\left(1+\frac{\text{tr}\left(\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\right)}{\nu-2}\right)}\frac{\partial\left(1+\frac{\text{tr}\left(\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\right)}{\nu-2}\right)}{\partial \text{vech}(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top)^\top}$$

$$= \text{vec}\left(-\frac{1}{2}\mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)^2\mathbf{C}^{-1} + \frac{\nu+p\bar{\mathbf{n}}}{2}\frac{1}{\left(1+\frac{\text{tr}(\text{dg}(\mathbf{n})\mathbf{Z})}{\nu-2}\right)}\frac{1}{\nu-2}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\mathbf{R}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^{-1}\right)^\top\mathbf{G}$$

$$= \text{vec}\left(-\frac{1}{2}\mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)^2\mathbf{C}^{-1} + \frac{\nu+p\bar{\mathbf{n}}}{2}\frac{1}{\left(1+\frac{\text{tr}(\text{dg}(\mathbf{n})\mathbf{Z})}{\nu-2}\right)}\frac{1}{\nu-2}\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{Z}\text{dg}(\mathbf{n})\mathbf{C}^{-1}\right)^\top\mathbf{G}$$

$$= \frac{1}{2}\text{vec}\left(\frac{\nu+p\bar{\mathbf{n}}}{(\nu-2+\text{tr}(\text{dg}(\mathbf{n})\mathbf{Z}))}\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{Z}\text{dg}(\mathbf{n})\mathbf{C}^{-1} - \mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)^2\mathbf{C}^{-1}\right)^\top\mathbf{G}.$$

Now, using (199) we have

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n},\nu)}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top} = \frac{\log p(\mathbf{R}|\mathbf{\Sigma},\mathbf{n})}{\partial \text{vech}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)^\top}\frac{\partial \text{vech}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top}$$

$$= \frac{1}{2}\text{vec}\left(\frac{\nu+p\bar{\mathbf{n}}}{(\nu-2+\text{tr}(\text{dg}(\mathbf{n})\mathbf{Z}))}\mathbf{C}^{-\top}\text{dg}(\mathbf{n})\mathbf{Z}\text{dg}(\mathbf{n})\mathbf{C}^{-1} - \mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)^2\mathbf{C}^{-1}\right)^\top\mathbf{G}$$

$$\times\frac{\partial \text{vech}\left(\mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)}{\partial \text{vech}\left(\mathbf{\Sigma}\right)^\top}$$

$$= \mathbf{G}^\top\text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\left(\frac{\nu+p\bar{\mathbf{n}}}{\nu-2+\text{tr}(\text{dg}(\mathbf{n})\mathbf{Z})}\mathbf{C}^{-\top}\text{dg}\left(\mathbf{n}\right)\mathbf{Z} - \mathbf{C}^{-\top}\text{dg}(\mathbf{n})\right)\right)\mathbf{C}^{-1}\right).$$

For the Fisher information matrix w.r.t. $\mathbf{\Sigma}$ we will again use $\mathbf{\Omega}$ as defined above and note that

$$\frac{\partial^2 \log\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} = \frac{\partial}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\left[\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-1}\frac{\partial \frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)}\right] \tag{325}$$

$$= -\frac{1}{\nu^2}\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-2}\frac{\partial \mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)}\frac{\partial \mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} \tag{326}$$

$$= -\frac{1}{\nu^2}\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-2}\mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}\mathbf{G}. \tag{327}$$

We will need the expectation of this expression. For better overview denote the normalizing constant of the $t$-Riesz distribution by

$$c(\nu,\mathbf{n},\boldsymbol{\Omega}) = \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma(\nu/2)\Gamma_p(\mathbf{n}/2)\nu^{p\bar{\mathbf{n}}/2}}|\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}. \tag{328}$$

Then

$$\mathbb{E}\left[\left(1+\frac{\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-2}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}\right] \tag{329}$$

$$= c(\nu,\mathbf{n},\mathbf{\Omega})\int_{\mathbf{R}>\mathbf{0}}\left(1+\frac{\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-2}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(1+\frac{\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}\mathrm{d}\mathbf{R} \tag{330}$$

$$= c(\nu,\mathbf{n},\mathbf{\Omega})\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(1+\frac{\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-\frac{\nu+p\bar{\mathbf{n}}+4}{2}}\mathrm{d}\mathbf{R} \tag{331}$$

$$= c(\nu,\mathbf{n},\mathbf{\Omega})\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(1+\frac{1}{\nu+4}\mathrm{tr}\left(\left(\frac{\nu}{\nu+4}\mathbf{\Omega}\right)^{-1}\mathbf{R}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}+4}{2}}\mathrm{d}\mathbf{R} \tag{332}$$

$$\begin{aligned}&= \frac{c(\nu,\mathbf{n},\mathbf{\Omega})}{c\left(\nu+4,\mathbf{n},\frac{\nu}{\nu+4}\mathbf{\Omega}\right)}\\ &\times \underbrace{c\left(\nu+4,\mathbf{n},\frac{\nu}{\nu+4}\mathbf{\Omega}\right)\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^{\top}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left(1+\frac{1}{\nu+4}\mathrm{tr}\left(\left(\frac{\nu}{\nu+4}\mathbf{\Omega}\right)^{-1}\mathbf{R}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}+4}{2}}\mathrm{d}\mathbf{R}}_{\mathbb{E}\left[\mathrm{vec}(\mathbf{R})\mathrm{vec}(\mathbf{R})^{\top}\right]\text{ if }\mathbf{R}\text{ follows a }t\text{-Riesz distribution with parameters }\nu+4,\ \mathbf{n}\text{ and }\mathbf{\Omega}_{new}=\frac{\nu}{\nu+4}\mathbf{\Omega}}\end{aligned} \tag{333}$$

$$= \frac{c(\nu,\mathbf{n},\mathbf{\Omega})}{c\left(\nu+4,\mathbf{n},\frac{\nu}{\nu+4}\mathbf{\Omega}\right)}\frac{(\nu+4)^2}{(\nu+2)\nu}\left(\frac{\nu}{\nu+4}\right)^2 \tag{334}$$

$$\times\,\mathbf{G}\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^{\top}\right)\left(\mathbf{C}_{\mathbf{\Omega}}^{\top}\otimes\mathbf{C}_{\mathbf{\Omega}}^{\top}\right)\left(\mathbf{G}\mathbf{G}^{+}\right)^{\top} \tag{335}$$

$$= \frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p\left(\mathbf{n}/2\right)\Gamma(\nu/2)}\frac{\Gamma_p\left(\mathbf{n}/2\right)\Gamma((\nu+4)/2)}{\Gamma(((\nu+4)+p\bar{\mathbf{n}})/2)}\frac{(\nu+4)^{p\bar{\mathbf{n}}/2}}{\nu^{p\bar{\mathbf{n}}/2}}\left(\frac{\nu+4}{\nu}\right)^{-p\bar{\mathbf{n}}/2}\frac{|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}}{|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}}\frac{\nu}{\nu+2}$$

$$\times\,\mathbf{G}\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^{\top}\right)\left(\mathbf{C}_{\mathbf{\Omega}}^{\top}\otimes\mathbf{C}_{\mathbf{\Omega}}^{\top}\right)\left(\mathbf{G}\mathbf{G}^{+}\right)^{\top} \tag{336}$$

$$= \frac{\nu\left(\nu+2\right)}{\left(\nu+p\bar{\mathbf{n}}\right)\left(\nu+p\bar{\mathbf{n}}+2\right)}\frac{\nu}{\nu+2} \tag{337}$$

$$\times\,\mathbf{G}\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^{\top}\right)\left(\mathbf{C}_{\mathbf{\Omega}}^{\top}\otimes\mathbf{C}_{\mathbf{\Omega}}^{\top}\right)\left(\mathbf{G}\mathbf{G}^{+}\right)^{\top} \tag{338}$$

$$= \frac{\nu^2}{\left(\nu+p\bar{\mathbf{n}}\right)\left(\nu+p\bar{\mathbf{n}}+2\right)} \tag{339}$$

$$\times\,\mathbf{G}\mathbf{G}^{+}\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^{\top}\right)\left(\mathbf{C}_{\mathbf{\Omega}}^{\top}\otimes\mathbf{C}_{\mathbf{\Omega}}^{\top}\right)\left(\mathbf{G}\mathbf{G}^{+}\right)^{\top} \tag{340}$$

(341)

with

$$
\begin{aligned}
&\mathbf{Cov}\left(\mathrm{vech}\left(\mathbf{R}\right)\right) && (342)\\
&= \mathbb{E}\left[\frac{1}{w^2}\right]\mathbf{Cov}\left(\mathrm{vech}(\mathbf{A})\right) + \mathbf{Var}\left(\frac{1}{w}\right)\mathbb{E}\left[\mathrm{vech}\left(\mathbf{A}\right)\right]\mathbb{E}\left[\mathrm{vech}\left(\mathbf{A}\right)^\top\right] && (343)\\
&= 2\frac{\nu^2}{(\nu-2)(\nu-4)}\mathbf{G}^+\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^\top\otimes\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top && (344)\\
&\quad + 2\frac{\nu^2}{(\nu-4)(\nu-2)^2}\mathbf{G}^+\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}^\top\otimes\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top, && (345)
\end{aligned}
$$

where

$$
\mathbf{Var}\left(\frac{1}{w}\right) = \frac{(\nu/2)^2}{(\nu/2-1)^2(\nu/2-2)} = 2\frac{\nu^2}{(\nu-2)^2(\nu-4)} \tag{346}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\frac{1}{w^2}\right] &= \mathbf{Var}\left(\frac{1}{w}\right) + \mathbb{E}\left[\frac{1}{w^2}\right] = \left(\frac{\nu}{\nu-2}\right)^2 + 2\frac{\nu^2}{(\nu-2)^2(\nu-4)} && (347)\\
&= \left(\frac{\nu}{\nu-2}\right)^2\left(1+\frac{2}{\nu-4}\right) = \left(\frac{\nu}{\nu-2}\right)^2\frac{\nu-2}{\nu-4} = \frac{\nu^2}{(\nu-2)(\nu-4)}. && (348)
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathbb{E}\left[\operatorname{vech}(\mathbf{R})\operatorname{vech}(\mathbf{R})^\top\right] \\
&= \mathbf{Cov}\left(\operatorname{vech}(\mathbf{R})\right) + \mathbb{E}\left[\operatorname{vech}(\mathbf{R})\right]\mathbb{E}\left[\operatorname{vech}(\mathbf{R})^\top\right] \\
&= 2\frac{\nu^2}{(\nu-2)(\nu-4)}\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\left(\operatorname{dg}(\mathbf{n})\otimes\mathbf{I}\right)\left(\mathbf{C}_{\mathbf{\Omega}}^\top\otimes\mathbf{C}_{\mathbf{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top \\
&\quad + 2\frac{\nu^2}{(\nu-4)(\nu-2)^2}\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)^\top\left(\mathbf{C}_{\mathbf{\Omega}}^\top\otimes\mathbf{C}_{\mathbf{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top \\
&\quad + 2\left(\frac{\nu}{\nu-2}\right)^2\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)^\top\left(\mathbf{C}_{\mathbf{\Omega}}^\top\otimes\mathbf{C}_{\mathbf{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top \\
&= \frac{\nu^2}{(\nu-2)(\nu-4)}\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right) \\
&\quad \times\left(2\left(\operatorname{dg}(\mathbf{n})\otimes\mathbf{I}\right) + \operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)^\top\right)\left(\mathbf{C}_{\mathbf{\Omega}}^\top\otimes\mathbf{C}_{\mathbf{\Omega}}^\top\right)\left(\mathbf{G}^+\right)^\top.
\end{aligned}
$$

and consequently

$$
\begin{aligned}
&\mathbb{E}\left[\operatorname{vec}(\mathbf{R})\operatorname{vec}(\mathbf{R})^\top\right] \\
&= \mathbf{G}\mathbb{E}\left[\operatorname{vech}(\mathbf{R})\operatorname{vech}(\mathbf{R})^\top\right]\mathbf{G}^\top \\
&= \frac{\nu^2}{(\nu-2)(\nu-4)}\mathbf{G}\mathbf{G}^+\left(\mathbf{C}_{\mathbf{\Omega}}\otimes\mathbf{C}_{\mathbf{\Omega}}\right) \\
&\quad \times\left(2\left(\operatorname{dg}(\mathbf{n})\otimes\mathbf{I}\right) + \operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)\operatorname{vec}\left(\operatorname{dg}(\mathbf{n})\right)^\top\right)\left(\mathbf{C}_{\mathbf{\Omega}}^\top\otimes\mathbf{C}_{\mathbf{\Omega}}^\top\right)\left(\mathbf{G}\mathbf{G}^+\right)^\top.
\end{aligned}
$$

Using the results above we have

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Omega},\mathbf{n})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)^\top}\right]$$

$$= \frac{\partial^2 \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})^\top} - \frac{\nu + p\bar{\mathbf{n}}}{2}\mathbb{E}\left[\frac{\partial^2 \log\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)}{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})^\top}\right] \tag{349}$$

$$= \frac{\partial^2 \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})^\top}$$

$$+ \frac{\nu + p\bar{\mathbf{n}}}{2}\frac{1}{\nu^2}\mathbf{G}^\top\mathbb{E}\left[\left(1+\frac{\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}{\nu}\right)^{-2}\mathrm{vec}\left(\mathbf{R}\right)\mathrm{vec}\left(\mathbf{R}\right)^\top\right]\mathbf{G} \tag{350}$$

$$= \frac{\partial^2 \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})^\top} + \frac{\nu + p\bar{\mathbf{n}}}{2}\frac{1}{\nu^2}\frac{\nu^2}{(\nu+p\bar{\mathbf{n}})(\nu+p\bar{\mathbf{n}}+2)}$$

$$\times \mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^\top\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^\top\otimes\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)\mathbf{G} \tag{351}$$

$$= \frac{\partial^2 \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1})^\top} + \frac{1}{2}\frac{1}{\nu+p\bar{\mathbf{n}}+2}$$

$$\times \mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{C}_{\boldsymbol{\Omega}}\right)\left(2\left(\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)+\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)\mathrm{vec}\left(\mathrm{dg}(\mathbf{n})\right)^\top\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^\top\otimes\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)\mathbf{G} \tag{352}$$

$$= -\frac{1}{2}\mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}\,\mathrm{dg}(\mathbf{n})\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}\mathbf{G}^\top\left(\boldsymbol{\Omega}\otimes\boldsymbol{\Omega}\right)\mathbf{G} + \frac{1}{2}\frac{1}{\nu+p\bar{\mathbf{n}}+2}$$

$$\times \mathbf{G}^\top\left(2\left(\mathbf{C}_{\boldsymbol{\Omega}}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\boldsymbol{\Omega}}^\top\otimes\boldsymbol{\Omega}\right)+\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\boldsymbol{\Omega}}^\top\right)^\top\right)\mathbf{G} \tag{353}$$

$$= -\frac{1}{2}\left(\frac{\nu-2}{\nu}\right)^2\mathbf{G}^\top\left(\mathbf{C}\,\mathrm{dg}(\mathbf{n})^{1/2}\otimes\mathbf{I}\right)\mathbf{F}^\top\left(\mathbf{G}^\top\left(\mathbf{C}\,\mathrm{dg}(\mathbf{n})^{-1/2}\otimes\mathbf{I}\right)\mathbf{F}^\top\right)^{-1}$$

$$\times \mathbf{G}^\top\left(\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\otimes\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)\mathbf{G}$$

$$+ \frac{1}{2}\left(\frac{\nu-2}{\nu}\right)^2\frac{1}{\nu+p\bar{\mathbf{n}}+2}\mathbf{G}^\top\left(2\left(\boldsymbol{\Sigma}\otimes\mathbf{C}\mathrm{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)+\mathrm{vec}\left(\boldsymbol{\Sigma}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}\right)^\top\right)\mathbf{G}, \tag{354}$$

such that using Lemma (**??**) we have

$$\mathcal{I} = -\left(\frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top}\right)^\top\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Omega},\mathbf{n})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)^\top}\right]\frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top}$$

with

$$\frac{\partial \mathrm{vech}\left(\mathbf{\Omega}^{-1}\right)}{\partial \mathrm{vech}\left(\mathbf{\Sigma}\right)^{\top}} = -\frac{\nu}{\nu-2}\mathbf{G}^{+}\left(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{+}\left(\mathbf{C}^{-\top}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{F}^{\top}\right)^{-1} \times \mathbf{G}^{+}\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{G}. \tag{355}$$

Furthermore

$$\begin{aligned}
\frac{\partial \log p}{\partial \mathbf{n}^{\top}} &= \frac{\log\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)-\log\Gamma_p(\frac{\mathbf{n}}{2})-p\bar{\mathbf{n}}/2\log\nu}{\partial \mathbf{n}^{\top}} \\
&+ \frac{\log|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}+\log|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}-\frac{\nu+p\bar{\mathbf{n}}}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right)}{\partial \mathbf{n}^{\top}} \\
&= \frac{1}{2}\psi((\nu+p\bar{\mathbf{n}})/2)-\frac{1}{2}\boldsymbol{\psi}(\mathbf{n}/2)-\frac{1}{2}\log\nu-\log\mathrm{vecd}\left(\mathbf{C}_{\mathbf{\Omega}}\right)+\log\mathrm{vecd}\left(\mathbf{C}_{\mathbf{R}}\right) \\
&\quad -\frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) \\
&= \frac{1}{2}\left(\psi((\nu+p\bar{\mathbf{n}})/2)-\boldsymbol{\psi}(\mathbf{n}/2)-\log\left(\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)\right) \\
&\quad -\log\mathrm{vecd}\left(\mathbf{C}_{\mathbf{\Omega}}\right)+\log\mathrm{vecd}\left(\mathbf{C}_{\mathbf{R}}\right)
\end{aligned}$$

$$\frac{\partial^2 \log p}{\partial \mathbf{n}\partial \mathbf{n}^{\top}} = \frac{1}{4}\mathrm{dg}\left(\psi'((\nu+p\bar{\mathbf{n}})/2)-\boldsymbol{\psi}'(\mathbf{n}/2)\right)$$

$$\frac{\partial^2 \log p}{\partial \mathbf{n}\partial \nu} = \left(\frac{1}{4}\psi'((\nu+p\bar{\mathbf{n}})/2)-\left(\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)^{-1}\right)\odot\mathbf{1}$$

$$
\begin{aligned}
\frac{\partial \log p}{\partial \nu} &= \frac{\log\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right) - \log\Gamma(\frac{\nu}{2}) - p\bar{\mathbf{n}}/2\log\nu - \frac{\nu+p\bar{\mathbf{n}}}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right)}{\partial\nu} \\
&= \frac{1}{2}\psi((\nu+p\bar{\mathbf{n}})/2) - \frac{1}{2}\psi(\nu/2) \\
&\quad - \frac{p\bar{\mathbf{n}}}{2\nu} - \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) + \frac{\nu+p\bar{\mathbf{n}}}{2}\frac{1}{1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}}\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu^2} \\
&= \frac{1}{2}\left(\psi((\nu+p\bar{\mathbf{n}})/2) - \psi(\nu/2)\right) \\
&\quad - \frac{p\bar{\mathbf{n}}}{2\nu} - \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) + \left(\frac{1}{2}+\frac{p\bar{\mathbf{n}}}{2\nu}\right)\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)} \\
&= \frac{1}{2}\left(\psi((\nu+p\bar{\mathbf{n}})/2) - \psi(\nu/2)\right) \\
&\quad - \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) + \frac{1}{2}\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)} + \frac{p\bar{\mathbf{n}}}{2\nu}\left(\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)} - 1\right) \\
&= \frac{1}{2}\left(\psi((\nu+p\bar{\mathbf{n}})/2) - \psi(\nu/2)\right) \\
&\quad - \frac{1}{2}\log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) + \frac{1}{2}\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)} - \frac{p\bar{\mathbf{n}}}{2\nu}\left(\frac{\nu}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}\right) \\
&= \frac{1}{2}\left(\psi((\nu+p\bar{\mathbf{n}})/2) - \psi(\nu/2) - \log\left(1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}\right) - \frac{p\bar{\mathbf{n}}-\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \log p}{\partial\nu\partial\nu} &= \frac{1}{2}\left(\frac{1}{2}\psi'((\nu+p\bar{\mathbf{n}})/2) - \frac{1}{2}\psi('\nu/2) + \frac{\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu^2}}{1+\frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\nu}} + \frac{p\bar{\mathbf{n}}-\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\left(\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)^2}\right) \\
&= \frac{1}{2}\left(\frac{1}{2}\psi'((\nu+p\bar{\mathbf{n}})/2) - \frac{1}{2}\psi('\nu/2) + \frac{\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)/\nu}{\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)} + \frac{p\bar{\mathbf{n}}-\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)}{\left(\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)^2}\right)
\end{aligned}
$$

$$
\frac{\partial^2 \log p}{\partial\nu\partial\mathbf{n}^\top} = \left(\frac{1}{4}\psi'((\nu+p\bar{\mathbf{n}})/2) - \frac{1}{2}\left(\nu+\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)^{-1}\right)\odot\mathbf{1}^\top
$$

□

### 7.6.6 t-Wishart

**Theorem 7.10.** *Let*

$$\mathbf{R} = \frac{\nu - 2}{\nu n} \Gamma^{-1}_{\left(\frac{\nu}{2}, \frac{2}{\nu}\right)} \underset{\sim}{\mathcal{B}} \underset{\sim}{\mathcal{B}}^\top, \tag{356}$$

*then* $\mathbf{R}$ *is said to follow a standardized Wishart distribution based on the t-model (Sutradhar and Ali 1989), which we simply call t-Wishart distribution and denote by* $\mathbf{R} \sim t\mathcal{W}(\mathbf{\Sigma}, n, \nu)$. *The probability density function of* $\mathbf{R}$ *is*

$$f(\mathbf{R}|\mathbf{\Sigma}, n, \nu) = \left(\frac{n}{\nu - 2}\right)^{pn/2} \frac{\Gamma((\nu + pn)/2)}{\Gamma_p(n/2)\,\Gamma(\nu/2)} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{\frac{n}{2}} \left(1 + \frac{n}{\nu - 2} \mathrm{tr}\,(\mathbf{Z})\right)^{-\frac{\nu + pn}{2}}. \tag{357}$$

*The expected value* $\mathbb{E}[\mathbf{R}]$ *obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma}, \tag{358}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \frac{1}{2} \mathbf{G}^\top \mathrm{vec}\left(n \frac{\nu + pn}{\nu - 2 + n\mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{R})} \mathbf{\Sigma}^{-1}\mathbf{R}\mathbf{\Sigma}^{-1} - n\mathbf{\Sigma}^{-1}\right), \tag{359}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = \frac{n}{2} \mathbf{G}^\top \left(\frac{\nu + pn}{\nu + pn + 2} \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) - \frac{n}{\nu + pn + 2} \mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right) \mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right)^\top\right) \mathbf{G}. \tag{360}$$

*Proof.* We base the proof on the standardized t-Riesz $I$ distribution by setting $n_i = n$ for all $i$. The probability density function follows easily. The expected value is the same as for the t-Riesz $I$ distribution since its independent of $\mathbf{n}$. For the score w.r.t. $\mathbf{\Sigma}$ start from equation (325) to get

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma}, n, \nu)}{\partial \text{vech}\,(\mathbf{\Sigma})^\top} = \frac{1}{2}\text{vec}\left(\frac{\nu + pn}{\nu - 2 + n\text{tr}(\mathbf{Z})} n^2 \mathbf{C}^{-\top}\mathbf{Z}\mathbf{C}^{-1} - n^2\mathbf{\Sigma}^{-1}\right)^\top \mathbf{G} n^{-1} \tag{361}$$

$$= \frac{1}{2}\text{vec}\left(\frac{\nu + pn}{\nu - 2 + n\text{tr}(\mathbf{\Sigma}^{-1}\mathbf{R})} n\mathbf{\Sigma}^{-1}\mathbf{R}\mathbf{\Sigma}^{-1} - n\mathbf{\Sigma}^{-1}\right)^\top \mathbf{G}. \tag{362}$$

For the Fisher information matrix w.r.t. $\mathbf{\Sigma}$ notice that if $n_i = n$ for all $i$ equation (355) reduces to

$$\mathcal{I} = -\left(\frac{n\nu}{\nu - 2}\right)^2 \left(\frac{\partial \text{vech}\left(\mathbf{\Sigma}^{-1}\right)}{\partial \text{vech}\,(\mathbf{\Sigma})^\top}\right)^\top \mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\mathbf{\Omega}, \mathbf{n})}{\partial \text{vech}\left(\mathbf{\Omega}^{-1}\right) \partial \text{vech}\left(\mathbf{\Omega}^{-1}\right)^\top}\right] \frac{\partial \text{vech}\left(\mathbf{\Sigma}^{-1}\right)}{\partial \text{vech}\,(\mathbf{\Sigma})^\top} \tag{363}$$

with according to equation (354)

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\mathbf{\Omega}, \mathbf{n})}{\partial \text{vech}\left(\mathbf{\Omega}^{-1}\right) \partial \text{vech}\left(\mathbf{\Omega}^{-1}\right)^\top}\right] \tag{364}$$

$$= -\frac{1}{2}\left(\frac{\nu - 2}{\nu}\right)^2 n^{-1}\mathbf{G}^\top (\mathbf{C} \otimes \mathbf{I}) \mathbf{F}^\top \left(\mathbf{G}^\top (\mathbf{C} \otimes \mathbf{I}) \mathbf{F}^\top\right)^{-1} \mathbf{G}^\top (\mathbf{\Sigma} \otimes \mathbf{\Sigma}) \mathbf{G}$$

$$+ \frac{1}{2}\left(\frac{\nu - 2}{\nu}\right)^2 \frac{1}{\nu + pn + 2}\mathbf{G}^\top \left(\frac{2}{n}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \text{vec}\,(\mathbf{\Sigma})\,\text{vec}\,(\mathbf{\Sigma})^\top\right) \mathbf{G} \tag{365}$$

$$= \frac{1}{2}\left(\frac{\nu - 2}{\nu}\right)^2 \mathbf{G}^\top \left(-n^{-1}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \frac{1}{\nu + pn + 2}\left(\frac{2}{n}(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + \text{vec}\,(\mathbf{\Sigma})\,\text{vec}\,(\mathbf{\Sigma})^\top\right)\right) \mathbf{G} \tag{366}$$

and

$$\frac{\partial \text{vech}\left(\mathbf{\Sigma}^{-1}\right)}{\partial \text{vech}\,(\mathbf{\Sigma})^\top} = -\mathbf{G}^+ \left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) \mathbf{G}, \tag{367}$$

such that altogether

$$\mathcal{I} = -\left(\frac{n\nu}{\nu-2}\right)^2 \left(-\mathbf{G}^+ \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{G}\right)^\top \tag{368}$$

$$\times \frac{1}{2}\left(\frac{\nu-2}{\nu}\right)^2 \mathbf{G}^\top \left(-n^{-1}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right) + \frac{1}{\nu+pn+2}\left(\frac{2}{n}\left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}\right) + \operatorname{vec}\left(\boldsymbol{\Sigma}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}\right)^\top\right)\right)\mathbf{G} \tag{369}$$

$$\times \left(-\mathbf{G}^+ \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{G}\right) \tag{370}$$

$$= -\frac{1}{2}n^2 \mathbf{G}^\top \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \left(\mathbf{G}^+\right)^\top \mathbf{G}^\top \tag{371}$$

$$\times \left(-n^{-1}\left(\mathbf{I} \otimes \mathbf{I}\right) + \frac{1}{\nu+pn+2}\left(\frac{2}{n}\left(\mathbf{I} \otimes \mathbf{I}\right) + \operatorname{vec}\left(\boldsymbol{\Sigma}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\right)\mathbf{G} \tag{372}$$

$$= -\frac{1}{2}\mathbf{G}^\top \left(-n\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + \frac{n^2}{\nu+pn+2}\left(\frac{2}{n}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\right)\mathbf{G} \tag{373}$$

$$= -\frac{n}{2}\mathbf{G}^\top \left(-\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + \frac{2}{\nu+pn+2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + \frac{n}{\nu+pn+2}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G} \tag{374}$$

$$= \frac{n}{2}\mathbf{G}^\top \left(\frac{\nu+pn}{\nu+pn+2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{n}{\nu+pn+2}\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}. \tag{375}$$

*Commented out are the stand-alone proofs, i.e. not based on the t-Riesz distribution.*

□

### 7.6.7 Inverse t-Riesz

**Theorem 7.11.** *Let*

$$\mathbf{R} = \Gamma_{\left(\frac{n}{2},\frac{2}{n}\right)}\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\mathbf{C}^{\top}. \tag{376}$$

*Then* $\mathbf{R}$ *is said to follow a standardized inverse t-Riesz distribution of type* $II$*, denoted by* $\mathbf{R} \sim it\mathcal{R}^{II}(\boldsymbol{\Sigma}, \mathbf{n}, \nu)$ *with* $\mathbf{m}^{i\mathcal{R}^{II}}$ *given in (274). The probability density function of* $\mathbf{R}$ *is*

$$\begin{aligned} p(\mathbf{R}|\boldsymbol{\Sigma}, n, \boldsymbol{\nu}) = & \frac{\prod_{i=1}^{p}(m_i^{i\mathcal{R}^{II}})^{-\nu_i/2}}{n^{p\bar{\boldsymbol{\nu}}/2}}\frac{\Gamma\left((n+p\bar{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma(n/2)}\left|\mathbf{R}\right|^{-\frac{p+1}{2}}\left|\mathbf{Z}\right|_{-\frac{\boldsymbol{\nu}}{2}} \\ & \times\left(1+\frac{1}{n}\mathrm{tr}\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{Z}^{-1}\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}}. \end{aligned} \tag{377}$$

*The expected value* $\mathbb{E}\left[\mathbf{R}\right]$ *obtains as*

$$\mathbb{E}\left[\mathbf{R}\right] = \boldsymbol{\Sigma}, \tag{378}$$

*the score w.r.t.* $\boldsymbol{\Sigma}$ *is*

$$\nabla = \mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathbf{X}\right)\mathbf{C}^{-1}\right) \tag{379}$$

*with*

$$\mathbf{X} = \mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}\left(\boldsymbol{\nu}\right) - \frac{n+p\bar{\boldsymbol{\nu}}}{n+\ \mathrm{tr}(\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\mathbf{Z}^{-1})}\mathbf{R}^{-1}\mathbf{C}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\right).$$

*Proof.* For the probability density function, note that in the stochastic representation $\mathbf{A} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\mathbf{C}^{\top} \sim i\mathcal{R}^{II}\left(\boldsymbol{\Sigma}, \boldsymbol{\nu}\right)$ and denote $w \sim \Gamma_{\left(\frac{n}{2},\frac{2}{n}\right)}$ such that $\mathbf{R} = w\mathbf{A}$. Denote for better readability $\boldsymbol{\Omega} = \mathbf{C}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\mathbf{C}^\top$. Then

$$
\begin{aligned}
&p_{it\mathcal{R}^{II}}(\mathbf{R}|\boldsymbol{\Sigma}, n, \boldsymbol{\nu}) = \int_0^\infty p_\Gamma\left(w \middle| \frac{n}{2}, \frac{2}{n}\right) p_{i\mathcal{R}^{II}}(w^{-1}\mathbf{R}|\boldsymbol{\Sigma}, \boldsymbol{\nu}) J(w^{-1}\mathbf{R} \to \mathbf{R}) dw \\
&= \int_0^\infty \frac{1}{\Gamma\left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^{\frac{n}{2}}} w^{\frac{n}{2}-1} e^{-w\frac{n}{2}} \\
&\qquad \times \frac{1}{2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |w^{-1}\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}} \mathrm{etr}\left(-\frac{1}{2}w\boldsymbol{\Omega}\mathbf{R}^{-1}\right) \\
&\qquad \times w^{-\frac{p(p+1)}{2}} dw \\
&= \frac{|\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}}}{\Gamma\left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^{\frac{n}{2}} 2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} \int_0^\infty w^{\frac{n+p\bar{\boldsymbol{\nu}}}{2}-1} \exp\left(-w\frac{n}{2}\left(1 + \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right)\right) dw \\
&= \frac{|\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}}}{\Gamma\left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^{\frac{n}{2}} 2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} \Gamma\left(\frac{n+p\bar{\boldsymbol{\nu}}}{2}\right) \left(\frac{n}{2}\left(1 + \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}} \\
&= \frac{1}{n^{p\bar{\boldsymbol{\nu}}/2}} \frac{\Gamma\left(\frac{n+p\bar{\boldsymbol{\nu}}}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}+p+1}{2}} \left(1 + \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}} \\
&= \frac{\prod_{i=1}^p (m_i^{i\mathcal{R}^{II}})^{-\nu_i/2}}{n^{p\bar{\boldsymbol{\nu}}/2}|\mathbf{R}|^{\frac{p+1}{2}}} \frac{\Gamma\left((n+p\bar{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma(n/2)} |\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}} \left(1 + \frac{1}{n}\mathrm{tr}\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{Z}^{-1}\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}},
\end{aligned}
$$

where we used https://dlmf.nist.gov/5.9#i. For the expected value note that when $w$ follows a gamma distribution with $\gamma = \nu/2$ and $\beta = 2/\nu$, then $\mathbb{E}[w] = 1$ (Thom 1958). Furthermore recall that $\mathbb{E}[\mathbf{A}] = \boldsymbol{\Sigma}$, such that

$$
\mathbb{E}[\mathbf{R}] = \mathbb{E}[w]\,\mathbb{E}[\mathbf{A}] = \boldsymbol{\Sigma}.
$$

For the score we have

$$
\begin{aligned}
&\frac{\partial \log p_{it\mathcal{R}^{II}}(\mathbf{R}|\boldsymbol{\Sigma}, n, \boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} = \frac{\partial \log |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)} - \frac{n+p\bar{\boldsymbol{\nu}}}{2} \frac{1/n}{1 + 1/n\,\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})} \frac{\partial \mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)} \\
&= \mathrm{vec}\left(\frac{1}{2}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\boldsymbol{\nu}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1} - \frac{n+p\bar{\boldsymbol{\nu}}}{2}\frac{1}{n +\ \mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}\mathbf{R}^{-1}\right)^\top \mathbf{G} \\
&= \frac{1}{2}\mathrm{vec}\left(\mathbf{C}^{-\top}\mathrm{dg}\left(\boldsymbol{\nu} \odot \mathbf{m}^{i\mathcal{R}^{II}}\right)\mathbf{C}^{-1} - \frac{n+p\bar{\boldsymbol{\nu}}}{n +\ \mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}\mathbf{R}^{-1}\right)^\top \mathbf{G}, \qquad (380)
\end{aligned}
$$

such that using Lemma 7.16 we have

$$\frac{\partial \log p_{it\mathcal{R}^{II}}(\mathbf{R}|\boldsymbol{\Sigma}, n, \boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} = \frac{\partial \log p_{it\mathcal{R}^{II}}(\mathbf{R}|\boldsymbol{\Sigma}, n, \boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} \frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^\top} \tag{381}$$

$$= \mathrm{vec}\left(\mathbf{C}^{-\top} \Phi\left(\mathbf{C}^\top \mathbf{X}\right) \mathbf{C}^{-1}\right)^\top \mathbf{G}$$

with

$$\mathbf{X} = \mathrm{tril}\left(\left(\mathbf{C}^{-\top} \mathrm{dg}\left(\boldsymbol{\nu} \odot \mathbf{m}^{i\mathcal{R}^{II}}\right) \mathbf{C}^{-1} - \frac{n + p\bar{\boldsymbol{\nu}}}{n + \mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})} \mathbf{R}^{-1}\right) \mathbf{C}\right) \mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}$$

$$= \mathrm{tril}\left(\mathbf{C}^{-\top} \mathrm{dg}\left(\boldsymbol{\nu}\right) - \frac{n + p\bar{\boldsymbol{\nu}}}{n + \mathrm{tr}(\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1} \mathbf{Z}^{-1})} \mathbf{R}^{-1} \mathbf{C} \mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-1}\right).$$

For the Fisher information matrix, proceed similarly as for the inverse t-Wishart distribution, where first $\mathbb{E}\left[\mathrm{vech}\left(\mathbf{R}\right) \mathrm{vech}\left(\mathbf{R}\right)^\top\right]$ for $\mathbf{R} \sim t\mathcal{R}^{II}$ has to be derived similarly to the expectation in $\mathbf{R} \sim t\mathcal{R}^{I}$.

Furthermore

$$\frac{\partial \log p}{\partial n} = \frac{\log \Gamma\left((n + p\bar{\boldsymbol{\nu}})/2\right) - \log \Gamma(n/2) - p\bar{\boldsymbol{\nu}}/2 \log n}{\partial n}$$

$$- \frac{\frac{n+p\bar{\boldsymbol{\nu}}}{2} \log\left(1 + \frac{\mathrm{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{n}\right)}{\partial \boldsymbol{\nu}}$$

$$= \frac{1}{2}\psi((n + p\bar{\boldsymbol{\nu}})/2) - \frac{1}{2}\psi(n/2) - \frac{p\bar{\boldsymbol{\nu}}}{2n} - \frac{1}{2}\log\left(1 + \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right)$$

$$+ \frac{n + p\bar{\boldsymbol{\nu}}}{2} \frac{\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)/n^2}{1 + \mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)/n}$$

and

$$
\begin{aligned}
\frac{\partial \log p}{\partial \boldsymbol{\nu}} &= \frac{\log \Gamma\left((n+p\bar{\boldsymbol{\nu}})/2\right) - \log \Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2) - p\bar{\boldsymbol{\nu}}/2 \log n}{\partial \boldsymbol{\nu}} \\
&\quad + \frac{\log |\boldsymbol{\Omega}|_{\frac{\nu}{2}} + \log |\mathbf{R}|_{-\frac{\nu+p+1}{2}} - \frac{n+p\bar{\nu}}{2} \log\left(1 + \frac{\operatorname{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right)}{\partial n} \\
&= \frac{1}{2}\psi((n+p\bar{\boldsymbol{\nu}})/2) - \frac{1}{2}\overleftarrow{\boldsymbol{\psi}(\overleftarrow{\boldsymbol{\nu}}/2)} - \frac{1}{2}\log n + \log \operatorname{vecd}\left(\mathbf{C}_{\boldsymbol{\Omega}}\right) - \log \operatorname{vecd}\left(\mathbf{C}_{\mathbf{R}}\right) \\
&\quad - \frac{1}{2}\log\left(1 + \frac{\operatorname{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)}{n}\right) \\
&= \frac{1}{2}\left(\psi((n+p\bar{\boldsymbol{\nu}})/2) - \overleftarrow{\boldsymbol{\psi}(\overleftarrow{\boldsymbol{\nu}}/2)} - \log\left(n + \operatorname{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)\right)\right) \\
&\quad + \log \operatorname{vecd}\left(\mathbf{C}_{\boldsymbol{\Omega}}\right) - \log \operatorname{vecd}\left(\mathbf{C}_{\mathbf{R}}\right)
\end{aligned}
$$

$\square$

### 7.6.8 Inverse t-Wishart

**Theorem 7.12.** *Let*

$$\mathbf{R} = (\nu - p - 1)\Gamma_{\left(\frac{n}{2}, \frac{2}{n}\right)}\mathbf{C}\bar{\mathcal{B}}^{-\top}\bar{\mathcal{B}}^{-1}\mathbf{C}^{\top}. \tag{382}$$

*Then* $\mathbf{R}$ *is said to follow a standardized inverse t-Wishart distribution denoted by* $\mathbf{R} \sim it\mathcal{W}(\mathbf{\Sigma}, n, \nu)$*. Denote* $\mathbf{\Omega} = (\nu - p - 1)\mathbf{\Sigma}$*. The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma}, n, \nu) = \left(\frac{\nu - p - 1}{n}\right)^{\nu p/2} \frac{\Gamma((n + p\nu)/2)}{\Gamma_p\left(\nu/2\right)\Gamma(n/2)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{-\frac{\nu}{2}} \left(1 + \frac{\nu - p - 1}{n}\mathrm{tr}(\mathbf{Z}^{-1})\right)^{-\frac{n+p\nu}{2}} \tag{383}$$

$$= \frac{\Gamma((n + p\nu)/2)}{n^{\nu p/2}\Gamma_p\left(\nu/2\right)\Gamma(n/2)}|\mathbf{\Omega}|^{\frac{\nu}{2}}|\mathbf{R}|^{-\frac{\nu+p+1}{2}} \left(1 + \frac{\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})}{n}\right)^{-\frac{n+p\nu}{2}}, \tag{384}$$

*the expected value* $\mathbb{E}\left[\mathbf{R}\right]$ *obtains as*

$$\mathbb{E}\left[\mathbf{R}\right] = \mathbf{\Sigma} = \frac{1}{\nu - p - 1}\mathbf{\Omega}, \tag{385}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \frac{1}{2}\mathbf{G}^{\top}\mathrm{vec}\left(\nu\mathbf{\Sigma}^{-1} - \frac{(n + p\nu)(\nu - p - 1)}{n + (\nu - p - 1)\mathrm{tr}(\mathbf{\Sigma}\mathbf{R}^{-1})}\mathbf{R}^{-1}\right) \tag{386}$$

*and the Fisher Information Matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = -\frac{\nu}{2}\mathbf{G}^{\top}\left(\frac{n + p\nu}{n + p\nu + 2}\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right) - \frac{\nu}{(n + p\nu + 2)}\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right)\mathrm{vec}\left(\mathbf{\Sigma}^{-1}\right)^{\top}\right)\mathbf{G}. \tag{387}$$

*Proof.* We base the proof on the standardized inverse t-Riesz $II$ distribution by setting $\nu_i = \nu$ for all $i$. The probability density function and the expected value follow easily.

For the score w.r.t. $\boldsymbol{\Sigma}$ use equations (380) and (381), and

$$\frac{\partial \text{vech}\left(\boldsymbol{\Omega}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} = (\nu - p - 1)\mathbf{I} \tag{388}$$

to get

$$\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \text{vec}\left(\frac{1}{2}\frac{\nu}{\nu - p - 1}\boldsymbol{\Sigma}^{-1} - \frac{n + p\nu}{2}\frac{1}{n + \text{ tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}\mathbf{R}^{-1}\right)^\top \mathbf{G}(\nu - p - 1) \tag{389}$$

$$= \frac{1}{2}\text{vec}\left(\nu\boldsymbol{\Sigma}^{-1} - \frac{(n + p\nu)(\nu - p - 1)}{n + (\nu - p - 1)\text{tr}(\boldsymbol{\Sigma}\mathbf{R}^{-1})}\mathbf{R}^{-1}\right)^\top \mathbf{G}. \tag{390}$$

The proof of the Fisher information matrix w.r.t $\boldsymbol{\Sigma}$ follows closely the proof for the Fisher information matrix if the t-Wishart distribution. We have

$$\mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu)}{\partial \text{vech}\left(\boldsymbol{\Omega}\right) \partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top}\right] \tag{391}$$

$$= \frac{\nu}{2}\frac{\partial \log |\boldsymbol{\Omega}|}{\partial \text{vech}\left(\boldsymbol{\Omega}\right) \partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top} - \frac{n + p\nu}{2}\mathbb{E}\left[\frac{\partial^2 \log\left(1 + \frac{1}{n}\text{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)}{\partial \text{vech}\left(\boldsymbol{\Omega}\right) \partial \text{vech}\left(\boldsymbol{\Omega}\right)^\top}\right] \tag{392}$$

$$\begin{aligned}&= -\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right)\mathbf{G}\\ &\quad + \frac{n + p\nu}{2}\left(\frac{1}{n}\right)^2 \mathbf{G}^\top\mathbb{E}\left[\left(1 + \frac{1}{n}\text{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-2} \text{vec}\left(\mathbf{R}^{-1}\right)\text{vec}\left(\mathbf{R}^{-1}\right)^\top\right]\mathbf{G}\end{aligned} \tag{393}$$

$$\begin{aligned}&= -\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right)\mathbf{G}\\ &\quad + \frac{n + p\nu}{2}\left(\frac{1}{n}\right)^2 \frac{n^2\nu^2}{(n + p\nu)\left(n + p\nu + 2\right)}\frac{1}{(1)^2}\end{aligned} \tag{394}$$

$$\quad \times \mathbf{G}^\top\left(\text{vec}\left(\boldsymbol{\Omega}^{-1}\right)\text{vec}\left(\boldsymbol{\Omega}^{-1}\right)^\top + \frac{2}{\nu}\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right)\right)\mathbf{G} \tag{395}$$

$$= -\frac{\nu}{2}\mathbf{G}^\top\left(\frac{n + p\nu}{n + p\nu + 2}\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right) - \frac{\nu}{(n + p\nu + 2)}\text{vec}\left(\boldsymbol{\Omega}^{-1}\right)\text{vec}\left(\boldsymbol{\Omega}^{-1}\right)^\top\right)\mathbf{G} \tag{396}$$

$$= -\frac{\nu}{2(n - p - 1)^2}\mathbf{G}^\top\left(\frac{n + p\nu}{n + p\nu + 2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{\nu}{(n + p\nu + 2)}\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\text{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}. \tag{397}$$

where

$$\begin{aligned}
\frac{\partial^2 \log\left(1+\frac{1}{n}\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})\right)}{\partial \mathrm{vech}\,(\mathbf{\Omega})\,\partial \mathrm{vech}\,(\mathbf{\Omega})^\top} &= \frac{\partial}{\partial \mathrm{vech}\,(\mathbf{\Omega})^\top}\left[\left(1+\frac{1}{n}\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})\right)^{-1}\frac{\partial \frac{1}{n}\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})}{\partial \mathrm{vech}\,(\mathbf{\Omega})}\right] \\
&= -\left(\frac{1}{n}\right)^2\left(1+\frac{1}{n}\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})\right)^{-2}\frac{\partial \mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})}{\partial \mathrm{vech}\,(\mathbf{\Omega})}\frac{\partial \mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})}{\partial \mathrm{vech}\,(\mathbf{\Omega})^\top} \\
&= -\left(\frac{1}{n}\right)^2\left(1+\frac{1}{n}\mathrm{tr}(\mathbf{\Omega}\mathbf{R}^{-1})\right)^{-2}\mathbf{G}^\top \mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^\top\mathbf{G},
\end{aligned}$$

and, denoting

$$c(n,\nu,\mathbf{\Omega}) = \frac{\Gamma((n+p\nu)/2)}{n^{p\nu/2}\Gamma_p\left(\nu/2\right)\Gamma(n/2)}|\mathbf{\Omega}|^{\frac{\nu}{2}},$$

$$
\begin{aligned}
&\mathbb{E}\left[\left(1+\frac{1}{n}\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-2}\mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^{\top}\right]\\
&=c(n,\nu,\boldsymbol{\Omega})\int_{\mathbf{R}>\mathbf{0}}\left(1+\frac{1}{n}\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-2}\mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^{\top}|\mathbf{R}|^{-\frac{\nu+p+1}{2}}\\
&\quad\times\left(1+\frac{1}{n}\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-\frac{n+p\nu}{2}}\mathrm{d}\mathbf{R}\\
&=c(n,\nu,\boldsymbol{\Omega})\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^{\top}|\mathbf{R}|^{-\frac{\nu+p+1}{2}}\left(1+\frac{1}{n}\mathrm{tr}\left(\boldsymbol{\Omega}\mathbf{R}^{-1}\right)\right)^{-\frac{n+p\nu+4}{2}}\mathrm{d}\mathbf{R}\\
&=c(n,\nu,\boldsymbol{\Omega})\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^{\top}|\mathbf{R}|^{-\frac{\nu+p+1}{2}}\\
&\quad\times\left(1+\frac{1}{n+4}\mathrm{tr}\left(\frac{n+4}{n}\boldsymbol{\Omega}\mathbf{R}^{-1}\right)\right)^{-\frac{n+p\nu+4}{2}}\mathrm{d}\mathbf{R}\\
&=\frac{c(n,\nu,\boldsymbol{\Omega})}{c\left(n+4,\nu,\frac{n+4}{n}\boldsymbol{\Omega}\right)}\\
&\quad\times c\left(n+4,\nu,\frac{n+4}{n}\boldsymbol{\Omega}\right)\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}\left(\mathbf{R}^{-1}\right)\mathrm{vec}\left(\mathbf{R}^{-1}\right)^{\top}|\mathbf{R}|^{-\frac{\nu+p+1}{2}}\\
&\quad\times\left(1+\frac{1}{n+4}\mathrm{tr}\left(\frac{n+4}{n}\boldsymbol{\Omega}\mathbf{R}^{-1}\right)\right)^{-\frac{n+p\nu+4}{2}}\mathrm{d}\mathbf{R}\\
&=\frac{c(n,\nu,\boldsymbol{\Omega})}{c\left(n+4,\nu,\frac{n+4}{n}\boldsymbol{\Omega}\right)}\frac{n\nu^2}{n+2}\left(\frac{1}{\nu}\left(\mathbf{I}_{p^2}+\mathbf{K}_{pp}\right)\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)+\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}\right)\\
&=\frac{\Gamma((n+p\nu)/2)}{n^{p\nu/2}\Gamma_p\left(\nu/2\right)\Gamma(n/2)}|\boldsymbol{\Omega}^{-1}|^{\frac{\nu}{2}}\frac{(n+4)^{p\nu/2}\Gamma_p\left(\nu/2\right)\Gamma((n+4)/2)}{\Gamma((n+4+p\nu)/2)}\left|\frac{n}{n+4}\boldsymbol{\Omega}^{-1}\right|^{\frac{\nu}{2}}\frac{n\nu^2}{n+2}\\
&\qquad\times\left(\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}+\frac{1}{\nu}\left(\mathbf{I}_{p^2}+\mathbf{K}_{pp}\right)\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)\right)\text{See iphone photo 22.06.2021}\\
&=\frac{\Gamma((n+p\nu)/2)}{\Gamma((n+4+p\nu)/2)}\frac{\Gamma((n+4)/2)}{\Gamma(n/2)}\frac{n\nu^2}{n+2}\\
&\quad\times\left(\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}+\frac{1}{\nu}\left(\mathbf{I}_{p^2}+\mathbf{K}_{pp}\right)\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)\right)\\
&=\frac{n\left(n+2\right)}{\left(n+p\nu\right)\left(n+p\nu+2\right)}\frac{n\nu^2}{n+2}\left(\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}+\frac{1}{\nu}\left(\mathbf{I}_{p^2}+\mathbf{K}_{pp}\right)\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)\right)\\
&=\frac{n^2\nu^2}{\left(n+p\nu\right)\left(n+p\nu+2\right)}\frac{1}{(1)^2}\left(\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}+\frac{1}{\nu}\left(\mathbf{I}_{p^2}+\mathbf{K}_{pp}\right)\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)\right).
\end{aligned}
$$

Thus, finally

$$
\begin{aligned}
\mathcal{I} &= \left(\frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}}\right)^{\top} \mathbb{E}\left[\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma}, n, n)}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right) \partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^{\top}}\right] \frac{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} \\
&= -\frac{\nu(n-p-1)^2}{2(n-p-1)^2} \\
&\quad \times \mathbf{G}^{\top}\left(\frac{n+p\nu}{n+p\nu+2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{\nu}{(n+p\nu+2)} \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}\right) \mathbf{G} \\
&= -\frac{\nu}{2} \mathbf{G}^{\top}\left(\frac{n+p\nu}{n+p\nu+2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - \frac{\nu}{(n+p\nu+2)} \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right) \mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^{\top}\right) \mathbf{G}.
\end{aligned}
$$

*Commented out is the covariance matrix and the standalone derivation of score and expected value.*

□

#### 7.6.9 F-Riesz

**Theorem 7.13.** *Let*

$$\mathbf{R} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-\frac{1}{2}}\mathbf{C}^{\top}, \tag{398}$$

[35] *then* $\mathbf{R}$ *is said to follow a standardized F-Riesz distribution of type I denoted by* $\mathbf{R} \sim \mathcal{FR}^{I}(\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})$, *with* $\mathbf{m}^{\mathcal{FR}'} = (m_1^{\mathcal{FR}'}, m_2^{\mathcal{FR}'}, \ldots, m_p^{\mathcal{FR}'})^{\top}$,

$$m_i^{\mathcal{FR}'} = \begin{cases} \frac{n_1}{\nu_1 - p - 1}, & \text{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(n_i + \sum_{i=1}^{i-1} m_i^{\mathcal{FR}'}\right) & \text{for } i = 2, \ldots, p. \end{cases} \tag{399}$$

*Denote* $\boldsymbol{\Omega} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-1}\mathbf{C}^{\top}$. *The probability density function of* $\mathbf{R}$ *is*

$$p\left(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu}\right) = \prod_{i=1}^{p}(m_i^{\mathcal{FR}'})^{n_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}} + \overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\left|\mathbf{R}\right|^{-\frac{p+1}{2}}\left|\mathbf{Z}\right|_{\frac{\mathbf{n}}{2}}\left|\mathbf{I}_p + \mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{1/2}\mathbf{Z}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{1/2}\right|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}} \tag{400}$$

[36]

*The expected value* $\mathbb{E}\left[\mathbf{R}\right]$ *obtains as*

$$\mathbb{E}\left[\mathbf{R}\right] = \boldsymbol{\Sigma} = \mathbf{C}_{\boldsymbol{\Omega}}\,\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})\mathbf{C}_{\boldsymbol{\Omega}}^{\top}, \tag{401}$$

*and the score w.r.t.* $\boldsymbol{\Sigma}$ *is*

$$\nabla = \mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\boldsymbol{\nu}) - \mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-\top}\mathrm{dg}(\boldsymbol{\nu}+\mathbf{n})\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-1}\mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-1}\right)\right)\mathbf{C}^{-1}\right). \tag{402}$$

*Proof.* Note that our stochastic representation is equivalent to the generation of the

35. See photo 29.03.2021 for stochastic representation.
36. Note that $\left|\mathbf{I}_p + \boldsymbol{\Omega}^{-1}\mathbf{R}\right|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$ can not be calculated using the Cholesky-based calculation of the generalized power function, since $\mathbf{I}_p + \boldsymbol{\Omega}^{-1}\mathbf{R}$ is not symmetric positive definite.

F-Riesz $I$ distribution in Blasques et al. 2021 with $\mathbf{\Sigma}_{Blasques} = \mathbf{\Omega}$. For the probability density function we thus take their derived probability density function and rewrite

$$
\begin{aligned}
p(\mathbf{R}|\mathbf{\Sigma}, \mathbf{n}, \boldsymbol{\nu}) &= \frac{\Gamma_p^U\left((\mathbf{n}+\boldsymbol{\nu})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p^U(\boldsymbol{\nu}/2)}|\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}|\mathbf{\Omega}+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}} \qquad (403)\\
&= \frac{\Gamma_p^U\left((\mathbf{n}+\boldsymbol{\nu})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p^U(\boldsymbol{\nu}/2)}|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}\left|\mathbf{I}_p+\mathbf{\Omega}^{-1}\mathbf{R}\right|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}\\
&= \prod_{i=1}^p (\mathbf{m}_i^{\mathcal{FR}^I})^{-\nu_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{\Sigma}|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}|\mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-1}\mathbf{C}^\top+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}\\
&= \prod_{i=1}^p (\mathbf{m}_i^{\mathcal{FR}^I})^{-\nu_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{\Sigma}|_{\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}|\mathbf{R}|^{-(p-1)/2}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}|\mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-1}\mathbf{C}^\top+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}\\
&= \prod_{i=1}^p (\mathbf{m}_i^{\mathcal{FR}^I})^{-\nu_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{R}|^{-(p-1)/2}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}|\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-1}+\mathbf{Z}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}\\
&= \prod_{i=1}^p (\mathbf{m}_i^{\mathcal{FR}^I})^{-\nu_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})|_{\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}|\mathbf{R}|^{-(p-1)/2}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}\\
&\quad \times |\mathbf{I}+\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{1/2}\mathbf{Z}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{1/2}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}\\
&= \prod_{i=1}^p (\mathbf{m}^{\mathcal{FR}^I})_i^{n_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{R}|^{-(p-1)/2}|\mathbf{Z}|_{\frac{\mathbf{n}}{2}}|\mathbf{I}+\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{1/2}\mathbf{Z}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{1/2}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}},
\end{aligned}
$$

where we have used Lemma 7.5 and Blasques et al. 2021 Lemma 3 (v). For the expected value Blasques et al. 2021 derive in Theorem 10, that

$$
\mathbb{E}\left[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\right] = \mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I}),
$$

such that

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{R}\right] &= \mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\mathbb{E}\left[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\right]\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\mathbf{C}^\top\\
&= \mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})\mathrm{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\mathbf{C}^\top = \mathbf{\Sigma}.
\end{aligned}
$$

For the score w.r.t. $\boldsymbol{\Sigma}$ we have

$$
\begin{aligned}
\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\boldsymbol{\nu})}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} &= \frac{\partial \log |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} + \frac{\partial \log |\boldsymbol{\Omega}+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} \\
&= \frac{\partial \log |\boldsymbol{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} + \frac{\partial \log |\boldsymbol{\Omega}+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega}+\mathbf{R})^\top} \frac{\partial \mathrm{vech}(\boldsymbol{\Omega}+\mathbf{R})}{\partial \mathrm{vech}\,(\boldsymbol{\Omega})^\top} \\
&= \frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu})\,\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} - \frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu}+\mathbf{n})\,\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-1}\right)^\top \mathbf{G} \\
&= \frac{1}{2}\mathrm{vec}\left(\mathbf{C}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu}\otimes\mathbf{m}^{\mathcal{FR}'})\,\mathbf{C}^{-1} - \mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu}+\mathbf{n})\,\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-1}\right)^\top \mathbf{G},
\end{aligned} \tag{404}
$$

such that using Lemma 7.16

$$
\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\boldsymbol{\nu})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^\top} \tag{405}
$$

$$
= \mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\mathrm{tril}(\left(\mathbf{C}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu}\otimes\mathbf{m}^{\mathcal{FR}'})\,\mathbf{C}^{-1} - \mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-\top}\mathrm{dg}\,(\boldsymbol{\nu}+\mathbf{n})\,\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-1}\right)\mathbf{C})\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-1}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{406}
$$

$$
= \mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\boldsymbol{\nu}) - \mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-\top}\mathrm{dg}(\boldsymbol{\nu}+\mathbf{n})\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}}^{-1}\mathbf{C}\mathrm{dg}(\mathbf{m}^{\mathcal{FR}'})^{-1}\right)\right)\mathbf{C}^{-1}\right)^\top \mathbf{G}. \tag{407}
$$

For the score w.r.t $\mathbf{n}$ we have, using Lemma 7.5, equation (403), Lemma 7.3 and Lemma 7.6,

$$
\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\boldsymbol{\nu})}{\partial \mathbf{n}^\top} \tag{408}
$$

$$
= \frac{\log\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right) - \log\Gamma_p(\mathbf{n}/2) + \log|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} + \log|\boldsymbol{\Omega}+\mathbf{R}|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}}{\partial \mathbf{n}^\top} \tag{409}
$$

$$
= \frac{1}{2}\overleftarrow{\boldsymbol{\psi}\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}^\top - \frac{1}{2}\boldsymbol{\psi}\,(\mathbf{n}/2)^\top + \log\mathrm{vecd}(\mathbf{C}_{\mathbf{R}})^\top - \log\mathrm{vecd}(\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}})^\top \tag{410}
$$

$$
\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\boldsymbol{\nu})}{\partial \mathbf{n}\partial \mathbf{n}^\top} = \frac{1}{4}\mathrm{dg}\left(\overleftarrow{\boldsymbol{\psi}'\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)} - \boldsymbol{\psi}'\,(\mathbf{n}/2)\right) \tag{411}
$$

$$\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}^\top} = \frac{1}{2}\overleftarrow{\boldsymbol{\psi}\left((\overleftarrow{\mathbf{n}} + \overleftarrow{\boldsymbol{\nu}})/2\right)}^\top - \frac{1}{2}\overleftarrow{\boldsymbol{\psi}\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}^\top - \log \operatorname{vecd}(\mathbf{C}_{\boldsymbol{\Omega}})^\top - \log \operatorname{vecd}(\mathbf{C}_{\boldsymbol{\Omega}+\mathbf{R}})^\top \tag{412}$$

$$\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu} \partial \boldsymbol{\nu}^\top} = \frac{1}{4}\operatorname{dg}\left(\overleftarrow{\boldsymbol{\psi}'\left((\overleftarrow{\mathbf{n}} + \overleftarrow{\boldsymbol{\nu}})/2\right)} - \overleftarrow{\boldsymbol{\psi}'\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}\right) \tag{413}$$

$\square$

Note that it does not hold that $\mathbf{R} \sim \mathcal{FR}(\boldsymbol{\Omega}^{-1}, \mathbf{n}, \boldsymbol{\nu}) \Rightarrow \mathbf{R}^{-1} \sim \mathcal{FR}(\boldsymbol{\Omega}, \boldsymbol{\nu}, \mathbf{n})$ for either type.[37] This is in contrast to the F distribution. Also note that the standardized F-Riesz type $I$ distribution cannot be obtained by mixing a standardized Riesz type $I$ with a standardized inverse Riesz type $II$, but only by mixing the non-standardized versions and then standardizing the resulting distribution, as done above. This is also in contrast to the F distribution and can be seen, since

$$\begin{aligned}
&\mathbf{C}\operatorname{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\operatorname{dg}(\mathbf{m}^{\mathcal{FR}^I})^{-\frac{1}{2}}\mathbf{C}^\top && (414)\\
&\neq \mathbf{C}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-1}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\mathbf{C}^\top. && (415)
\end{aligned}$$

*Commented out are the Hessian with some propositions for approximation of the Fisher information matrix, the derivation of the F-Riesz based on the Riesz type $I$ and inverse Riesz type $II$, the derivation of the inverse F-Riesz type $I$ distribution and an attempt to derive the distribution of* $\mathbf{C}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-1}\operatorname{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\mathbf{C}^\top$.

37. See the derivation of the inverse F-Riesz type $II$ below. The derivation of the inverse F-Riesz type $I$ is very similar.

### 7.6.10 Inverse F-Riesz

**Theorem 7.14.** *Let*

$$\mathbf{R} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-\frac{1}{2}}\mathbf{C}^{\top}, \tag{416}$$

*then $\mathbf{R}$ is said to follow a standardized inverse F-Riesz distribution of type $II$ denoted by $\mathbf{R} \sim i\mathcal{F}\mathcal{R}^{II}(\mathbf{\Sigma}, \mathbf{n}, \boldsymbol{\nu})$, with $\mathbf{m}^{i\mathcal{R}^{II}} = (m_1^{i\mathcal{F}\mathcal{R}^{II}}, \ldots, m_p^{i\mathcal{F}\mathcal{R}^{II}})^{\top}$ given by*

$$m_i^{i\mathcal{R}^{II}} = \begin{cases} n_1 m_1^{i\mathcal{R}^{II}}, & \textit{for } i = 1 \\ \sum_{j=1}^{i-1} m_j^{i\mathcal{R}^{II}} + (n_i + i - 1)m_i^{i\mathcal{R}^{II}}, & \textit{for } i > 1, \end{cases} \tag{417}$$

*The probability density function of $\mathbf{R}$ is*

$$p(\mathbf{R}|\mathbf{\Sigma}, \mathbf{n}, \boldsymbol{\nu}) = \frac{\Gamma_p((\boldsymbol{\nu} + \mathbf{n})/2) \prod_{i=1}^{p} (m_i^{i\mathcal{R}^{II}})^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}} \left|\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}}) + \mathbf{Z}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}. \tag{418}$$

*The expected value $\mathbb{E}[\mathbf{R}]$ obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma} \tag{419}$$

*the score w.r.t. $\mathbf{\Sigma}$ is*

$$\nabla = -\mathbf{G}^{\top}\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n}) - \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\mathbf{C}_{\mathbf{B}}\mathrm{dg}(\mathbf{n} + \boldsymbol{\nu})\mathbf{C}_{\mathbf{B}}^{\top}\mathbf{C}^{-\top}\right)\right)\mathbf{C}^{-1}\right), \tag{420}$$

*where $\mathbf{B} = \left(\mathbf{\Omega}^{-1} + \mathbf{R}^{-1}\right)^{-1}$.*

*Proof.* The stochastic representation $\mathbf{Y}$ following an F-Riesz distribution of type $II$ with scale matrix $\mathbf{\Omega}^{-1}$, and degree of freedom parameter vectors $\boldsymbol{\nu}$ and $\mathbf{n}$ is $\mathbf{Y} = \mathbf{U}_{\mathbf{\Omega}^{-1}}\underline{\mathbf{B}}^{-1}\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}\underline{\mathbf{B}}^{-1}\mathbf{U}_{\mathbf{\Omega}^{-1}}^{\top}$, where $\mathbf{U}_{\mathbf{\Omega}^{-1}}$ is the upper Cholesky factor of $\mathbf{\Omega}^{-1}$.[38] Thus

38. See Blasques et al. 2021.

the stochastic representation of the inverse F-Riesz distribution of type $II$ is given by

$$\mathbf{R} = \mathbf{Y}^{-1} = \mathbf{C}_{\boldsymbol{\Omega}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\mathbf{C}_{\boldsymbol{\Omega}}^{\top}, \tag{421}$$

which translate to $\mathbf{R} \sim i\mathcal{R}^{II}(\mathbf{Y}, \boldsymbol{\nu})$, $\mathbf{Y} \sim \mathcal{R}^{I}(\boldsymbol{\Omega}, \mathbf{n})$.[39] For the probability density function we can consequently use

$$\begin{aligned}
p_{{}_{i\mathcal{FR}^{II}}}(\mathbf{R}|\boldsymbol{\Omega}, \mathbf{n}, \boldsymbol{\nu}) &= \int_{\mathbf{Y}>\mathbf{0}} p_{{}_{i\mathcal{R}^{II}}}(\mathbf{R}|\mathbf{Y}, \boldsymbol{\nu}) p_{{}_{\mathcal{R}^{I}}}(\mathbf{Y}|\boldsymbol{\Omega}, \boldsymbol{\nu}) d\mathbf{Y} & (422)\\
&= \int_{\mathbf{Y}>\mathbf{0}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}} |\mathbf{Y}|_{\frac{\boldsymbol{\nu}}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}\mathbf{R}^{-1}\right) \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2) 2^{p\bar{\boldsymbol{\nu}}/2}} & (423)\\
&\quad \times |\mathbf{Y}|_{\frac{\mathbf{n}-p-1}{2}} |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}} \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}^{-1}\mathbf{Y}\right) \frac{1}{\Gamma_p\left(\mathbf{n}/2\right) 2^{p\bar{\mathbf{n}}/2}}\, d\mathbf{Y} & (424)\\
&= \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right) 2^{p\boldsymbol{\nu}\bar{+}\mathbf{n}/2}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}} |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}} \int_{\mathbf{Y}>\mathbf{0}} |\mathbf{Y}|_{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1})\right) d\mathbf{Y} & (425)\\
\text{Theorem 7.4:} &= \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right) 2^{p\boldsymbol{\nu}\bar{+}\mathbf{n}/2}} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}} |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}} 2^{p\boldsymbol{\nu}\bar{+}\mathbf{n}/2}\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2) \left|\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} & (426)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)} |\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}} |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}} \left|\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}. & (427)
\end{aligned}$$

Now, it will be shown below that $\mathbb{E}\left[\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\right] = \mathrm{dg}(\mathbf{m}^{{}_{i\mathcal{R}^{II}}})$ is diagonal, such that $\boldsymbol{\Sigma} =: \mathbb{E}\left[\mathbf{R}\right] = \mathbf{C}_{\boldsymbol{\Omega}}\mathrm{dg}(\mathbf{m}^{{}_{i\mathcal{R}^{II}}})\mathbf{C}_{\boldsymbol{\Omega}}^{\top}$ and consequently $\boldsymbol{\Omega} = \mathbf{C}\mathrm{dg}(\mathbf{m}^{{}_{i\mathcal{R}^{II}}})^{-1}\mathbf{C}^{\top}$. Finally the standardized inverse F-Riesz distribution of type $II$ probability density function is then

39. Recall that $\mathbf{U}_{\boldsymbol{\Omega}^{-1}}^{-\top} = \mathbf{C}_{\boldsymbol{\Omega}}$.

given by

$$
\begin{aligned}
p(\mathbf{R}|\boldsymbol{\Sigma},\mathbf{n},\boldsymbol{\nu}) &= p_{i\mathcal{F}\mathcal{R}^{II}}(\mathbf{R}|\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top,\mathbf{n},\boldsymbol{\nu}) && (428)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}}|\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top|_{-\frac{\mathbf{n}}{2}}\left|\left(\left(\mathbf{C}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top\right)^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (429)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}}|\boldsymbol{\Sigma}|_{-\frac{\mathbf{n}}{2}}\left|\left(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (430)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}}|\boldsymbol{\Sigma}|_{-\frac{\mathbf{n}}{2}}\left|\left(\mathbf{C}^{-\top}\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})+\mathbf{Z}^{-1}\right)\mathbf{C}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (431)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}}|\boldsymbol{\Sigma}|_{-\frac{\mathbf{n}}{2}}\left|\mathbf{C}\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})+\mathbf{Z}^{-1}\right)^{-1}\mathbf{C}^\top\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (432)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{R}|_{-\frac{\boldsymbol{\nu}}{2}}|\boldsymbol{\Sigma}|_{\frac{\boldsymbol{\nu}}{2}}\left|\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})+\mathbf{Z}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (433)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{n_i/2}}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}}\left|\left(\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})+\mathbf{Z}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (434)\\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)\prod_{i=1}^p\left(m_i^{i\mathcal{R}^{II}}\right)^{\nu_i/2}}|\mathbf{R}|^{-\frac{p+1}{2}}|\mathbf{Z}|_{-\frac{\boldsymbol{\nu}}{2}}\left|\left(\mathbf{I}+\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-\frac{1}{2}}\mathbf{Z}^{-1}\mathrm{dg}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{-\frac{1}{2}}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}} && (435)
\end{aligned}
$$

For the expected value we have due to independence

$$
\mathbb{E}\left[\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^\top\right]=\mathbb{E}\left[\underline{\mathbf{B}}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\underline{\mathbf{B}}^\top\right], \tag{436}
$$

where $\mathbf{m}^{i\mathcal{R}^{II}}$ is given in (274). Denote

$$
\mathbf{T}=\underline{\mathbf{B}}\left(\mathbf{m}^{i\mathcal{R}^{II}}\right)^{1/2}, \tag{437}
$$

i.e.

$$
\underline{\mathbf{T}} = \begin{bmatrix}
\sqrt{\chi^2_{n_1-1+1}}\sqrt{m_1^{i\mathcal{R}^{II}}} & 0 & \dots & 0 & 0 \\
\mathcal{N}(0,1)\sqrt{m_1^{i\mathcal{R}^{II}}} & \sqrt{\chi^2_{n_2-2+1}}\sqrt{m_2^{i\mathcal{R}^{II}}} & \ddots & 0 & 0 \\
\mathcal{N}(0,1)\sqrt{m_1^{i\mathcal{R}^{II}}} & \mathcal{N}(0,1)\sqrt{m_2^{i\mathcal{R}^{II}}} & \ddots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\mathcal{N}(0,1)\sqrt{m_1^{i\mathcal{R}^{II}}} & \mathcal{N}(0,1)\sqrt{m_2^{i\mathcal{R}^{II}}} & \ddots & \sqrt{\chi^2_{n_{p-1}-(p-1)+1}}\sqrt{m_{p-1}^{i\mathcal{R}^{II}}} & 0 \\
\mathcal{N}(0,1)\sqrt{m_1^{i\mathcal{R}^{II}}} & \mathcal{N}(0,1)\sqrt{m_2^{i\mathcal{R}^{II}}} & \dots & \mathcal{N}(0,1)\sqrt{m_3^{i\mathcal{R}^{II}}} & \sqrt{\chi^2_{n_p-p+1}}\sqrt{m_p^{i\mathcal{R}^{II}}}
\end{bmatrix}, \tag{438}
$$

with elements $\mathbf{T}_{ij} = \underline{\mathbf{B}}_{ij}\sqrt{m_j^{i\mathcal{R}^{II}}}$. The $(i,j)$ element of $\mathbf{R} = \mathbf{T}\mathbf{T}^\top$ is

$$
\mathbf{R}_{i,j} = \sum_{k=1}^{p} \mathbf{T}_{ik} \left(\mathbf{T}\right)^\top_{kj} = \sum_{k=1}^{p} \mathbf{T}_{ik}\mathbf{T}_{jk} = \sum_{k=1}^{p} m_k^{i\mathcal{R}^{II}} \underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}, \tag{439}
$$

which for $i \neq j$ we have

$$
\mathbb{E}\left[\mathbf{R}_{i,j}\right] = \sum_{k=1}^{p} m_k^{i\mathcal{R}^{II}} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}\right] = \sum_{k=1}^{p} m_k^{i\mathcal{R}^{II}} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\right]\mathbb{E}\left[\underline{\mathbf{B}}_{jk}\right] = 0, \tag{440}
$$

because of independence of the elements in $\underline{\mathbf{B}}$ and the fact that at least one of the elements in each summand is mean zero. Furthermore, for $i = j$ we have

$$
\mathbb{E}\left[\mathbf{R}_{i,i}\right] = \sum_{k=1}^{p} m_k^{i\mathcal{R}^{II}} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}^2\right] = \sum_{k \leq i} m_k^{i\mathcal{R}^{II}} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}^2\right], \tag{441}
$$

with

$$
\mathbb{E}\left[\underline{\mathbf{B}}_{ik}^2\right] = \begin{cases} 1, & \text{for } i \neq k \\ n_k - k + 1 & \text{for } i = k. \end{cases} \tag{442}
$$

Thus the elements of $\mathbb{E}_{i\mathcal{F}\mathcal{R}^{II}}[\mathbf{R}] = \mathrm{dg}\left(\mathbf{m}^{i\mathcal{F}\mathcal{R}^{II}}\right)$ are given by

$$\mathbb{E}[\mathbf{R}_{1,1}] = (n_1 - 1 + 1)m_1^{i\mathcal{R}^{II}}, \tag{443}$$

$$\mathbb{E}[\mathbf{R}_{2,2}] = m_1^{i\mathcal{R}^{II}} + (n_2 - 2 + 1)m_2^{i\mathcal{R}^{II}}, \tag{444}$$

$$\mathbb{E}[\mathbf{R}_{2,2}] = m_1^{i\mathcal{R}^{II}} + m_2^{i\mathcal{R}^{II}} + (n_3 - 3 + 1)m_3^{i\mathcal{R}^{II}}, \tag{445}$$

$$\vdots \tag{446}$$

or

$$m_i^{i\mathcal{R}^{II}} = \sum_{j=1}^{i-1} m_j^{i\mathcal{R}^{II}} + (n_i + i - 1)m_i^{i\mathcal{R}^{II}} \tag{447}$$

or more precisely

$$m_i^{i\mathcal{R}^{II}} = \begin{cases} n_1 m_1^{i\mathcal{R}^{II}}, & \text{for } i = 1 \\ \sum_{j=1}^{i-1} m_j^{i\mathcal{R}^{II}} + (n_i + i - 1)m_i^{i\mathcal{R}^{II}}, & \text{for } i > 1, \end{cases} \tag{448}$$

which for $n_i = n$ and $n_i = n$ for all $i$ equals $\frac{n}{\nu - p - 1}$.

For the score w.r.t. $\boldsymbol{\Sigma}$ start from equation (427), such that

$$\frac{\partial \log p_{iFR^{II}}(\mathbf{R}|\boldsymbol{\Omega},\mathbf{n},\boldsymbol{\nu})}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} = \frac{\partial \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} + \frac{\partial \log \left|\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} \tag{449}$$

$$= \frac{\partial \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} + \frac{\partial \log \left|\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}}{\partial \mathrm{vech}(\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1})^\top} \frac{\partial \mathrm{vech}(\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} \tag{450}$$

$$= -\frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} + \frac{1}{2}\mathrm{vec}\left(\mathbf{C}^{-\top}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}^{-1}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\right)^\top \mathbf{G} \tag{451}$$

$$\times \frac{\partial \mathrm{vech}(\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^\top} \frac{\partial \mathrm{vech}(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1})}{\partial \mathrm{vech}\left(\boldsymbol{\Omega}\right)^\top} \tag{452}$$

$$= -\frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} + \frac{1}{2}\mathrm{vec}\left(\mathbf{C}^{-\top}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}^{-1}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\right)^\top \mathbf{G} \tag{453}$$

$$\times (-1)\mathbf{G}^+\left(\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1} \otimes \left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\right)\mathbf{G}(-1)\mathbf{G}^+\left(\boldsymbol{\Omega}^{-1}\otimes\boldsymbol{\Omega}^{-1}\right)\mathbf{G} \tag{454}$$

$$= -\frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} + \frac{1}{2}\mathrm{vec}\left(\mathbf{C}^{-\top}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}^{-1}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\right)^\top \mathbf{G} \tag{455}$$

$$\times \mathbf{G}^+\left(\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\boldsymbol{\Omega}^{-1} \otimes \left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\boldsymbol{\Omega}^{-1}\right)\mathbf{G} \tag{456}$$

$$= -\frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} \tag{457}$$

$$+ \frac{1}{2}\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\mathbf{C}^{-\top}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}^{-1}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}\boldsymbol{\Omega}^{-1}\right)^\top \mathbf{G} \tag{458}$$

$$= -\frac{1}{2}\mathrm{vec}\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\mathrm{dg}\left(\mathbf{n}\right)\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right)^\top \mathbf{G} \tag{459}$$

$$+ \frac{1}{2}\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\mathbf{C}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}^{\top}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\boldsymbol{\Omega}^{-1}\right)^\top \mathbf{G}. \tag{460}$$

Thus using Lemma (7.16) we have

$$\frac{\log p(\mathbf{R}|\mathbf{\Sigma},\boldsymbol{\nu})}{\partial \text{vech}\,(\mathbf{\Sigma})^\top} = \frac{\partial \log p_{i\mathcal{FR}^{II}}(\mathbf{R}|\mathbf{\Omega},\mathbf{n},\boldsymbol{\nu})}{\partial \text{vech}(\mathbf{\Omega})^\top} \frac{\partial \text{vech}(\mathbf{C}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{\Sigma})^\top} \tag{461}$$

$$= -\frac{1}{2}\text{vec}\left(\mathbf{C}_{\mathbf{\Omega}}^{-\top}\text{dg}\,(\mathbf{n})\,\mathbf{C}_{\mathbf{\Omega}}^{-1} - \mathbf{\Omega}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{\Omega}^{-1}\right)^\top \mathbf{G} \tag{462}$$

$$\times \frac{\partial \text{vech}(\mathbf{C}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{\Sigma})^\top} \tag{463}$$

$$= -\text{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^\top\text{tril}\,(\mathbf{XC})\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1}\right)\mathbf{C}^{-1}\right)^\top \mathbf{G} \tag{464}$$

with

$$\mathbf{XC} = \mathbf{C}_{\mathbf{\Omega}}^{-\top}\text{dg}\,(\mathbf{n})\,\mathbf{C}_{\mathbf{\Omega}}^{-1}\mathbf{C} - \mathbf{\Omega}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{\Omega}^{-1}\mathbf{C} \tag{465}$$

$$= \mathbf{C}_{\mathbf{\Omega}}^{-\top}\text{dg}\,(\mathbf{n})\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2} - \mathbf{\Omega}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{C}^{-\top}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}}), \tag{466}$$

such that

$$\text{tril}\,(\mathbf{XC})\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1} \tag{467}$$

$$= \text{tril}\left(\mathbf{C}_{\mathbf{\Omega}}^{-\top}\text{dg}\,(\mathbf{n})\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1/2} - \mathbf{\Omega}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{C}^{-\top}\right) \tag{468}$$

$$= \text{tril}\left(\mathbf{C}^{-\top}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{1/2}\text{dg}\,(\mathbf{n})\,\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})^{-1/2} - \mathbf{C}^{-\top}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{C} \tag{469}$$

$$= \text{tril}\left(\mathbf{C}^{-\top}\text{dg}\,(\mathbf{n}) - \mathbf{C}^{-\top}\text{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\mathbf{C}_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\text{dg}\,(\mathbf{n}+\boldsymbol{\nu})\,\mathbf{C}^\top_{\left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}}\mathbf{C}^{-\top}\right) \tag{470}$$

and finally

$$
\frac{\log p(\mathbf{R}|\boldsymbol{\Sigma}, \boldsymbol{\nu})}{\partial \mathrm{vech}\left(\boldsymbol{\Sigma}\right)^{\top}} \tag{471}
$$

$$
= -\mathrm{vec}\left(\mathbf{C}^{-\top}\Phi\left(\mathbf{C}^{\top}\mathrm{tril}\left(\mathbf{C}^{-\top}\mathrm{dg}\left(\mathbf{n}\right) - \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{m}^{i\mathcal{R}^{II}})\mathbf{C}^{-1}\mathbf{C}_{\mathbf{B}}\mathrm{dg}\left(\mathbf{n}+\boldsymbol{\nu}\right)\mathbf{C}_{\mathbf{B}}^{\top}\mathbf{C}^{-\top}\right)\right)\mathbf{C}^{-1}\right)^{\top}\mathbf{G}, \tag{472}
$$

where $\mathbf{B} = \left(\boldsymbol{\Omega}^{-1} + \mathbf{R}^{-1}\right)^{-1}$.

For the score w.r.t $\mathbf{n}$ we have, using Lemma 7.5, equation (403), Lemma 7.3 and Lemma 7.6,

$$
\begin{aligned}
&\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \mathbf{n}^{\top}} \\
&= \frac{\log\Gamma_p\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \log\Gamma_p(\mathbf{n}/2) + \log|\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}} + \log|\left(\boldsymbol{\Omega}^{-1} + \mathbf{R}^{-1}\right)^{-1}|_{\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}}{\partial \mathbf{n}^{\top}} \\
&= \frac{1}{2}\boldsymbol{\psi}\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \frac{1}{2}\boldsymbol{\psi}\left(\mathbf{n}/2\right) - \mathrm{vecd}(\mathbf{C}_{\boldsymbol{\Omega}})^{\top} + \mathrm{vecd}(\mathbf{C}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}})^{\top}
\end{aligned}
$$

and

$$
\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \mathbf{n}\partial \mathbf{n}^{\top}} = \frac{1}{4}\left(\boldsymbol{\psi}'\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \boldsymbol{\psi}'\left(\mathbf{n}/2\right)\right).
$$

Furthermore

$$
\begin{aligned}
&\frac{\partial \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}^{\top}} \\
&= \frac{\log\Gamma_p\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \log\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2) + \log|\mathbf{R}|_{-\frac{\boldsymbol{\nu}-p-1}{2}} + \log|\left(\boldsymbol{\Omega}^{-1} + \mathbf{R}^{-1}\right)^{-1}|_{\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}}{\partial \mathbf{n}^{\top}} \\
&= \frac{1}{2}\boldsymbol{\psi}\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \frac{1}{2}\overleftarrow{\boldsymbol{\psi}\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} - \mathrm{vecd}(\mathbf{C}_{\mathbf{R}})^{\top} + \mathrm{vecd}(\mathbf{C}_{\left(\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1}})^{\top}
\end{aligned}
$$

and

$$
\frac{\partial^2 \log p(\mathbf{R}|\boldsymbol{\Sigma}, \mathbf{n}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}\partial \boldsymbol{\nu}^{\top}} = \frac{1}{4}\left(\boldsymbol{\psi}'\left((\mathbf{n}+\boldsymbol{\nu})/2\right) - \overleftarrow{\boldsymbol{\psi}'\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}\right).
$$

□

### 7.6.11 Matrix-F

**Theorem 7.15.** *Let*

$$\mathbf{R} = \frac{\nu - p - 1}{n}\mathbf{C}\bar{\mathcal{B}}^{-\top}\underline{\mathcal{B}}\underline{\mathcal{B}}^{\top}\bar{\mathcal{B}}^{-1}\mathbf{C}^{\top} \tag{473}$$

$$= \frac{\nu - p - 1}{n}\mathbf{C}\underline{\mathcal{B}}\bar{\mathcal{B}}^{-\top}\bar{\mathcal{B}}^{-1}\underline{\mathcal{B}}^{\top}\mathbf{C}^{\top}, \tag{474}$$

*then* $\mathbf{R}$ *is said to follow a standardized matrix-variate F distribution denoted by* $\mathbf{R} \sim \mathcal{F}(\mathbf{\Sigma}, n, \nu)$*. The probability density function of* $\mathbf{R}$ *is*

$$p(\mathbf{R}|\mathbf{\Sigma}, n, \nu) \tag{475}$$

$$= \left(\frac{n}{\nu - p - 1}\right)^{pn/2} \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Sigma}|^{-\frac{n}{2}} |\mathbf{R}|^{\frac{n-p-1}{2}} \left|\mathbf{I}_p + \frac{n}{\nu - p - 1}\mathbf{\Sigma}^{-1}\mathbf{R}\right|^{-\frac{n+\nu}{2}} \tag{476}$$

$$= \left(\frac{\nu - p - 1}{n}\right)^{\nu n/2} \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Sigma}|^{\frac{\nu}{2}} |\mathbf{R}|^{\frac{n-p-1}{2}} \left|\frac{\nu - p - 1}{n}\mathbf{\Sigma} + \mathbf{R}\right|^{-\frac{n+\nu}{2}} \tag{477}$$

$$= \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Omega}|^{\frac{\nu}{2}} |\mathbf{R}|^{\frac{n-p-1}{2}} |\mathbf{\Omega} + \mathbf{R}|^{-\frac{n+\nu}{2}}, \tag{478}$$

*where* $\mathbf{\Omega} = \frac{\nu-p-1}{n}\mathbf{\Sigma}$*. The expected value* $\mathbb{E}[\mathbf{R}]$ *obtains as*

$$\mathbb{E}[\mathbf{R}] = \mathbf{\Sigma} = \frac{n}{\nu - p - 1}\mathbf{\Omega} \tag{479}$$

*the score w.r.t.* $\mathbf{\Sigma}$ *is*

$$\nabla = \frac{1}{2}\mathbf{G}^{\top}\mathrm{vec}\left(\nu\mathbf{\Sigma}^{-1} - (\nu + n)\left(\mathbf{\Sigma} + \frac{n}{\nu - p - 1}\mathbf{R}\right)^{-1}\right) \tag{480}$$

*and the Fisher information matrix w.r.t.* $\mathbf{\Sigma}$ *is given by*

$$\mathcal{I} = . \tag{481}$$

*Proof.* Recall that the stochastic representations above are equal to the ones of the standardized F-Riesz $I$ and standardized inverse F-Riesz $II$ distribution, respectively, by setting $n_i = n$ and $\nu_i = \nu$ for all $i$. Next we show, that they both yield the same probability density function for $\mathbf{R}$ and are thus equivalent. For the probability density function of the first stochastic representation start from the standardized F-Riesz $I$ probability density function to get

$$p\left(\mathbf{R}|\mathbf{\Sigma}, n, \nu\right) = \left(\frac{n}{\nu - p - 1}\right)^{pn/2} \frac{\Gamma_p\left((n+\nu)/2\right)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} \left|\mathbf{R}\right|^{-\frac{p+1}{2}} \left|\mathbf{Z}\right|^{\frac{n}{2}} \left|\mathbf{I} + \frac{n}{\nu - p - 1}\mathbf{Z}\right|^{-\frac{n+\nu}{2}}. \tag{482}$$

Now, starting from the standardized inverse F-Riesz $II$ probability density function we have

$$p\left(\mathbf{R}|\mathbf{\Sigma}, n, \nu\right) = \left(\frac{n}{\nu - p - 1}\right)^{pn/2} \frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(\nu/2)\Gamma_p\left(n/2\right)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{-\frac{\nu}{2}} \left|\left(\frac{n}{\nu - p - 1} + \mathbf{Z}^{-1}\right)^{-1}\right|^{\frac{\nu+n}{2}} \tag{483}$$

$$= \left(\frac{n}{\nu - p - 1}\right)^{pn/2} \frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(\nu/2)\Gamma_p\left(n/2\right)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{-\frac{\nu+n}{2}} |\mathbf{Z}|^{\frac{n}{2}} \left|\frac{n}{\nu - p - 1} + \mathbf{Z}^{-1}\right|^{-\frac{\nu+n}{2}} \tag{484}$$

$$= \left(\frac{n}{\nu - p - 1}\right)^{pn/2} \frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(\nu/2)\Gamma_p\left(n/2\right)} |\mathbf{R}|^{-\frac{p+1}{2}} |\mathbf{Z}|^{\frac{n}{2}} \left|\frac{n}{\nu - p - 1}\mathbf{Z} + \mathbf{I}\right|^{-\frac{\nu+n}{2}}, \tag{485}$$

which is the same probability density function as in (482). Thus both stochastic representations are equivalent.

For the score we have start from (404) to get

$$\frac{\partial \log p\left(\mathbf{R}|\mathbf{\Sigma}, n, \nu\right)}{\partial \mathrm{vech}(\mathbf{\Omega})^\top} = \frac{1}{2}\mathrm{vec}\left(\frac{n\nu}{\nu - p - 1}\mathbf{\Sigma}^{-1} - (\nu + n)\left(\frac{\nu - p - 1}{n}\mathbf{\Sigma} + \mathbf{R}\right)^{-1}\right)^\top \mathbf{G}, \tag{486}$$

such that

$$\frac{\partial \log p\left(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu\right)}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^\top} = \frac{\partial p\left(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu\right)}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} \frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})} \tag{487}$$

$$= \frac{1}{2}\mathrm{vec}\left(\frac{n\nu}{\nu-p-1}\boldsymbol{\Sigma}^{-1} - (\nu+n)\left(\frac{\nu-p-1}{n}\boldsymbol{\Sigma}+\mathbf{R}\right)^{-1}\right)^\top \mathbf{G}\frac{\nu-p-1}{n} \tag{488}$$

$$= \frac{1}{2}\mathrm{vec}\left(\nu\boldsymbol{\Sigma}^{-1} - (\nu+n)\left(\boldsymbol{\Sigma}+\frac{n}{\nu-p-1}\mathbf{R}\right)^{-1}\right)^\top \mathbf{G}. \tag{489}$$

For the Fisher information matrix we have

$$\frac{\partial^2 \log p\left(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu\right)}{\partial \mathrm{vech}(\boldsymbol{\Sigma}) \partial \mathrm{vech}(\boldsymbol{\Sigma})^\top} \tag{490}$$

$$= \frac{\nu}{2}\frac{\partial^2 \log|\boldsymbol{\Sigma}|}{\partial \mathrm{vech}(\boldsymbol{\Sigma}) \partial \mathrm{vech}(\boldsymbol{\Sigma})^\top} - \frac{n+\nu}{2}\frac{\partial^2 \log\left|\frac{\nu-p-1}{n}\boldsymbol{\Sigma}+\mathbf{R}\right|}{\partial \mathrm{vech}(\boldsymbol{\Sigma}) \partial \mathrm{vech}(\boldsymbol{\Sigma})^\top} \tag{491}$$

$$= -\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G} + \frac{n+\nu}{2}\left(\frac{\nu-p-1}{n}\right)^2\mathbf{G}^\top\left(\left(\frac{\nu-p-1}{n}\boldsymbol{\Sigma}+\mathbf{R}\right)^{-1}\otimes\left(\frac{\nu-p-1}{n}\boldsymbol{\Sigma}+\mathbf{R}\right)^{-1}\right)\mathbf{G} \tag{492}$$

$$= -\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G} + \frac{n+\nu}{2}\left(\frac{\nu-p-1}{n}\right)^2\mathbf{G}^\top\left((\boldsymbol{\Omega}+\mathbf{R})^{-1}\otimes(\boldsymbol{\Omega}+\mathbf{R})^{-1}\right)\mathbf{G} \tag{493}$$

$$= -\frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{G} + \frac{n+\nu}{2}\left(\frac{\nu-p-1}{n}\right)^2\mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\otimes\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)\left(\left(\mathbf{I}+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)^{-1}\otimes\left(\mathbf{I}+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)^{-1}\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\otimes\mathbf{C}_{\boldsymbol{\Omega}}^{-} \tag{494}$$

Now, $\mathbf{Z}_{Kollo} = \mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}$ follows a matrix F distribution with scale matrix $\mathbf{I}$ as defined in Theorem 2.4.9. of Kollo and Rosen 2005. They derive on p. 265,

$$
\begin{aligned}
&\mathbb{E}\left[\left(\left(\mathbf{I}+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)^{-1}\otimes\left(\mathbf{I}+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)^{-1}\right)\right] && (495)\\
&\quad=\left(c_3\mathbf{I}+c_4\mathbf{K}_{pp}+c_4\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^{\top}\right) && (496)\\
&\quad=\left((c_3-c_4)\mathbf{I}+c_4\left(\mathbf{I}+\mathbf{K}_{pp}\right)+c_4\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^{\top}\right) && (497)\\
&\quad=\left((c_3-c_4)\mathbf{I}+2c_4\mathbf{G}\mathbf{G}^{+}+c_4\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^{\top}\right), && (498)
\end{aligned}
$$

with according to their p. 263

$$
\begin{aligned}
c_4&=\frac{n-p-1}{(n+\nu-1)(n+\nu+2)}\left(\left(n-p-2+\frac{1}{n+\nu}\right)c_2-\left(1+\frac{n-p-1}{n+\nu}\right)c_1\right), && (499)\\
c_3&=\frac{n-p-1}{n+\nu}\left((n-p-2)\,c_2-c_1\right)-(n+\nu+1)\,c_4, && (500)\\
c_2&=\frac{n(\nu-p-2)+n^2+n}{(\nu-p)(\nu-p-1)(\nu-p-3)}, && (501)\\
c_1&=\frac{n^2(\nu-p-2)+2n}{(\nu-p)(\nu-p-1)(\nu-p-3)}. && (502)
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\mathcal{I} - \mathbb{E}\left[\frac{\partial^2 \log p\left(\mathbf{R}|\boldsymbol{\Sigma}, n, \nu\right)}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \mathrm{vech}(\boldsymbol{\Sigma})^\top}\right] && (503)\\
&\quad = \frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{G}\\
&\qquad - \frac{n+\nu}{2}\left(\frac{\nu-p-1}{n}\right)^2 \mathbf{G}^\top\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-\top} \otimes \mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right)\left((c_3-c_4)\mathbf{I} + 2c_4\mathbf{G}\mathbf{G}^+ + c_4\mathrm{vec}\left(\mathbf{I}\right)\mathrm{vec}\left(\mathbf{I}\right)^\top\right)\left(\mathbf{C}_{\boldsymbol{\Omega}}^{-1} \otimes \mathbf{C}_{\boldsymbol{\Omega}}^{-1}\right. && (504)\\
&\quad = \frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{G}\\
&\qquad - \frac{n+\nu}{2}\left(\frac{\nu-p-1}{n}\right)^2 \mathbf{G}^\top\left((c_3-c_4)\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right) + 2c_4\mathbf{G}\mathbf{G}^+\left(\boldsymbol{\Omega}^{-1} \otimes \boldsymbol{\Omega}^{-1}\right) + c_4\mathrm{vec}\left(\boldsymbol{\Omega}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Omega}^{-}\right.\right. && (505)\\
&\quad = \frac{\nu}{2}\mathbf{G}^\top\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{G}\\
&\qquad - \frac{n+\nu}{2}\mathbf{G}^\top\left((c_3-c_4)\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + 2c_4\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) + c_4\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G} && (506)\\
&= \frac{1}{2}\mathbf{G}^\top\left((\nu-(n+\nu)(c_3+c_4))\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) - (n+\nu)c_4\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\left(\boldsymbol{\Sigma}^{-1}\right)^\top\right)\mathbf{G}. && (507)
\end{aligned}
$$

□

*Commented out is the proof that an inverse matrix-F distribution with* $\boldsymbol{\Sigma}^{-1}$ *expectation is again a matrix-F distribution with* $\boldsymbol{\Sigma}$ *expectation and switched degree of freedoms. For the covariance matrix use Kollo and Rosen 2005, Theorem 2.4.15. together with (*??*.*