Generalized Autoregressive Score Models for Realized Covariance Matrices

Michael Stollenwerk (Heidelberg University)

November 18, 2023

Abstract

In Generalized Autoregressive Score (GAS) models the scaled score of the assumed probability distribution drives the time-variation in the distributions' parameters. The score contains information about the entire shape of the probability density function and thus the parameter updating dynamics are richer than in "traditional" observation-driven models. Probability distributions for RCs can be characterized by their symmetric positive definite expected value matrix and a set of d.o.f. parameters. A central contribution of our paper is to endow the distribution-specific degree of freedom (d.o.f.) parameters with GAS dynamics to make them time-varying. Empirically, we show that the time-variation in the d.o.f. parameters significantly improves fit and forecast performance, and a statistical test confirms that models without time-varying d.o.f. parameters are too restrictive. Furthermore, the fitted values of the d.o.f. parameters make economic sense as they imply more fat-tailed distributions during economic crisis periods. The second main contribution is to derive the scores and Fisher information matrices of all hitherto used probability distributions and to derive a general across-distributions formula for the scaled scores that reduces the computational burden so that the models can be applied to dimensions of up to 50 assets. The comparison across distributions reveals that the Inverse t-Riesz distribution offers the best in- and out-of-sample fit, closely followed by the F-Riesz distribution.

Key words: Realized Covariance Measures, Generalized Autoregressive Score Models, Matrix Distributions, Time-Series Models, Riesz Distributions

1 Introduction

The covariance matrix of financial asset returns is a crucial concept in financial econometrics, with direct implications for efficient portfolio allocation, risk management, and derivative pricing. A *realized covariance matrix* (RC) is an ex-post estimate of the daily integrated covariance matrix constructed from high-frequency data. It can be interpreted as providing an "effectively observable" daily covariance matrix of the underlying financial asset returns, and consequently, directly modeling the time-series of RCs has been advocated in the literature (e.g., Andersen et al., 2001, Andersen et al., 2006, McAleer and Medeiros, 2008, Chiriac and Voev, 2011).

One way to model time-series of RCs is the class of *observation-driven models* as defined by Cox (1981). In this class of models, the RC of a given day is assumed to follow a conditional probability distribution with time-varying parameters, where the parameters are updated using the previous RC realization(s). Probability distributions for RCs can be characterized by their $p \times p$ symmetric positive definite *expected value matrix* and a set of *degree of freedom* (d.o.f.) parameters (see Stollenwerk, 2023). In the literature on observation-driven models for RCs, it is assumed that only the expected value matrix is time-varying, while the d.o.f. parameters are constant (see e.g. Golosnoy, Gribisch, and Liesenfeld, 2012, Noureldin, Shephard, and Sheppard, 2012 and Opschoor et al., 2018).

To make the models applicable to reasonable data dimension the updating process typically has to be restricted to scalar dynamics. That is, next day's expected value matrix is determined by the previous RC(s) multiplied by scalar parameters, which implies that the conditional expected value of any element in the RC on a given day depends only on the realizations of this element in the previous RCs. This represents a very strong assumption, as there can be no spillovers between the realized (co)variances.

A very general subclass of observation-driven models that avoids this assumption is given by the *generalized autoregressive score* (GAS) models, introduced by Creal, Koopman, and Lucas (2011), Creal, Koopman, and Lucas (2013) and Harvey (2013). In GAS models, information about the entire shape of the assumed conditional probability distribution is incorporated directly into the pa-

rameter updating process, by using the distribution-specific *score*, often scaled by the corresponding inverse of the *Fisher information matrix* (FIM), as the forcing variable.[1] Advantages of GAS models over traditional observation-driven models are as follows. First, GAS models feature rich dynamic updating dynamics even if scalar dynamics are assumed. Second, they offer an intuitive way to make *any* parameter of a probability distribution time-varying, not just those that have an obvious observable forcing variable like the expected value matrix of distributions for RCs. Furthermore, GAS models have been shown to possess desirable information-theoretic optimality (see Blasques, Koopman, and Lucas, 2015). Finally, it is noteworthy that GAS models contain some traditional observation-driven models as special cases. One of these special cases is discovered in this paper.[2]

While the general literature on GAS models is very extensive (see Harvey, 2022 and www.gasmodel.com), and for time-series of RCs, the literature on traditional observation-driven is quite large (see references above and, among others, Asai and So, 2013, Jin and Maheu, 2016 and Zhou et al., 2019), the literature on GAS models for time-series of RCs is very scarce. To the best of our knowledge, there are only two papers. The first is by Opschoor et al. (2018), who propose to model the daily return vectors and the RCs jointly. They assume that the daily return vectors follow a multivariate standardized t-distribution with a covariance matrix that is equal to the expected value matrix of the standardized matrix-F distribution, assumed for the RCs. This parameter matrix follows GAS dynamics. The authors show that their model significantly outperforms all previously proposed competitors in terms of in-sample fit and out-of-sample forecasting ability. The second paper on GAS models for RCs is by Gorgi et al. (2019). They propose a model that is similar to the one in Opschoor et al. (2018). However, they assume a multivariate standardized normal (nested by the t) and Wishart (nested by the matrix-F) distribution for the daily returns and RCs, respectively. They generalize the GAS dynamics of Opschoor et al. (2018) by assuming that the covariance matrix of the normal distribution is equal to the *scaled* expected value matrix of the Wishart, and by assuming full GAS rather than scalar-GAS dynamics. While

1. We call the variable that drives the updating process the "forcing variable".
2. See Lemma 3.1.

the full GAS dynamics make their model more flexible, it also makes it infeasible even for moderate cross-sectional dimensions.

The scarcity of literature on GAS models for RCs is despite there having been proposed many different probability distributions for RCs (see Stollenwerk, 2023). Each distribution would entail a different parameter updating forcing variable in a GAS model since the score is distribution-specific. This paper aims to extend the literature by deriving GAS models for RCs for all probability distributions used in the literature. This extension is important because, in GAS models the natural question arises which distributions and their scores work best in practice. We contribute to the literature by deriving a general (across all distributions) representation of the scaled score w.r.t. the expected value matrix that makes the GAS models computationally feasible for medium to large RCs (say, five to 50 assets).

Furthermore, we provide empirical evidence consistent with economic intuition, which suggests that the d.o.f. parameters should also be modeled as time-varying. Consequently, another contribution we make is to assume time-varying d.o.f. parameters by endowing them with GAS dynamics. For realized variances, i.e. one-dimensional RCs, time-varying d.o.f. parameters with GAS dynamics have been proposed by Opschoor and Lucas (2022), who consider the special case of the F distribution. Using various loss functions, they show that their model outperforms competitors without time-varying d.o.f. parameters in an out-of-sample forecasting exercise. For RCs there is, to the best of our knowledge, no study to date considering time-varying d.o.f. parameters.

We contribute by deriving all the necessary theoretical inputs for GAS models with time-varying expected value matrix and time-varying d.o.f. parameters. That is, we derive the scores and the FIMs w.r.t. the expected value matrix and the d.o.f. parameters for all probability distributions for RCs.[3]. To the best of our knowledge, only a very small subset of these quantities (e.g. the score and FIM w.r.t. the expected value matrix of the Wishart and the score w.r.t. the expected value matrix of the matrix-F) have been derived before.

Finally, in the empirical section, likelihood ratio tests and forecast comparisons

3. We do not obtain the FIMs w.r.t. the expected value matrix for the Inverse t-Riesz and the (Inverse) F-Riesz distributions.

reveal that time-varying d.o.f. parameters are indeed important for all distributions. Furthermore, we show that in our model the t-Riesz distribution family and the F-Riesz distribution family are the best distributions in terms of fit and forecasting performance.

The rest of this paper is structured as follows. The next section examines the drawbacks of non-GAS observation-driven models for RCs and explains how the GAS models alleviate them. Section 3 introduces the GAS model framework for time-series of RCs in general and presents, in particular, (the derivations for) our GAS models for all distributions for RCs. Section 4 contains the empirical application with in-sample fit and out-of-sample forecasting comparison that investigates among others the relevance of time-varying d.o.f. parameters across all distributions. Section 5 concludes.

2 Drawbacks of Traditional Observation-Driven Models for RCs

To be more precise on the traditional observation-driven model setup, that is a model without GAS dynamics, remember that all probability distributions $\mathcal{D}$ for RCs can be characterized by a symmetric positive definite expected value matrix $\mathbf{\Sigma}$ and their distribution-specific d.o.f. parameters, which we collect in $\boldsymbol{\theta}_{\mathcal{D}}$. Thus, the distributional assumption can be written as,

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

where $\mathbf{R}_t$ denotes an RC at time t and $\mathcal{F}_{t-1} = \{\mathbf{R}_{t-1}, \mathbf{R}_{t-2} \ldots\}$. The scalar dynamics for the conditional expected value matrix in a traditional observation-driven model might be, for example, a scalar-BEKK specification:

$$\mathbf{\Sigma}_{t+1} = (1 - a - b)\mathbf{\Xi} + a\mathbf{R}_t + b\mathbf{\Sigma}_t. \tag{2}$$

This scalar-dynamic assumption, which is necessary to make the models applicable to more than, say, five assets, restrictively implies that each element in $\mathbf{\Sigma}_{t+1}$ only depends on the corresponding previous element in $\mathbf{\Sigma}_t$ and the corresponding pre-

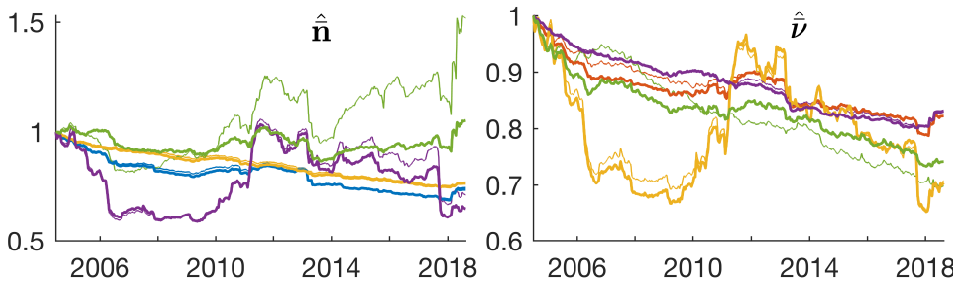

Figure 1: (Average) estimated d.o.f. parameters $\mathbf{n}$ and $\boldsymbol{\nu}$ from estimation on a moving window of 1250 observations of scalar-BEKK model for all distributions for the ten-dimensional dataset described in 4.1. All time-series are normed to the first estimated (average) degree of freedom. The plotted points correspond to the mid-point of the moving window sample. The color coding is: Riesz, Wishart, Inv.Riesz, Inv.Wishart, t-Riesz, t-Wishart, Inv.t-Riesz, Inv.t-Wishart, F-Riesz, F. The thick line corresponds to the Riesz-type distributions, for which we plot the average of their d.o.f. parameter vectors, the thin line to the corresponding Wishart-named (and the F) ones.

vious realization in $\mathbf{R}_t$. That is, there is no possibility for (co)volatility spillovers, which is a very strong assumption. In reality there is complex non-linear interdependence between different firms. On the other hand, a scalar-GAS model, where $\mathbf{R}_t$ in equation (2) is replaced by the score of the observational density, naturally accommodates volatility spillovers, since $\mathbf{R}_t$ enters non-linearly in the score.

The second drawback of traditional observation-driven models for RCs is the central assumption of constant d.o.f. parameters. To illustrate this we estimated the scalar-BEKK model given by equations (1) and (2) for all distributions for RCs $\mathcal{D}$ on a rolling window of 1250 observations for the ten-dimensional dataset described in Section 4.1. We plot the resulting d.o.f. parameter estimates in Figure 1. We see that the estimated d.o.f. parameters vary substantially over time for all distributions, even if we introduce time-variation in the expected value matrix ($\boldsymbol{\Sigma}_t$) via the scalar-BEKK specification. For an intuition, recall that the d.o.f. parameter ν of the t-Wishart distribution regulates its fat-tailedness, with lower values implying a more fat-tailed distribution. In Figure 1, we indeed see, as we would expect, that $\hat{\nu}$ of the t-Wishart decreases substantially (around 30 %) during the 2008 global financial crisis and in the COVID-19 crisis. We conclude that the assumption of static d.o.f. parameters is restrictive and that there is statistical and economic value in exploring time-variation in the d.o.f. parameters. For the d.o.f. parameters in an observation-driven model, there is no observable

realization to use in an updating mechanism. Here, GAS models, with their use of the lagged score as the forcing variable, offer an intuitive way to dynamically update the d.o.f. parameters.

3 GAS Model Setup

As mentioned above, probability distributions for RCs can be characterized by a symmetric positive definite expected value matrix $\boldsymbol{\Sigma}$ and their distribution-specific d.o.f. parameters (one or two of the set $(n, \nu, \mathbf{n}, \boldsymbol{\nu})$), which we collect in the vector $\boldsymbol{\theta}_{\mathcal{D}}$. For example, for the F-Riesz distribution we have $\boldsymbol{\theta}_{F\mathcal{R}} = (\mathbf{n}^\top, \boldsymbol{\nu}^\top)^\top$ with d.o.f. parameter vectors $\mathbf{n} = (n_1, \ldots, n_p)^\top$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^\top$, the t-Riesz distribution has d.o.f. parameter vector $\mathbf{n}$ and scalar d.o.f. ν, thus $\boldsymbol{\theta}_{t\mathcal{R}} = (\mathbf{n}^\top, \nu)^\top$, and the Wishart distribution has just one d.o.f. parameter, thus $\boldsymbol{\theta}_{\mathcal{W}} = n$. Assuming that all parameters are time-varying, the conditional distribution can be written as

$$\mathbf{R}_t | \mathcal{F}_{t-1} \sim \mathcal{D}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t}). \tag{3}$$

In the GAS model framework, the parameter updating at time $t+1$ is based on the score vector of the respective distribution $\mathcal{D}$ at time t,

$$\nabla_{\mathcal{D},t} = \left(\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t} | \mathbf{R})}{\partial \text{vech}(\boldsymbol{\Sigma}_t)^\top}, \frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t} | \mathbf{R})}{\partial \boldsymbol{\theta}_{\mathcal{D},t}^\top} \right)^\top,$$

where $p_{\mathcal{D}}(\cdot)$ is the probability density function (p.d.f.) of distribution $\mathcal{D}$, and the vech$(\cdot)$-operator takes a lower-triangular or symmetric matrix as input and stacks its elements on and below the main diagonal column-wise into a column vector. We denote the distinct parts of the score by

$$\nabla^{\Sigma}_{\mathcal{D},t} = \left(\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t} | \mathbf{R})}{\partial \text{vech}(\boldsymbol{\Sigma}_t)^\top} \right)^\top \text{ and}$$
$$\nabla^{\theta}_{\mathcal{D},t} = \left(\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t} | \mathbf{R})}{\partial \boldsymbol{\theta}_{\mathcal{D},t}^\top} \right)^\top.$$

See Section 6.1.4 in the appendix for a treatment of multivariate derivatives.[4]

The time-t score vector gives the parameter updating direction in which the log-likelihood at time t can be improved most. Thus, it is a natural choice for a forcing variable in a dynamic updating equation. However, empirically, it behaves too erratically in order to extract valuable signals, and thus it is usually advocated to scale the score in some way (c.f. Creal, Koopman, and Lucas, 2013). A widely used scaling matrix is the inverse Fisher information matrix (FIM) (Gorgi et al., 2019, Blasques et al., 2022), which embodies information about the curvature of the log-likelihood,

$$\mathcal{I}_{\mathcal{D},t} = \mathbb{E}\begin{bmatrix} \nabla^{\Sigma}_{\mathcal{D},t}(\nabla^{\Sigma}_{\mathcal{D},t})^{\top} & \nabla^{\Sigma}_{\mathcal{D},t}(\nabla^{\theta}_{\mathcal{D},t})^{\top} \\ \nabla^{\theta}_{\mathcal{D},t}(\nabla^{\Sigma}_{\mathcal{D},t})^{\top} & \nabla^{\theta}_{\mathcal{D},t}(\nabla^{\theta}_{\mathcal{D},t})^{\top} \end{bmatrix}, \tag{4}$$

to obtain the scaled score vector

$$\mathbf{s}_{\mathcal{D},t} = \mathcal{I}^{-1}_{\mathcal{D},t}\nabla_{\mathcal{D},t}.$$

4. It is not immediately obvious how to arrange the individual derivatives $\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D},t}|\mathbf{R})}{\partial(\boldsymbol{\Sigma}_t)_{ij}}$. Also, it is important to take the derivative w.r.t. vech($\boldsymbol{\Sigma}$), not vec($\boldsymbol{\Sigma}$), to take the symmetry of $\boldsymbol{\Sigma}$ into account. Both points are explained in more detail in Section 6.1.4.

In order to be able to estimate the parameters of the updating equation, one has to choose scalar dynamics, akin to the scalar-BEKK model, because full matrix dynamics suffer from the *curse of dimensionality*. However, since it seems unrealistic that $\boldsymbol{\Sigma}_t$ and the d.o.f. parameters $(\mathbf{n}_t, n_t, \boldsymbol{\nu}_t, \nu_t)$ follow the same scalar dynamics, we assume separate scalar-GAS updating equations for them. The scaled score w.r.t. $\boldsymbol{\Sigma}_t$ is defined as

$$\mathbf{S}^{\Sigma}_{\mathcal{D},t} = \text{ivech}((\mathcal{I}^{\Sigma}_{\mathcal{D},t})^{-1}\nabla^{\Sigma}_{\mathcal{D},t}), \tag{5}$$

$$\text{with } \mathcal{I}^{\Sigma}_{\mathcal{D},t} = \mathbb{E}[\nabla^{\Sigma}_{\mathcal{D},t}(\nabla^{\Sigma}_{\mathcal{D},t})^{\top}], \tag{6}$$

where ivech$(\cdot)$ is the inverse vech operator that creates a symmetric matrix from a vector of suitable size. For the d.o.f. parameters in $\boldsymbol{\theta}_{\mathcal{D},t}$ we assume separate scalar-GAS updating equations for $\mathbf{n}_t, n_t, \boldsymbol{\nu}_t$ and ν_t, with the scaled scores w.r.t. $\mathbf{n}_t$ and n_t given by

$$\mathbf{s}^{\mathbf{n}}_{\mathcal{D},t} = (\mathcal{I}^{\mathbf{n}}_{\mathcal{D},t})^{-1}\nabla^{\mathbf{n}}_{\mathcal{D},t}$$
$$\text{and } s^{n}_{\mathcal{D},t} = \frac{\nabla^{n}_{\mathcal{D},t}}{\mathcal{I}^{n}_{\mathcal{D},t}},$$

where

$$\nabla^{\mathbf{n}}_{\mathcal{D},t} = \left(\frac{\partial \log p_{\mathcal{D}}(\cdot)}{\partial \mathbf{n}_t^{\top}}\right)^{\top},$$
$$\nabla^{n}_{\mathcal{D},t} = \frac{\partial \log p_{\mathcal{D}}(\cdot)}{\partial n_t},$$

and

$$\mathcal{I}^{\mathbf{n}}_{\mathcal{D},t} = \mathbb{E}[\nabla^{\mathbf{n}}_{\mathcal{D},t}(\nabla^{\mathbf{n}}_{\mathcal{D},t})^{\top}],$$
$$\mathcal{I}^{n}_{\mathcal{D},t} = \mathbb{E}[(\nabla^{n}_{\mathcal{D},t})^{2}].$$

The scaled scores w.r.t $\boldsymbol{\nu}_t$ and ν_t are defined analogously to those of $\mathbf{n}_t$ and n_t. Which of the d.o.f. parameter updating equations is present in the respective model depends, of course, on the chosen distribution.

Distribution	$2\triangle^{\Omega}_{\mathcal{D}}$
Wishart	$\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - n\boldsymbol{\Omega}^{-1}$
Riesz	$\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}$
Inv.Wishart	$\nu\boldsymbol{\Omega}^{-1} - \mathbf{R}^{-1}$
Inv.Riesz	$\mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathbf{C}_{\Omega}^{-1} - \mathbf{R}^{-1}$
t-Wishart	$\frac{\nu+pn}{1+\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - n\boldsymbol{\Omega}^{-1}$
t-Riesz	$\frac{\nu+p\bar{\mathbf{n}}}{1+\mathrm{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})}\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}$
Inv.t-Wishart	$\nu\boldsymbol{\Omega}^{-1} - \frac{\nu+pn}{1+\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}\mathbf{R}^{-1}$
Inv.t-Riesz	$\mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathbf{C}_{\Omega}^{-1} - \frac{n+p\bar{\nu}}{1+\mathrm{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})}\mathbf{R}^{-1}$
F	$\mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathbf{C}_{\Omega}^{-1} - \mathbf{C}_{H}^{-\top}\mathrm{dg}(\boldsymbol{\nu}+\mathbf{n})\mathbf{C}_{H}^{-1}$
F-Riesz	$\mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\boldsymbol{\nu})\mathbf{C}_{\Omega}^{-1} - \mathbf{C}_{H}^{-\top}\mathrm{dg}(\boldsymbol{\nu}+\mathbf{n})\mathbf{C}_{H}^{-1}$
Inv.F-Riesz	$\mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{C}_{J}\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_{J}^{\top}\boldsymbol{\Omega}^{-1}$

Table 1: $\triangle^{\Omega}_{\mathcal{D}}$ is the $p\times p$ score matrix w.r.t. $\boldsymbol{\Omega}$, as defined in equation (33) in the appendix, with subscripts t omitted for readability. The scores w.r.t. $\boldsymbol{\Omega}$ are given by $\nabla^{\Omega}_{\mathcal{D}} = \mathbf{G}^{\top}\mathrm{vec}(\triangle^{\Omega}_{\mathcal{D}})$ and the ones w.r.t. $\boldsymbol{\Sigma}$ then easily follow from Lemma 6.9. The proofs are straightforward using Lemmas 6.6 and 6.7. As an example, see Section 6.2 in the appendix for the derivation of the F-Riesz score. $\mathbf{C}_H$ and $\mathbf{C}_J$ denote the lower Cholesky factor of $\mathbf{H} = \boldsymbol{\Omega} + \mathbf{R}$ and $\mathbf{J} = (\boldsymbol{\Omega}^{-1} + \mathbf{R}^{-1})^{-1}$, respectively.

3.1 Expected Value Matrix Recursion

The scores and FIMs w.r.t. $\boldsymbol{\Sigma}_t$ for the different distributions are given in Tables 1 and 2. The derivations are given in Section 6.2 in the appendix. The formulas for the FIMs contain $p^2 \times p^2$ matrix multiplications and inversions which makes the direct calculation of the scaled score (equation (5)) prohibitively slow. Thus, we propose to scale the score w.r.t. $\boldsymbol{\Sigma}_t$ of all Riesz-type distributions with the FIMs of the corresponding Wishart-type distributions. That is, for example, we scale the F-Riesz distribution score ($\nabla^{\Sigma}_{F\mathcal{R},t}$) by the F distribution inverse FIM ($\mathcal{I}^{\Sigma}_{F,t}$) instead of its own ($\mathcal{I}^{\Sigma}_{F\mathcal{R},t}$), by setting $n_{F,t} = \bar{\mathbf{n}}_{F\mathcal{R},t}$ and $\nu_{F,t} = \bar{\boldsymbol{\nu}}_{F\mathcal{R},t}$.

We can be flexible with the scaling of the scores since the most important information is contained in the score itself, which at time t defines the steepest ascend direction of the log-likelihood in which the parameters should be updated in $t+1$. The scaling just serves to stabilize the time-series behavior of the scores. In fact, in a small dimensional example using the t-Riesz distribution we found

Distribution	$\mathcal{I}_{\mathcal{D}}^{\Sigma}$
Wishart	$\frac{n}{2}\mathbf{G}^{\top}\boldsymbol{\Sigma}^{-\otimes 2}\mathbf{G}$
Inv.Wishart	$\frac{\nu}{2}\mathbf{G}^{\top}\boldsymbol{\Sigma}^{-\otimes 2}\mathbf{G}$
t-Wishart	$\frac{n}{2}\mathbf{G}^{\top}\left(\frac{\nu+pn}{\nu+pn+2}\boldsymbol{\Sigma}^{-\otimes 2}-\frac{n}{(\nu+pn+2)}\text{vec}^2\left(\boldsymbol{\Sigma}^{-1}\right)\right)\mathbf{G}$
Inv.t-Wishart	$\frac{\nu}{2}\mathbf{G}^{\top}\left(\frac{n+p\nu}{n+p\nu+2}\boldsymbol{\Sigma}^{-\otimes 2}-\frac{\nu}{(n+p\nu+2)}\text{vec}^2\left(\boldsymbol{\Sigma}^{-1}\right)\right)\mathbf{G}$
F	$\frac{1}{2}\mathbf{G}^{\top}\left(\left(\nu+(n+\nu)(c_3+c_4)\right)\boldsymbol{\Sigma}^{-\otimes 2}+(n+\nu)c_4\text{vec}^2\left(\boldsymbol{\Sigma}^{-1}\right)\right)\mathbf{G}$

Table 2: Fisher information matrices w.r.t $\boldsymbol{\Sigma}$ of Wishart-type distributions. The subscripts t are omitted for readability. For the derivations (and the definitions of c_3 and c_4) see Section 6.2.3 in the appendix. The Fisher information matrices of the (inverse) Riesz and t-Riesz distributions are also derived in the appendix.

that the $\boldsymbol{\Sigma}_t$-dynamics obtained from scaling with the "correct" t-Riesz FIM are very similar to the ones obtained from scaling with the t-Wishart FIM. Opschoor et al. (2018) scale the matrix F distribution with the inverse FIM of the Wishart distribution. Furthermore, Blasques, Francq, and Laurent (2022) show that it is possible to disentangle the distributional assumption from the score dynamics completely, i.e. one could update the parameters with a scaled score from a different distribution than the conditional distributional assumption on the data.

In Theorem 3.1 we show that in the formulas for the scaled scores that result from this scaling scheme, only $p \times p$ matrix operations remain, which makes the models empirically feasible again.

Theorem 3.1. *Consider the scaled score w.r.t.* $\mathbf{\Sigma}_t$*,*

$$\mathbf{S}^{\Sigma}_{\mathcal{D},t} = \text{ivech}((\mathcal{I}^{\Sigma}_{\mathcal{D},t})^{-1}\nabla^{\Sigma}_{\mathcal{D},t}),$$

as defined in equation (5). If, for any Riesz-type distribution, we use $\mathcal{I}^{\Sigma}_{\mathcal{D},t}$ *of its Wishart-type counterpart, instead of its own, by setting the degree(s) of freedom equal to the average of the corresponding d.o.f. parameter vector(s). Then, for all distributions for RCs the scaled score w.r.t.* $\mathbf{\Sigma}_t$ *can be written as*

$$\dot{\mathbf{S}}^{\Sigma}_{\mathcal{D},t} = \alpha_{\mathcal{D}}\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t}\mathbf{\Sigma}_t + \beta_{\mathcal{D}}\text{tr}\left(\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t}\right)\mathbf{\Sigma}_t,$$

where $\triangle^{\Sigma}_{\mathcal{D},t}$ *is the* $p \times p$ *score matrix w.r.t.* $\mathbf{\Sigma}_t$ *ignoring symmetry (as defined in equation (30)), and* $\alpha_{\mathcal{D}}$ *and* $\beta_{\mathcal{D}}$ *are scalars that depend only on the d.o.f. parameters of the respective distribution. Proof in Appendix.*

We omit the FIMs w.r.t. $\mathbf{\Sigma}_t$ for the Riesz-type distributions in Table 2 since using Theorem 3.1 we don't need them and their formulas are long and cumbersome.[5]

Now we present our specification for the expected value matrix equation. We assume

$$\begin{aligned}\mathbf{\Sigma}_{t+1} = (1 - b_1 - b_2 - b_3)\mathbf{\Xi} + a_1\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t}\mathbf{\Sigma}_t + a_2\text{tr}(\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t})\mathbf{\Sigma}_t \\ + b_1\mathbf{\Sigma}_t + b_2\bar{\mathbf{\Sigma}}_{t:t-4} + b_3\bar{\mathbf{\Sigma}}_{t:t-21},\end{aligned} \tag{7}$$

where $\mathbf{\Xi}$ is a $p \times p$ symmetric positive definite parameter matrix and a_1, a_2, b_1, b_2 and b_3 are scalar parameters. Following Opschoor et al. (2018), we assume a HAR-type (see Corsi, 2009) structure, where $\bar{\mathbf{\Sigma}}_{t:t-4} = 1/5\sum_{j=0}^{4}\mathbf{\Sigma}_{t-j}$ and $\bar{\mathbf{\Sigma}}_{t:t-21} = 1/22\sum_{j=0}^{21}\mathbf{\Sigma}_{t-j}$ are the average expected value matrix of the previous trading week and month, respectively. In our empirical analysis, the HAR structure improved the fit and forecasting ability substantially over the version where only $\mathbf{\Sigma}_t$ goes into the updating process. Note that our scaling nest the one in Opschoor et al. (2018) by setting $a_2 = 0$. It is important to see, that instead of straightforwardly incorporating Theorem 3.1 by using $a\dot{\mathbf{S}}^{\Sigma}_{\mathcal{D},t} = a(\alpha_{\mathcal{D}}\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t}\mathbf{\Sigma}_t + \beta_{\mathcal{D}}\text{tr}(\mathbf{\Sigma}_t\triangle^{\Sigma}_{\mathcal{D},t})\mathbf{\Sigma}_t)$ as

5. You can find the FIMs for some Riesz-type distributions expressions in Section 6.2.4 in the appendix.

the forcing variable in our updating equation, we allow for different stand-alone parameters (a_1,a_2) for the two terms that constitute $\dot{\mathbf{S}}^{\Sigma}_{\mathcal{D},t}$. This comes at the small cost of adding one more parameter but has the advantage of adding flexibility and making the updating equation completely independent of the FIMs such that we don't need to calculate the complicated formulas for $\alpha_{\mathcal{D}}$ and $\beta_{\mathcal{D}}$.[6] Note that under this small generalization the mean-zero property of the two scaled score terms is preserved since $\mathbb{E}[\triangle^{\Sigma}_{\mathcal{D},t}|\mathcal{F}_{t-1}] = 0$. Furthermore, it is directly visible that for all distributions, even for scalar parameters there are spillovers of the realized (co)variances of all assets to the expected realized (co)variances of all other assets.

Note that our specification endows the expected value matrix $\mathbf{\Sigma}_t$ with GAS dynamics which is different from a model that assumes GAS dynamics for time-varying parameter matrix $\mathbf{\Omega}_t$ (recall that the probability distributions can be parameterized in terms of expected value matrix $\mathbf{\Sigma}_t$ or parameter matrix $\mathbf{\Omega}_t$; see Stollenwerk, 2023). This is because the score w.r.t. $\mathbf{\Omega}$ is not simply a multiple of the score w.r.t. $\mathbf{\Sigma}$. In contrast, in a scalar-BEKK specification as in Blasques et al. (2021), it does *not* matter if the distributions are parameterized in terms of $\mathbf{\Omega}_t$ or $\mathbf{\Sigma}_t$. We choose the model $\mathbf{\Sigma}_t$ with a GAS specification instead of $\mathbf{\Omega}_t$ because there are several advantages to it. First, unlike $\mathbf{\Omega}_t$, the $\mathbf{\Sigma}_t$ parameter has the same meaning across distributions and an intuitive interpretation of yielding a mean-shifting process. Second, the distributions when parameterized in terms of $\mathbf{\Sigma}_t$ nest each other according to the distribution family tree in Stollenwerk (2023), 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 GAS model with nested distributions as starting points for the estimation of the GAS models with nesting distributions. Third, the targeting estimation of the intercept matrix in the $\mathbf{\Sigma}_t$-GAS recursion is feasible and very easy. In fact, if we introduce time-variation in the d.o.f. parameters, like we do in the next section, it is impossible to target the intercept matrix of a $\mathbf{\Omega}_t$-GAS specification since the unconditional mean of $\mathbf{\Omega}_t$ depends on the d.o.f. parameters.

As an interesting side note, in Lemma 3.1, we show that, for the case of the Riesz (and Wishart) distribution, the original scaled score $\mathbf{S}^{\Sigma}_{\mathcal{D},t}$ boils down to the

6. These formulas are especially complicated for the F distribution.

difference between $\mathbf{R}_t$ and $\mathbf{\Sigma}_t$ (and thus also does not require $p^2 \times p^2$ matrix multiplications).

Lemma 3.1. *For the Riesz (and consequently also the Wishart) distribution, we have that the scaled score w.r.t.* $\mathbf{\Sigma}$, $\mathbf{S}^{\Sigma}_{\mathcal{D},t}$, *in equation (5) can be rewritten as*

$$\mathbf{S}^{\Sigma}_{\mathcal{R},t} = \mathbf{R}_t - \mathbf{\Sigma}_t.$$

Proof in Appendix.

Consequently, a scalar-GAS (1,1) specification would simplify to

$$\begin{aligned}\mathbf{\Sigma}_{t+1} &= (1-b)\mathbf{\Xi} + a\mathbf{S}^{\Sigma}_{\mathcal{R},t} + b\mathbf{\Sigma}_t \\ &= (1-b)\mathbf{\Xi} + a\mathbf{R}_t + (b-a)\mathbf{\Sigma}_t,\end{aligned}$$

which is equivalent to the well-known scalar-BEKK specification in equation (2). For the other Riesz-type distributions, however, such simplifications are not possible.

3.2 Degree of Freedom Parameters Recursion

We assume that the time-varying d.o.f. parameters in $\boldsymbol{\theta}_{\mathcal{D},t}$ follow GARCH-type recursions,

$$\mathbf{n}_{t+1} = (1-b^{\mathbf{n}})\boldsymbol{\xi}^{\mathbf{n}} + a^{\mathbf{n}}\mathbf{s}^{\mathbf{n}}_{\mathcal{D},t} + b^{\mathbf{n}}\mathbf{n}_t, \tag{8}$$

$$n_{t+1} = (1-b^{n})\xi^{n} + a^{n}s^{n}_{\mathcal{D},t} + b^{n}n_t, \tag{9}$$

where $\boldsymbol{\xi}^{\mathbf{n}}$ is a parameter vector of size $p \times 1$, $\xi^n, a^{\mathbf{n}}, b^{\mathbf{n}}, a^n$ and b^n are scalar parameters, and $\mathbf{s}^{\mathbf{n}}_{\mathcal{D},t}$ ($p \times 1$) and $\mathbf{s}^{n}_{\mathcal{D},t}$ (scalar) are the scaled scores with respect to $\mathbf{n}_t$ and n_t, respectively.[7] The specifications for $\boldsymbol{\nu}_t$ and ν_t are analogous to those of $\mathbf{n}_t$ and n_t. Of course, which of the recursions ($\mathbf{n}_t, \boldsymbol{\nu}_t, n_t$ and/or ν_t) are present in the respective model depends on the chosen distribution $\mathcal{D}$. We collect all distribution-specific d.o.f.-recursion parameters in the vector $\boldsymbol{\vartheta}_{\mathcal{D}}$. For example,

7. To mirror the $\mathbf{\Sigma}_t$ specification, we also tried HAR-type recursions for the d.o.f. parameters. However, in the empirical application, it turned out that the lagged weekly and monthly average d.o.f. parameters were insignificant, and their estimates were often close to zero.

$\boldsymbol{\vartheta}_{\mathcal{W}} = (\xi^n, a^n, b^n)^\top$ and $\boldsymbol{\vartheta}_{F\mathcal{R}} = ((\boldsymbol{\xi}^{\mathbf{n}})^\top, a^{\mathbf{n}}, b^{\mathbf{n}}, (\boldsymbol{\xi}^{\boldsymbol{\nu}})^\top, a^{\boldsymbol{\nu}}, b^{\boldsymbol{\nu}})^\top$. The scores and FIMs w.r.t. $\mathbf{n}_t, n_t, \boldsymbol{\nu}_t$ and ν_t, which are needed to construct the forcing variables (scaled scores) in the recursions, are listed in Tables 3 and 4, respectively, and are derived in Section 6.2 in the appendix. Note that for the special case that the non-intercept parameters $(a^{\mathbf{n}}, a^n, a^{\boldsymbol{\nu}}, a^\nu, b^{\mathbf{n}}, b^n, b^{\boldsymbol{\nu}}, b^\nu)$ are set to equal zero, we recover a restricted GAS model with constant d.o.f. parameters (GAS dynamics only in $\boldsymbol{\Sigma}_t$).

To give some intuition on the interpretation of the d.o.f. parameters, consider omitting the subscript t, the one-dimensional case, i.e. $\mathbf{R}$ being of size 1×1. In that case the stochastic representation of all distributions[8] can be written as

$$\mathbf{R} = \begin{cases} \boldsymbol{\Sigma}\, \Gamma(\frac{n}{2}, \frac{n}{2}) & (\mathcal{W}, \mathcal{R}) \\ \boldsymbol{\Sigma}\, \Gamma^{-1}(\frac{\nu}{2}, \frac{\nu-2}{2}) & (i\mathcal{W}, i\mathcal{R}) \\ \boldsymbol{\Sigma}\, \Gamma(\frac{n}{2}, \frac{n}{2})\Gamma^{-1}(\frac{\nu}{2}, \frac{\nu-2}{2}) & (t\mathcal{W}, t\mathcal{R}, it\mathcal{W}, it\mathcal{R}, F, F\mathcal{R}, iF\mathcal{R}), \end{cases}$$

where $\boldsymbol{\Sigma}$ is the 1×1 expected value and $\Gamma(n/2, n/2)$ and $\Gamma^{-1}(\nu/2, (\nu-2)/2)$ are mean-zero Gamma and Inverse Gamma distributions, respectively. The variance and excess kurtosis of both the $\Gamma(n/2, n/2)$ and the $\Gamma^{-1}(\nu/2, (\nu-2)/2)$ increases, if the d.o.f. parameters n and ν decrease. That is, the distributions get "wider" and more "fat-tailed", which carries over to the distribution of the one-dimensional $\mathbf{R}_t$. This is economically important because if we are interested in predicting $\mathbf{R}_{t+1}$, simple point predictions are of limited use without an accompanying statement about the distributional properties of this forecast, either in the form of confidence bands or by stating variance and tail-measures. Because of the outlined interpretation, we expect the d.o.f. parameters to fall during times of crisis. Thus forecasts of the d.o.f. parameters could also be used to identify upcoming market turmoil.

8. Stollenwerk, 2023: $\mathbf{R} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}} \boldsymbol{\mathcal{K}}_{\mathcal{D}} \mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}} \mathbf{C}^\top$.

Distribution	$2\nabla^{\mathbf{n}}$ or $2\nabla^{n}$
Wishart	$-p\log(2) - \sum_{j=1}^{p}\psi(\frac{n-j+1}{2}) + \log\lvert\mathbf{\Omega}^{-1}\mathbf{R}\rvert$
Riesz	$-\log(2) - \boldsymbol{\psi}(\frac{\mathbf{n}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{\Omega}^{-1}\mathbf{C}_{R})$
Inv.Wishart	-
Inv.Riesz	-
t-Wishart	$p\psi(\frac{\nu+pn}{2}) - \sum_{j=1}^{p}\psi(\frac{n-j+1}{2}) - \log\lvert\mathbf{\Omega}^{-1}\mathbf{R}\rvert - p\log(1+\operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))$
t-Riesz	$\psi(\frac{\nu+p\bar{\mathbf{n}}}{2}) - \boldsymbol{\psi}(\frac{\mathbf{n}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{\Omega}^{-1}\mathbf{C}_{R}) - \log(1+\operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))$
Inv.t-Wishart	$\psi(\frac{n+p\nu}{2}) - \psi(\frac{n}{2}) - \log(1+\operatorname{tr}(\mathbf{\Omega}\mathbf{R}^{-1}))$
Inv.t-Riesz	$\psi(\frac{n+p\bar{\boldsymbol{\nu}}}{2}) - \psi(\frac{n}{2}) - \log(1+\operatorname{tr}(\mathbf{\Omega}\mathbf{R}^{-1}))$
F	$\sum_{j=1}^{p}\psi(\frac{\nu+n-j+1}{2}) - \psi(\frac{n-j+1}{2}) + \log\lvert\mathbf{R}\rvert - \log\lvert\mathbf{H}\rvert$
F-Riesz	$\overleftarrow{\boldsymbol{\psi}}(\frac{\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}}}{2}) - \boldsymbol{\psi}(\frac{\mathbf{n}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{R}) - 2\log\operatorname{vecd}(\mathbf{C}_{H})$
Inv.F-Riesz	$\boldsymbol{\psi}(\frac{\mathbf{n}+\boldsymbol{\nu}}{2}) - \boldsymbol{\psi}(\frac{\mathbf{n}}{2}) - 2\log\operatorname{vecd}(\mathbf{C}_{\Omega}) + 2\log\operatorname{vecd}(\mathbf{C}_{J})$
Distribution	$2\nabla^{\boldsymbol{\nu}}$ or $2\nabla^{\nu}$
Wishart	-
Riesz	-
Inv.Wishart	$-p\log(2) - \sum_{j=1}^{p}\psi(\frac{\nu-j+1}{2}) + \log\lvert\mathbf{\Omega}\mathbf{R}^{-1}\rvert$
Inv.Riesz	$-\log(2) - \overleftarrow{\boldsymbol{\psi}}(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{R}^{-1}\mathbf{C}_{\Omega})$
t-Wishart	$\psi(\frac{\nu+pn}{2}) - \psi(\frac{\nu}{2}) - \log(1+\operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))$
t-Riesz	$\psi(\frac{\nu+p\bar{\mathbf{n}}}{2}) - \psi(\frac{\nu}{2}) - \log(1+\operatorname{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))$
Inv.t-Wishart	$p\psi(\frac{n+p\nu}{2}) - \sum_{j=1}^{p}\psi(\frac{\nu-j+1}{2}) + \log\lvert\mathbf{\Omega}\mathbf{R}^{-1}\rvert - p\log(1+\operatorname{tr}(\mathbf{\Omega}\mathbf{R}^{-1}))$
Inv.t-Riesz	$\psi(\frac{n+p\bar{\boldsymbol{\nu}}}{2}) - \overleftarrow{\psi}(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{R}^{-1}\mathbf{C}_{\Omega}) - \log(1+\operatorname{tr}(\mathbf{\Omega}\mathbf{R}^{-1}))$
F	$\sum_{j=1}^{p}\psi(\frac{\nu+n-j+1}{2}) - \psi(\frac{\nu-j+1}{2}) + \log\lvert\mathbf{\Omega}\rvert - \log\lvert\mathbf{H}\rvert$
F-Riesz	$\overleftarrow{\boldsymbol{\psi}}(\frac{\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}}}{2}) - \overleftarrow{\boldsymbol{\psi}}(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) + 2\log\operatorname{vecd}(\mathbf{C}_{\Omega}) - 2\log\operatorname{vecd}(\mathbf{C}_{H})$
Inv.F-Riesz	$\boldsymbol{\psi}(\frac{\mathbf{n}+\boldsymbol{\nu}}{2}) - \overleftarrow{\boldsymbol{\psi}}(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) - 2\log\operatorname{vecd}(\mathbf{C}_{R}) + 2\log\operatorname{vecd}(\mathbf{C}_{J})$

Table 3: Scores w.r.t. $\mathbf{n}$, n, $\boldsymbol{\nu}$ and ν of all distributions for RCs. $\mathbf{C}_H$ and $\mathbf{C}_J$ denote the lower Cholesky factors of $\mathbf{H} = \mathbf{\Omega}+\mathbf{R}$ and $\mathbf{J} = (\mathbf{\Omega}^{-1}+\mathbf{R}^{-1})^{-1}$, respectively. The derivations are straightforward by applying Lemmas 6.10 and 6.12 (appendix) directly on the log of the p.d.f.s given in Stollenwerk, 2023. Also note Lemma 6.11.

Distribution	$4\mathcal{I}^{\mathbf{n}}$ or $4\mathcal{I}^{n}$	$4\mathcal{I}^{\boldsymbol{\nu}}$ or $4\mathcal{I}^{\nu}$
Wishart	$\sum_{j=1}^{p}\psi'(\frac{n-j+1}{2})$	-
Riesz	$\boldsymbol{\psi}'(\frac{\mathbf{n}}{2})$	-
Inv.Wishart	-	$\sum_{j=1}^{p}\psi'(\frac{\nu-j+1}{2})$
Inv.Riesz	-	$\overleftarrow{\boldsymbol{\psi}}'(\frac{\overleftarrow{\boldsymbol{\nu}}}{2})$
t-Wishart	$\sum_{j=1}^{p}\psi'(\frac{n-j+1}{2}) - p^2\psi'(\frac{\nu+pn}{2})$	$\psi'(\frac{\nu}{2}) - \psi'(\frac{\nu+pn}{2})$
t-Riesz	$\boldsymbol{\psi}'(\frac{\mathbf{n}}{2}) - \psi'(\frac{\nu+p\bar{\mathbf{n}}}{2})$	$\psi'(\frac{\nu}{2}) - \psi'(\frac{\nu+p\bar{\mathbf{n}}}{2})$
Inv.t-Wishart	$\psi'(\frac{n}{2}) - \psi'(\frac{n+p\nu}{2})$	$\sum_{j=1}^{p}\psi'(\frac{\nu-j+1}{2}) - p^2\psi'(\frac{n+p\nu}{2})$
Inv.t-Riesz	$\psi'(\frac{n}{2}) - \psi'(\frac{n+p\bar{\boldsymbol{\nu}}}{2})$	$\overleftarrow{\psi}'(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) - \psi'(\frac{n+p\bar{\boldsymbol{\nu}}}{2})$
F	$\psi'(\frac{n-j+1}{2}) - \sum_{j=1}^{p}\psi'(\frac{\nu+n-j+1}{2})$	$\psi'(\frac{\nu-j+1}{2}) - \sum_{j=1}^{p}\psi'(\frac{\nu+n-j+1}{2})$
F-Riesz	$\boldsymbol{\psi}'(\frac{\mathbf{n}}{2}) - \overleftarrow{\boldsymbol{\psi}}'(\frac{\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}}}{2})$	$\overleftarrow{\boldsymbol{\psi}}'(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) - \overleftarrow{\boldsymbol{\psi}}'(\frac{\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}}}{2})$
Inv.F-Riesz	$\boldsymbol{\psi}'(\frac{\mathbf{n}}{2}) - \boldsymbol{\psi}'(\frac{\mathbf{n}+\boldsymbol{\nu}}{2})$	$\overleftarrow{\boldsymbol{\psi}}'(\frac{\overleftarrow{\boldsymbol{\nu}}}{2}) - \boldsymbol{\psi}'(\frac{\mathbf{n}+\boldsymbol{\nu}}{2})$

Table 4: Fisher information matrices w.r.t. $\mathbf{n}$, n, $\boldsymbol{\nu}$ and ν directly obtained by applying Lemma 6.12 (appendix) on the scores in Table 3.

3.3 Summary

To summarize, we collect all model parameters in the vector

$$\boldsymbol{\delta}_{\mathcal{D}} = (\text{vech}(\boldsymbol{\Xi})^{\top}, a_1, a_2, b_1, b_2, b_3, \boldsymbol{\vartheta}_{\mathcal{D}}^{\top})^{\top},$$

such that we can write

$$\mathbf{R}_t|\mathcal{F}_{t-1} \sim \mathcal{D}\left(\boldsymbol{\Sigma}_t(\boldsymbol{\delta}_{\mathcal{D}}), \boldsymbol{\theta}_{\mathcal{D},t}(\boldsymbol{\delta}_{\mathcal{D}})\right).$$

Both hyperparameters $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\theta}_{\mathcal{D},t}$ are dependent on all model parameters $\boldsymbol{\delta}_{\mathcal{D}}$ since their scaled scores on both $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\theta}_{\mathcal{D},t}$.

With our assumed specifications for $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\theta}_t$, we believe that we have made the time-variation of the parameter space as rich as feasible while avoiding the curse of dimensionality. Thus, the remaining differences in fit and forecasting ability reflect mainly the appropriateness of the distributional assumption.

4 Empirical Application

4.1 Data

Our original data are one-minute close prices from all trading days from 01 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. So our data can be described as previous tick interpolation on a fixed one-minute grid. We acquired the data from Quantquote[9], who combine, clean and process data directly obtained from various exchanges, where the biggest are NYSE, NASDAQ and AMEX[10].

The aim is to produce the longest possible time-series of accurately estimated daily integrated covariance estimators. We exclude dates before 01 January 2002, because the NYSE fully implemented decimal pricing in 2001[11] and there are numerous other trading irregularities during 2001[12]. This leaves 4808 trading days. We then exclude stocks that have not been traded on one of the remaining days in the sample, which leaves 465 of 983 stocks. We only keep observations from official trading hours to be consistent across trading days. We then choose the 100 stocks with the most one-minute close price observations. Of those, the one with the least observations has, on average, 385.18 one-minute close price observations per trading day. Since the typical trading day has 390 minutes, on average less than five close prices are missing per day. Excluding illiquid stocks is common practice in creating time-series of RCs (see e.g. Lunde, Shephard, and Sheppard, 2016). While this procedure biases the sample towards stocks that were very liquid over the entire sample period[13], it does ensure that the integrated covariance estimates are accurate for those stocks included.

We follow Opschoor et al. (2018) and Blasques et al. (2021) and use five-minute returns to construct the 100-dimensional RCs. In particular, we average for each trading day the five distinct RCs, obtained from constructing five-minute log-

9. The Caltech Quantitative Finance Group recommends the company, see http://quant.caltech.edu/historical-stock-data.html.

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

11. On 29 January 2001 to be precise.

12. For example the days surrounding the terrorist attacks on 11 September 2001 and "computer systems connectivity problems" on 08 June 2001.

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

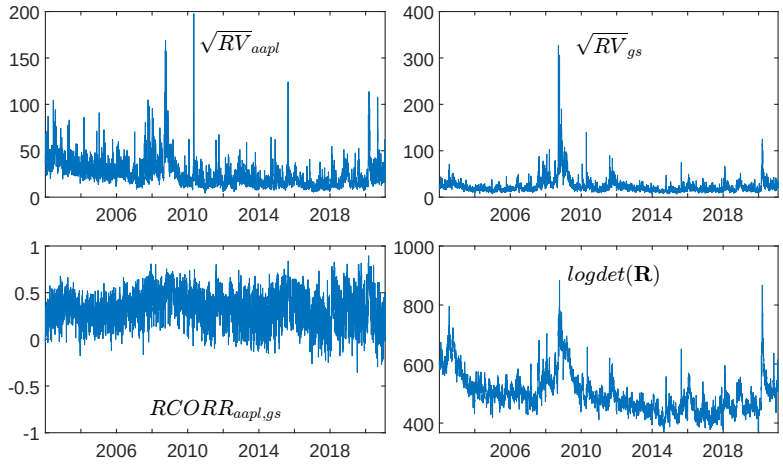

Figure 2: Top row: Annualized realized volatilities of Apple (aapl) and Goldman Sachs (gs), i.e. the square root of two of the elements on the main diagonal of the 100 asset $\mathbf{R}_t$ for the complete sample (01 January 2002 to 05 February 2021). Bottom row: Realized correlation between Apple and Goldman Sachs $RCORR_{aapl,gs} = \mathbf{R}_{aapl,gs}/(\sqrt{RV_{aapl}}\sqrt{RV_{gs}})$ and the natural logarithm of the determinant of the 100 assets $\mathbf{R}_t$ over time.

return vectors for each of the five distinct five-minute grids over the trading day.[14] This estimator is known as the subsampling realized covariance matrix and has been introduced by Zhang, Mykland, and Aït-Sahalia (2005) (see also Sheppard, 2012). It has the advantage of being more efficient than the simple realized covariance matrix since it uses all our data, not just the data of one of the grids. Furthermore, it produces positive definite matrices, even for high cross-sectional dimensions and low sampling frequencies.[15]

For a view of the data, see Figure 2, which shows the annualized realized volatility for Apple (appl) and Goldman Sachs (gs), as well as their realized correlation and the log-determinant of the 100-asset RCs. We see that the spikes in volatility are of similar magnitude for Apple and Goldman Sachs in the recent COVID-19-induced market turmoil, while the global financial crises of 2008/2009 caused volatility to spike much higher for Goldman Sachs than for Apple. The dot-com crisis (early 2000s), on the other hand, causes more volatility for Apple. We see that correlations are mainly positive and more stable around crisis periods.

14. The five time-grids start (on a typical trading day) at 09:00, 09:01, 09:02, 09:03 and 09:04, respectively.

15. On a typical trading day we have 390/5=70 intraday return vectors on a five-minute grid. This allows for a maximum 70 assets to generate positive definite RCs. With subsampling, however, RCs are based on 385 five-minute return vectors, so for up to 385 assets the resulting RCs are positive definite.

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-19, global financial crisis).

From the 100-dimensional dataset described, we randomly choose a 5, 10, and 25-dimensional principal submatrix, as well as the three principal submatrices corresponding to companies with SIC codes of the (1) Financial, Insurance and Real Estate, (2) Mining and (3) High-End Manufacturing division, for a total of six datasets. The division-specific datasets are chosen to investigate whether the tail-homogeneous distributions better fit more homogeneous data.

4.2 Estimation of the GAS Models

Since $\mathbb{E}[\triangle^{\Sigma}_{\mathcal{D},t}|\mathcal{F}_{t-1}] = \mathbf{0}$, it is easy to show (see equation (7)) that

$$\mathbb{E}[\mathbf{R}_t] = \boldsymbol{\Xi}. \tag{10}$$

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

$$\widehat{\boldsymbol{\Xi}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t, \tag{11}$$

and we estimate the rest of the parameters in a second step via standard numerical maximum likelihood estimation, conditional on our estimate for $\boldsymbol{\Xi}$. As in the previous chapter, we follow the algorithm proposed in Blasques et al. (2021) to optimize over the asset ordering for Riesz-type distributions.

Note that the overall number of parameters that must be estimated via numerical maximum likelihood is constant for Wishart-type distributions and is of order $\mathcal{O}(p)$ for Riesz-type distributions. This is because the intercepts of the d.o.f. recursions in the Riesz-type distributions ($\boldsymbol{\xi}^{\mathbf{n}}$ and/or $\boldsymbol{\xi}^{\boldsymbol{\nu}}$) contain p elements, while they are scalars for Wishart-type distributions.

In our empirical analysis, we noticed that the estimation of the dynamic d.o.f. recursion parameters (the a's and b's in equations (8) and (9)) is very sensitive to the chosen starting values since the likelihood is relatively flat in their directions. A "bottom-up" estimation method that proved to recover sensible esti-

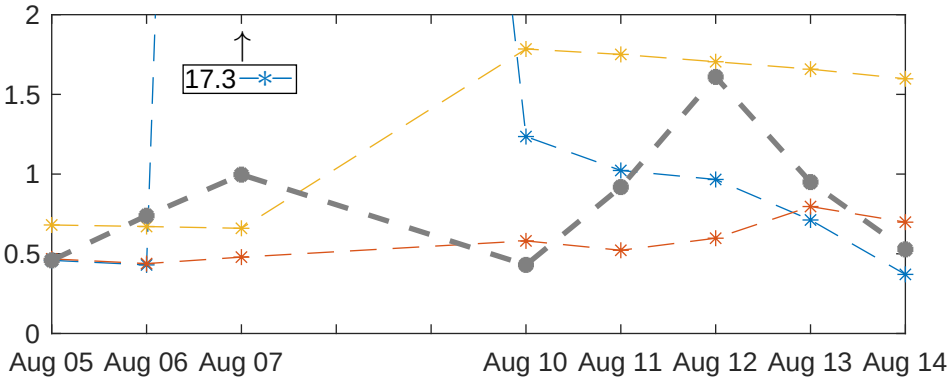

Figure 3: Time frame: 05 August 2015 - 14 August 2015. Blue line: Realized variance of American Express (AXP). Grey line: Median realized variance of the 15 stocks in *Financial* dataset. Both the yellow and the red line assume the best-fitting Inverse t-Riesz distribution ($\mathbf{R}_t|\mathcal{F}_{t-1} \sim it\mathcal{R}(\mathbf{\Sigma}_t, \mathbf{n}, \boldsymbol{\nu})$) and depict the estimated conditional expected variance $(\widehat{\mathbf{\Sigma}}_t)_{33}$.) Yellow line: scalar-BEKK model ($\mathbf{\Sigma}_{t+1} = (1-a-b)\mathbf{\Xi} + a\mathbf{R}_t + b\mathbf{\Sigma}_t$). Red line: scalar-GAS model ($\mathbf{\Sigma}_{t+1} = (1-a-b)\mathbf{\Xi} + a_1\mathbf{\Sigma}_t \triangle^{\Sigma}_{it\mathcal{R},t}\mathbf{\Sigma}_t + a_2 \mathrm{tr}\left(\mathbf{\Sigma}_t \triangle^{\Sigma}_{it\mathcal{R},t}\right)\mathbf{\Sigma}_t + b\mathbf{\Sigma}_t$).

mates relatively reliably is to first estimate the restricted model with constant d.o.f. parameters, that is, restricting all present a and b d.o.f. recursion parameters $(a^{\mathbf{n}}, a^{n}, a^{\boldsymbol{\nu}}, a^{\nu}, b^{\mathbf{n}}, b^{n}, b^{\boldsymbol{\nu}}, b^{\nu})$ to be equal to zero. This provides good estimates for the rest of the parameters (i.e. for $\mathbf{\Xi}$, a_1, a_2, b_1, b_2 and b_3 and for the intercepts in the d.o.f. recursions $\boldsymbol{\xi}^{\mathbf{n}}$, $\boldsymbol{\xi}^{\boldsymbol{\nu}}$, ξ^{n} and ξ^{ν}). Then, we set the obtained estimates as starting values for a second restriction-less optimization. As a useful by-product, we receive the estimated restricted GAS models with constant d.o.f. parameters, which we can compare via likelihood ratio test and forecast comparison to the non-restricted versions to determine the importance of time-varying d.o.f. parameters.

4.3 Illustration of GAS Advantages in $\mathbf{\Sigma}_t$ Updating Equation

In Figure 3, we plot an illustration of the advantages of scalar-GAS model over scalar-BEKK dynamics in the updating equation for $\mathbf{\Sigma}_t$. There, we take the 15-dimensional "Financial" dataset, assume the best-fitting distribution (Inverse t-Riesz, see Section 4), and consider the estimated model with scalar-BEKK and scalar-GAS dynamics for $\mathbf{\Sigma}_t$. The scalar-GAS model does not feature the HAR-type structure (only $\mathbf{\Sigma}_{t-1}$ goes into updating $\mathbf{\Sigma}_t$) and has constant d.o.f. parameters to make the comparison as fair as possible. The blue line corresponds to the

realized variance (RV) of American Express (AXP), and the yellow and red lines correspond to the estimated expected conditional RV $(\hat{\mathbf{\Sigma}}_t)_{33}$ of the scalar-BEKK and scalar-GAS model, respectively. The gray line depicts the median RV of the 15 financial assets. On 07 August 2015, "reports emerged activist investor Value Act Capital had acquired an undisclosed stake in the credit card issuer with plans to pursue shareholder-friendly changes at the company."[16] This constitutes an idiosyncratic one-day event, which only influences the RV of AXP on the 07 August 2015, and should not affect the financial system. Consequently, the conditional expected RV of AXP should not be updated strongly. However, the scalar-BEKK model cannot capture this, as it updates the RV too strongly. This overestimation takes a long time to revert back downwards. In contrast, the GAS model takes into account that the entire RC matrix on 07 August was not a tail event, as only AXP had a tail-realization and updates the conditional expected RV of AXP only slightly upwards. The ability to take into account the *entire* RC in order to update the individual constituent time-series is also visible in the more complex updating dynamics of the scalar-GAS model in the relatively calm period from 10 August to 14 August. The median RV increased relatively much from 11 to 12 August. The GAS model takes this into account and updates upwards its expected RV of AXP from 12 August to 13 August, even while AXP's RV dropped a little. The BEKK model on the other hand just incorporates the decrease in RV of AXP and updates its expected RV for AXP downwards accordingly.

In general, for fat-tailed distributions, GAS models inherently down-weight the impact of extreme realizations of the RCs on the updating process since extreme realizations are less unexpected by fat-tailed distributions and thus yield less extreme score realizations. This type of modeling behavior has also been advocated for in the literature on modeling RCs by e.g. Bollerslev, Patton, and Quaedvlieg (2018), whose *dynamic attenuation model* down-weights the impact of extreme RC realizations by incorporating the fact that they are more inaccurate estimates of integrated covariance.

16. See https://www.nasdaq.com/articles/financial-sector-update-08072015-axphtgcalex-2015-08-07.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart (GHJK)	-14742	-9977	1784	64253	349331	348292
Wishart	-13984	-8426	5831	83409	384626	382252
Riesz	-10915	-4285	18362	112218	462677	463780
Inv.Wishart	-8228	6367	46776	169822	634566	645663
Inv.Riesz	-6340	8319	52720	182317	658765	669844
t-Wishart	-3290	8612	39554	161069	496899	495073
t-Riesz	-1392	10657	48182	184176	554587	554640
Inv.t-Wishart	1510	15169	63864	216533	690170	702292
Inv.t-Riesz	3068	16519	69474	228078	708591	721055
F (OJLV)	-4231	5197	44130	161781	618250	626190
F	-3076	7132	49361	178930	650237	657571
F-Riesz	2691	15007	68534	213218	716322	726208
Inv.F-Riesz	1877	14259	66107	209555	710940	719874

Table 5: Log-likelihood values for the estimated restricted GAS model where d.o.f. parameters are restricted to be constant. As competitor models we have (GHJK), which resembles the model of Gorgi et al. (2019), and (OJLV) resembles the one of Opschoor et al. (2018). 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.

4.4 In-Sample Fit

For the in-sample fit comparison we add two competitor models, namely those of the only other two papers featuring GAS models for RCs. These are the model by Gorgi et al. (2019) based on the Wishart distribution and the one by Opschoor et al. (2018) based on the F distribution. Note that our model does not model the daily return vectors, but both of these competitors do. In order to keep comparability we drop the daily return vector equations from their models and only take their RC specification. Furthermore, note that the main difference to our model is that Gorgi et al. (2019) and Opschoor et al. (2018) take the Wishart distribution FIM to scale the score. Thus, their models are nested by our restricted GAS model. That is, if we impose the restriction that $\alpha_2 = 0$ in equation (7), we arrive at their models.[17]

Table 5 displays the log-likelihood values of our estimated restricted GAS model, i.e. the model where the d.o.f.s are restricted to be constant, for the different

17. Actually the model of Gorgi et al. (2019) does not feature the HAR structure, but we impose it for their model to keep the comparison fair.

datasets and distributions. The Bayes information criterion (BIC) rankings are equal 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 is of order p^2.

The first thing to note is that our restricted GAS model, i.e. including the a_2 parameter, assuming the Wishart and F distribution do indeed improve the fit of significantly over the models of Gorgi et al. (2019) and Opschoor et al. (2018) ($a_2 = 0$). When comparing the fit across distributions, note that nested distributions must have a lower estimated log-likelihood values than the nesting ones. However, it is not clear if the differences are statistically significant and how non-nested distributions, e.g. the (Inverse) F-Riesz versus the (Inverse) t-Riesz distributions compare to each other. The Inverse t-Riesz and the F-Riesz distributions emerge as the clear winners. The former wins the four lower-dimensional datasets, and the latter wins the two 25-dimensional datasets. Both distributions are very close in terms of fitted log-likelihood values, thus the difference in terms of fit between tail-homogeneity (Inverse t-Riesz) and tail-heterogeneity (F-Riesz) is very slim. The gap favoring the Inverse t-Riesz distribution is larger for the industry-specific Mining and Finance datasets. This aligns with our economic intuition that having homogeneous tails is beneficial in more homogeneous datasets. The F-Riesz distribution has the advantage for both high-dimensional (25 assets) datasets, suggesting that in higher dimensions, the tail-homogeneity imposed by the Inverse t-Riesz distribution could be too restrictive. The Riesz and its special case, the Wishart distribution, are unequivocally the worst-fitting distributions.

Now we return to our unrestricted GAS model, i.e. the one where the d.o.f. parameters are assumed to be time varying as well. Table 6 displays the corresponding log-likelihood values. Comparing this table to the one for our restricted GAS model, it becomes clear that the likelihood-gains from assuming time-varying d.o.f. parameters are very small. Furthermore, the ranking across distributions stays the same as for the restricted GAS model. Thus, it is not that some distributions gain more from time-varying d.o.f.s than others. This is surprising since one could think that the worst fitting distributions could gain more from dynamically changing their shape via time-varying d.o.f. parameters than those distributions whose shape fits the RC data already quite well.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	-13814	-8409	6501	83541	386600	387104
Riesz	-10741	-4230	18813	112455	465362	469443
Inv.Wishart	-6979	6425	46776	172471	635234	646137
Inv.Riesz	-6053	8342	52727	182444	658864	669930
t-Wishart	-3113	8870	40283	162996	500667	501338
t-Riesz	-1241	10760	48554	184606	557429	559547
Inv.t-Wishart	1641	15351	64294	217966	692004	704941
Inv.t-Riesz	3157	16590	69656	228469	709889	722258
F	-3030	7213	49563	180031	651463	658949
F-Riesz	2776	15044	68572	213372	716891	726828
Inv.F-Riesz	2014	14356	66405	210311	712416	721365

Table 6: Log-likelihood values for the estimated GAS models with time-varying d.o.f. parameters 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 gray-scaling in between. Largest values in red.

It is noteworthy that the Inverse t-Wishart distribution with only two d.o.f. parameters[18] provides a reasonably good fit across datasets and gets closer to the best-fitting distributions with increasing cross-sectional dimension. It is thus a natural candidate for empirical applications on vast-dimensional (say 50 or more assets) RCs, where, for the Riesz-type distributions, the number of d.o.f. recursion parameters (of order p) becomes problematic for numerical maximum likelihood estimation.

Although the likelihood gains from time-varying d.o.f.s are rather small we still want to test statistically whether the data supports the GAS models with the d.o.f. parameters constraint to be constant. To this end we perform the simple likelihood-ratio test with the test statistic

$$\lambda_{LR} = -2\left(\sum_{t=1}^{T}\log p_{\mathcal{D}}\left(\boldsymbol{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_{\mathcal{D}}^{R}),\boldsymbol{\theta}_{\mathcal{D},t}(\widehat{\boldsymbol{\vartheta}}_{\mathcal{D}}^{R})\right)-\sum_{t=1}^{T}\log p_{\mathcal{D}}\left(\boldsymbol{\Sigma}_t(\widehat{\boldsymbol{\vartheta}}_{\mathcal{D}}),\boldsymbol{\theta}_{\mathcal{D},t}(\widehat{\boldsymbol{\vartheta}}_{\mathcal{D}})\right)\right),$$

where $\widehat{\boldsymbol{\vartheta}}_{\mathcal{D}}^{R}$ is the parameter vector with the non-intercept d.o.f. recursion parame-

18. In total, the GAS models with Wishart-type distributions contain only 10 to 15 parameters (depending on whether they feature one or two d.o.f. parameters) that have to be estimated via numerical maximum likelihood.

ters restricted to be equal to zero. As $T \to \infty$,

$$\lambda_{LR} \sim \chi^2_d,$$

where $d = \dim(\boldsymbol{\vartheta}_{\mathcal{D}}) - \dim(\boldsymbol{\vartheta}^R_{\mathcal{D}})$ is the difference of dimensionality between the restricted and unrestricted parameter vector. Across all distributions and datasets the median (largest) p-value we obtain equals 0 (0.0009) up to computer precision. Thus, even though the likelihood gains are small from assuming time-varying d.o.f. parameters, we can still reject the models where the d.o.f. parameters are constant. This provides strong evidence in favor of time-varying d.o.f. parameters.

Next, we are interested in the time-evolution of the d.o.f. parameters to determine if it makes sense economically. For this, we present the estimated n_t and ν_t of all Wishart-based distributions from our estimated GAS model for the ten-dimensional dataset in Figure 4. The time-series are normed to their first values to make them comparable. We see substantial time-variation, supporting the importance of time-varying d.o.f. parameters. Furthermore, we see that the estimated $\hat{\nu}_t$ does indeed drop down substantially, making the distributions "wider" and more "fat-tailed", in economically sensible time-periods, for example, during the 2008 financial crisis, the 2015 European sovereign debt crisis and the 2020 COVID-19 crises.

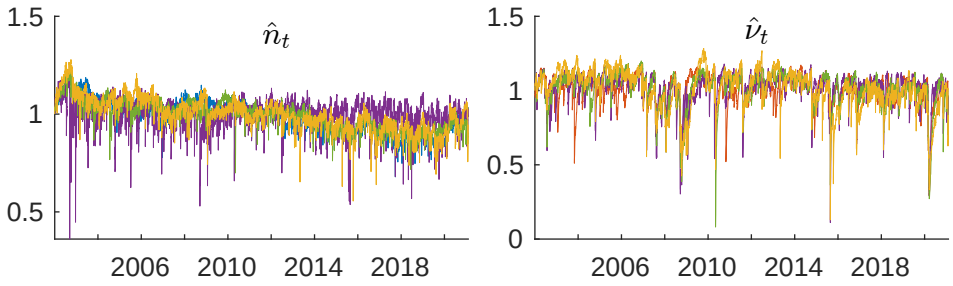

Figure 4: Estimated dynamic d.o.f. parameters $\hat{n}_t$ and $\hat{\nu}_t$ from estimation of our GAS model describe in Section 3 for all distributions for the ten-dimensional dataset described in 4.1. All time-series are normed to the first estimated d.o.f. The color coding is: Wishart, Inv.Wishart, t-Wishart, Inv.t-Wishart, F. The Riesz-type distributions have been left out for clarity of the plot, but their average d.o.f. vectors behave very similarly to their Wishart-based counterpart d.o.f. parameters.

Next, we look at the model parameter estimates. Table 7 shows the estimated score parameters $\hat{a}_1$ and $\hat{a}_2$. All are highly significant. It is interesting to note that the estimated parameters are similar across distributions. Furthermore, we see that the higher the cross-sectional dimension p, the smaller the estimated parameters, which, according to Pakel et al. (2021), is estimation bias caused by the targeting estimator for $\boldsymbol{\Xi}$. The estimated GARCH (b_1, b_2 and b_3) and persistence ($b_1 + b_2 + b_3$) parameters are highly significant as well and very close to each other across cross-sectional dimensions and distributions, with the lowest (highest) estimated persistence amounting to 0.983 (0.9994). See Table 13 in Section of the appendix. This aligns with the empirical observation that realized (co)variances exhibit persistent time-series behavior.

For the estimated score parameters of the d.o.f. parameter recursions, see Table 8. We note that even with our "bottom-up" estimation method described in Section 4.2 we can, in a few cases, not come up with sensible estimates. However, in the overwhelming majority of cases, we obtain sensible estimates, that is, the estimated d.o.f. recursions exhibit high persistence, as for the $\boldsymbol{\Sigma}_t$ recursion parameters. For these cases, estimated d.o.f. recursion parameters, that is, the intercepts, the score parameters, and the GARCH parameters are significant at the 1% level. We thus think that insignificant results are only due to the technical difficulty of obtaining reasonable estimates. We tried multiple solvers, solver specifications, and starting points but have yet to develop a robust estimation method. As we can see in Table 8, the F, F-Riesz, and Inverse F-Riesz distributions are the most difficult to estimate as the solver does not come up with sensible estimates for the d.o.f.-recursions most often for these distributions. In our opinion, this is an argument in favor of the (Inverse) t-Riesz or (Inverse) t-Wishart distribution, which are analytically (see the score and FIM derivations in the appendix) and numerically more tractable.

Finally, we estimated a more general GAS model for the Wishart distribution for the five-dimensional "Random" dataset, where we generalized the dynamics in the $\boldsymbol{\Sigma}_t$-equation to diagonal matrices instead of scalars. The Wishart, providing the worst fit for all distributions, should benefit most from such a generalization. We wanted to investigate if the significance of our dynamic d.o.f. specification comes from a too restrictive specification in the mean-shifting process $\boldsymbol{\Sigma}_t$. This

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			$\hat{a}_1 \times 100$			
Wishart	1.610	1.193	0.872	0.517	0.340	0.335
Riesz	1.135	0.809	0.588	0.336	0.236	0.218
Inv.Wishart	0.197	0.541	0.332	0.157	0.167	0.166
Inv.Riesz	0.121	0.480	0.299	0.113	0.144	0.155
t-Wishart	0.998	0.767	0.651	0.419	0.341	0.325
t-Riesz	0.813	0.663	0.527	0.292	0.247	0.233
Inv.t-Wishart	0.711	0.575	0.456	0.302	0.220	0.210
Inv.t-Riesz	0.636	0.527	0.398	0.258	0.192	0.185
F	0.890	0.704	0.536	0.332	0.248	0.235
F-Riesz	0.477	0.371	0.300	0.207	0.168	0.156
Inv.F-Riesz	0.548	0.416	0.344	0.221	0.169	0.166
			$\hat{a}_2 \times 100$			
Wishart	0.607	0.517	0.249	0.204	0.085	0.080
Riesz	0.791	0.630	0.346	0.251	0.103	0.097
Inv.Wishart	0.328	0.418	0.212	0.149	0.067	0.071
Inv.Riesz	0.399	0.435	0.237	0.149	0.074	0.076
t-Wishart	4.249	4.198	3.370	3.703	2.146	2.221
t-Riesz	4.084	4.285	3.283	3.950	2.238	2.184
Inv.t-Wishart	3.598	3.474	2.851	3.192	1.600	1.756
Inv.t-Riesz	3.409	3.231	2.768	3.134	1.493	1.686
F	0.623	0.477	0.276	0.179	0.089	0.086
F-Riesz	1.183	0.915	0.490	0.293	0.141	0.138
Inv.F-Riesz	1.121	0.884	0.478	0.280	0.133	0.129

Table 7: Estimated scaled score parameters $\hat{a}_1 \times 100$ and $\hat{a}_2 \times 100$. All estimates are highly significant. The median (smallest) t-statistic for $\hat{a}_1$ is 114 (68), and for $\hat{a}_2$ is 150 (42).

does not seem to be the case. In each diagonal parameter matrix, the estimated parameters are very close to each other, and the d.o.f. recursion parameters stayed highly significant.

4.5 Out-of-Sample Forecasting Performance

We re-estimate the models every ten trading days on a rolling window of 1250 observations (roughly five years of data). The forecasting window starts at observation 1251 (18 December 2006) and ends on 05 February 2021.

We take the log-score as the natural loss function since, as Hansen and Du-

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			$\widehat{a^{\mathbf{n}}} \times 100$ ($\widehat{a^{n}} \times 100$)			
Wishart	0.827	0.707	1.633	0.598	1.002	1.235
Riesz	0.863	1.096	1.292	0.507	0.691	0.917
Inv.Wishart	-	-	-	-	-	-
Inv.Riesz	-	-	-	-	-	-
t-Wishart	6.735	8.149	6.658	12.561	8.098	9.925
t-Riesz	1.533	1.500	0.955	1.905	0.829	1.002
Inv.t-Wishart	1.591	1.892	1.076	0.987	1.323	1.183
Inv.t-Riesz	1.824	1.904	0.992	1.093	1.295	1.223
F	13.723	5.032	0.570	0.000	2.893	0.003
F-Riesz	7.017	2.612	2.647	0.842	1.535	1.524
Inv.F-Riesz	2.549	5.600	0.273	0.001	3.422	3.325
			$\widehat{a^{\boldsymbol{\nu}}} \times 100$ ($\widehat{a^{\nu}} \times 100$)			
Wishart	-	-	-	-	-	-
Riesz	-	-	-	-	-	-
Inv.Wishart	3.077	3.595	2.794	2.408	1.230	3.486
Inv.Riesz	6.687	59.902	5.449	1.677	2.503	4.108
t-Wishart	1.071	1.078	1.548	1.816	1.274	2.108
t-Riesz	1.347	1.025	1.396	1.654	1.049	2.392
Inv.t-Wishart	12.924	12.572	14.124	20.952	16.173	19.947
Inv.t-Riesz	8.756	7.097	5.356	1.917	0.645	0.456
F	14.264	11.900	13.637	17.600	10.968	10.406
F-Riesz	9.910	7.857	0.934	4.683	1.093	2.213
Inv.F-Riesz	9.924	9.946	9.327	8.631	6.056	6.188

Table 8: Estimated score parameter $\widehat{a^{\mathbf{n}}} \times 100$ and $\widehat{a^{\boldsymbol{\nu}}} \times 100$ ($\widehat{a^{n}} \times 100$ and $\widehat{a^{\nu}} \times 100$ for Wishart-type distributions). No background color indicates significance at the 1% level, light gray and mid-light gray indicate significance at the 5% and 10% level, dark gray indicates insignificance.

mitrescu (2022) show, coherency between the estimation criterion, that is, the loss function chosen to obtain the estimates, and the actual objective, i.e. the out-of sample quantity we want to forecast, is essential. We only consider the Inverse t-Riesz and the F-Riesz distribution to save computational resources, and since it is evident from the in-sample analysis that these two distributions provide the best out-of-sample forecasting results. Table 9 contains the log-score one-day-ahead forecast losses for the entire forecasting window. The 90% model confidence

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			BEKK			
Inv.t-Riesz	3.70	10.67	-10.86	-53.90	-75.18	-110.57
F-Riesz	3.75	10.85	-10.81	-48.56	-76.43	-109.33
			Restricted GAS			
Inv.t-Riesz	3.13	10.05	-12.52	-57.12	-85.48	-118.53
F-Riesz	3.28	10.46	-12.12	-53.44	-86.20	-119.24
			GAS			
Inv.t-Riesz	3.11	10.05	-12.55	-57.19	-85.57	-118.56
F-Riesz	3.27	10.45	-12.13	-53.51	-86.23	-119.26

Table 9: Average of log-score loss, $-p_{\mathcal{D}}(\mathbf{R}_{t+1}|\widehat{\boldsymbol{\Sigma}}_{t+1}, \hat{\boldsymbol{\theta}}_{\mathcal{D},t+1})$, for the entire forecasting window, where each model is re-estimated every ten trading days. 90% model confidence sets in red. BEKK refers to the scalar-BEKK model of Stollenwerk (2023), Restricted GAS is our GAS model, where the d.o.f. parameters are constant, GAS is our full GAS model.

sets (MCS, see Hansen, Lunde, and Nason, 2011) are shaded in gray.[19]

We see that our unrestricted GAS model with dynamic d.o.f. parameters is the clear winner in this comparison as it is the only member of the MCS for all datasets except the Mining dataset, where also the GAS model with d.o.f. parameters restricted to be constant is a member of the MCS. The scalar-BEKK model, with its simple mean shifting dynamics, is rejected out-of-sample by the data. Furthermore, it is apparent that for the datasets with smaller cross-sectional dimensions p, the Inverse t-Riesz distribution provides the best forecasts. In contrast, for the 25-dimensional dataset, the F-Riesz distribution is the sole member of the MCS. This aligns with the in-sample fit (see Table 6). We conjecture that with increasing cross-sectional dimension, the assumption of tail-homogeneity (Inverse t-Riesz) might become too restrictive since it imposes too much dependence between the different elements in the RCs.

Finally, we consider an economic loss function by computing the realized variance of the forecasted global minimum variance portfolio (GMVP). That is, suppose we want to minimize the portfolio variance given our forecasted RC, $\widehat{\boldsymbol{\Sigma}}_{t+1}$,

19. For calculation of the MCS, we choose 5000 stationary bootstrap replications with block length set equal to the maximum number of consecutive significant partial autocorrelations of the losses. We use the MFE toolbox by Kevin Sheppard for MCS calculation.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			BEKK			
Inv.t-Riesz	1.368	2.046	0.736	1.403	0.687	0.836
F-Riesz	1.369	2.042	0.735	1.375	0.678	0.824
			Restricted GAS			
Inv.t-Riesz	1.368	2.045	0.738	1.441	0.698	0.841
F-Riesz	1.391	2.073	0.758	1.468	0.705	0.859
			GAS			
Inv.t-Riesz	1.368	2.045	0.738	1.440	0.698	0.841
F-Riesz	1.392	2.073	0.757	1.472	0.704	0.859

Table 10: Average GMVP loss, $RV_{GMPV,t+1}$, for the entire forecasting window, where each model is re-estimated every ten trading days. 90% model confidence sets in red. BEKK refers to the scalar-BEKK model of Stollenwerk (2023), Restricted GAS is our GAS model, where the d.o.f. parameters are constant, GAS is our full GAS model.

by choosing the optimal weights,

$$\min_{\mathbf{w}} \mathbf{w}^\top \widehat{\mathbf{\Sigma}}_{t+1} \mathbf{w}, \qquad \text{given } \mathbf{w}^\top \mathbf{1} = 1,$$

where $\mathbf{1}$ is an p-dimensional vector of ones. The optimal weights that solve this minimization problem are given by

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

Now for each forecast the GMVP loss is equal to the actually realized variance if an investor had chosen these weights,

$$RV_{GMPV,t+1} = \hat{\mathbf{w}}^\top \mathbf{R}_{t+1} \hat{\mathbf{w}}.$$

Table 10 contains the average GMVP loss for all datasets and models. We see that for the three datasets of dimension 10 or lower, there emerges no clear winner across time-series specifications since the BEKK, Restricted GAS and GAS all are members of the MCS. Comparing distributions, for these three datasets the Inverse t-Riesz distribution performs better since it is in the MCS for all three times versus one time for the F-Riesz. For the higher dimensional dataset the

picture is more clear. Here the F-Riesz with BEKK dynamics is the sole member of the MCS. However, due to the nonlinear formula of the GMVP loss function it is difficult to speculate which distributional features might lead to the advantage of the F-Riesz for high-dimensional datasets.

5 Conclusion

In this paper, we propose a novel GAS model for time-series of realized covariance matrices. This model features a time-varying expected value matrix and allows for time-varying degree of freedom parameters, thus making all distribution parameters dynamic. We derive the model for all hitherto used probability distributions in order to understand which distributions' score yields the best fit and forecasting performance. In particular we derive all necessary scores and Fisher information matrices for all probability distributions hitherto used for RCs. Our model features an easy-to-compute closed-form formula for the scaled scores, making the model computationally feasible.

Empirically, we find evidence for time-varying degree of freedom parameters. We show that their time-variation is economically interpretable and is in line with the notion that realized covariance matrices become more fat-tailed during crises periods. Our GAS model with time-varying d.o.f. parameters performs best in- and out-of-sample against the competitor with constant d.o.f. parameters using the log-likelihood as the loss-function. Using the economically relevant global minimum variance portfolio loss function we see no advantage for our GAS model. Furthermore, we confirm the finding from the last chapter that the Inverse t-Riesz and F-Riesz distributions exhibit the best in-sample fit for time-series of RCs. No clear winner between these two distributions can be made out, but there seems to be a relation to the cross-sectional dimension. In particular, the F-Riesz is more often preferred in higher-dimensional settings, whereas the t-Riesz outperforms in low- to medium-dimensional settings. Furthermore, we show using an illustrative example that GAS models imply richer dynamic updating properties than traditional models, even for scalar dynamics.

References

Andersen, Torben G., Tim Bollerslev, Peter F. Christoffersen, and Francis X. Diebold. 2006. "Volatility and Correlation Forecasting." In *Handbook of Economic Forecasting,* edited by G. Elliott, C. W. J. Granger, and A. Timmermann, 1:777–878. Elsevier.

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.

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

Blasques, Francisco, Christian Francq, and Sébastien Laurent. 2022. "Quasi Score-Driven Models." *Journal of Econometrics.*

Blasques, Francisco, Siem J. Koopman, and Andre Lucas. 2015. "Information-Theoretic Optimality of Observation-Driven Time Series Models for Continuous Responses." *Biometrika* 102 (2): 325–343.

Blasques, Francisco, Andre Lucas, Anne Opschoor, and Luca Rossini. 2021. "Tail Heterogeneity for Dynamic Covariance-Matrix-Valued Random Variables: The F-Riesz Distribution." Working Paper, Tinbergen Institute Discussion Paper.

Blasques, Francisco, Janneke van Brummelen, Siem J. Koopman, and André Lucas. 2022. "Maximum Likelihood Estimation for Score-Driven Models." *Journal of Econometrics* 227 (2): 325–346.

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.

Chiriac, Roxana, and Valeri Voev. 2011. "Modelling and Forecasting Multivariate Realized Volatility." *Journal of Applied Econometrics* 26 (6): 922–947.

Corsi, Fulvio. 2009. "A Simple Approximate Long-Memory Model of Realized Volatility." *Journal of Financial Econometrics* 7 (2): 174–196.

Cox, David R. 1981. "Statistical Analysis of Time Series: Some Recent Developments." *Scandinavian Journal of Statistics,* 93–115.

Creal, Drew, Siem J. 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.

———. 2013. "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics* 28 (5): 777–795.

Díaz-García, José A. 2013. "A Note on the Moments of the Riesz Distribution." *Journal of Statistical Planning and Inference* 143 (11): 1880–1886.

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.

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.

Gorgi, Paolo, 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.

Gribisch, Bastian, and Jan P. Hartkopf. 2022. "Modeling Realized Covariance Measures with Heterogeneous Liquidity: A Generalized Matrix-Variate Wishart State-Space Model." *Journal of Econometrics.*

Gupta, Arjun K., and Daya K. Nagar. 2000. *Matrix Variate Distributions.* Chapman and Hall/CRC.

Hansen, Peter R., and Elena-Ivona Dumitrescu. 2022. "How Should Parameter Estimation be Tailored to the Objective?" *Journal of Econometrics* 230 (2): 535–558.

Hansen, Peter R., Asger Lunde, and James M. Nason. 2011. "The Model Confidence Set." *Econometrica* 79 (2): 453–497.

Harvey, Andrew C. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series.* Vol. 52. Econometric Society Monographs. Cambridge University Press.

———. 2022. "Score-Driven Time Series Models." *Annual Review of Statistics and Its Application* 9 (1): 321–342.

Harville, David A. 1997. *Matrix Algebra From a Statisticians Perspective.* Springer.

Jin, Xin, and John M. Maheu. 2016. "Bayesian Semiparametric Modeling of Realized Covariance Matrices." *Journal of Econometrics* 192 (1): 19–39.

Kollo, Tõnu, and Dietrich von Rosen. 2005. *Advanced Multivariate Statistics with Matrices.* Springer.

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 and Economic Statistics* 34 (4): 504–518.

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.

———. 2005. *New Introduction to Multiple Time Series Analysis.* Springer.

Magnus, Jan R. 2010. "On the Concept of Matrix Derivative." *Journal of Multivariate Analysis* 101 (9): 2200–2206.

Magnus, Jan R., and Heinz Neudecker. 1980. "The Elimination Matrix: Some Lemmas and Applications." *SIAM Journal on Algebraic Discrete Methods* 1 (4): 422–449.

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

McAleer, Michael, and Marcelo C. Medeiros. 2008. "Realized Volatility: A Review." *Econometric Reviews* 27 (1-3): 10–45.

Murray, Ian. 2016. "Differentiation of the Cholesky Decomposition." Working Paper, arXiv.org.

Noureldin, Diaa, Neil Shephard, and Kevin Sheppard. 2012. "Multivariate High-Frequency-Based Volatility (HEAVY) Models." *Journal of Applied Econometrics* 27 (6): 907–933.

Opschoor, Anne, Pawel Janus, André Lucas, and Dick van Dijk. 2018. "New HEAVY Models for Fat-Tailed Realized Covariances and Returns." *Journal of Business and Economic Statistics* 36 (4): 643–657.

Opschoor, Anne, and André Lucas. 2022. "Time-Varying Variance and Skewness in Realized Volatility Measures." *International Journal of Forecasting.*

Pakel, Cavit, Neil Shephard, Kevin Sheppard, and Robert F. Engle. 2021. "Fitting Vast Dimensional Time-Varying Covariance Models." *Journal of Business and Economic Statistics* 39 (3): 652–668.

Sheppard, Kevin. 2012. "Forecasting High Dimensional Covariance Matrices." In *Handbook of Volatility Models and Their Applications,* edited by L. Bauwens, C. Hafner, and S. Laurent, 103–125. John Wiley and Sons.

Stollenwerk, Michael. 2023. "Probability Distributions for Realized Covariance Matrices." Working Paper, ssrn.com.

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia. 2005. "A Tale of Two Time Scales." *Journal of the American Statistical Association* 100 (472): 1394–1411.

Zhou, Jiayuan, Feiyu Jiang, Ke Zhu, and Wai K. Li. 2019. "Time Series Models for Realized Covariance Matrices based on the Matrix-F Distribution." Working Paper, arXiv.org.

6 Appendix

6.1 Preliminaries

6.1.1 Some 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)

$$\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^\top \otimes \mathbf{X})\text{vec}(\mathbf{Y}), \tag{12}$$

$$\text{tr}(\mathbf{XYZ}) = \text{tr}(\mathbf{YZX}) = \text{tr}(\mathbf{ZXY}), \tag{13}$$

$$\text{tr}(\mathbf{X}^\top\mathbf{Y}) = \text{vec}(\mathbf{X})^\top\text{vec}(\mathbf{Y}) \text{ and} \tag{14}$$

$$\text{tr}(\mathbf{WXYZ}) = \text{vec}(\mathbf{W})^\top\text{vec}(\mathbf{XYZ}) = \text{vec}(\mathbf{W})^\top(\mathbf{X} \otimes \mathbf{Z})\text{vec}(\mathbf{Y}), \tag{15}$$

where for the last equality we used (14) and (12).

6.1.2 Special Functions

Definition 6.1. *Let* $\mathbf{X}$ *be a square matrix, then the function* $\blacktriangle(\mathbf{X})$ *returns a lower triangular matrix by setting all elements of* $\mathbf{X}$ *above the main diagonal equal to zero,*

$$\blacktriangle_{ij}(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij} \text{ for } i \geq j \text{ and} \\ 0 \text{ for } i < j. \end{cases}$$

Definition 6.2. *Let* $\mathbf{X}$ *be a square matrix, then the function* $\bar{\blacktriangle}(\mathbf{X})$ *returns a lower triangular matrix by setting all elements of* $\mathbf{X}$ *above the main diagonal equal to zero and halving all elements on the main diagonal,*

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

That is, $\bar{\blacktriangle}(\mathbf{X}) = \blacktriangle\left(\mathbf{X} - \frac{1}{2}\mathbf{I} \odot \mathbf{X}\right)$.

Definition 6.3. *Let* $\mathbf{X}$ *be a square matrix, then the function* $\tilde{\blacktriangle}(\mathbf{X})$ *returns a symmetric matrix by setting all elements above the main diagonal equal to the elements of* $\mathbf{X}$ *below the diagonal,*

$$\tilde{\blacktriangle}_{ij}(\mathbf{X}) = \begin{cases} \mathbf{X}_{ij} \text{ for } i \geq j, \\ \mathbf{X}_{ji} \text{ for } i < j. \end{cases}$$

That is, $\tilde{\blacktriangle}(\mathbf{X}) = \bar{\blacktriangle}(\mathbf{X}) + \bar{\blacktriangle}(\mathbf{X})^\top$.

6.1.3 Duplication, Elimination, and Commutation Matrices

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

$$\text{vec}(\mathbf{X}) = \mathbf{G}_p \text{vech}(\mathbf{X}), \tag{16}$$

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 vec($\mathbf{X}$) to 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 $\mathbf{G}_p$,

$$\mathbf{G}_p^+ = (\mathbf{G}_p^\top \mathbf{G}_p)^{-1} \mathbf{G}_p^\top, \tag{17}$$

for which obviously

$$\mathbf{G}_p^+ \text{vec}(\mathbf{X}) = \mathbf{G}_p^+ \mathbf{G}_p \text{vech}(\mathbf{X}) = \text{vech}(\mathbf{X}). \tag{18}$$

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)) that the *unique elimination* and *duplication matrices* are given by

$$\text{vec}(\mathbf{Y}) = \mathbf{F}_p^\top \text{vech}(\mathbf{Y}) \text{ and} \tag{19}$$

$$\text{vech}(\mathbf{Y}) = \mathbf{F}_p \text{vec}(\mathbf{Y}). \tag{20}$$

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

$$\text{vec}(\mathbf{Z}^\top) = \mathbf{K}_{pq} \text{vec}(\mathbf{Z}), \tag{21}$$

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

$$(\mathbf{I} + \mathbf{K}) = 2\mathbf{G}\mathbf{G}^+. \tag{22}$$

Furthermore, it holds that

$$\mathbf{G}\mathbf{G}^+ = \mathbf{G}(\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top = (\mathbf{G}(\mathbf{G}^\top \mathbf{G})^{-1} \mathbf{G}^\top)^\top = (\mathbf{G}\mathbf{G}^+)^\top \tag{23}$$

and

$$(\mathbf{G}^+)^\top \mathbf{G}^\top \text{vec}(\mathbf{X}) = ((\mathbf{G}^\top \mathbf{G})^{-1}\mathbf{G}^\top)^\top \mathbf{G}^\top \mathbf{G}\text{vech}(\mathbf{X}) \tag{24}$$

$$= \mathbf{G}\text{vech}(\mathbf{X}) = \text{vec}(\mathbf{X}). \tag{25}$$

For square matrix $\mathbf{X}$ we have (see Magnus and Neudecker, 2019, p.57)

$$\mathbf{G}\mathbf{G}^+\mathbf{X}^{\otimes 2} = \mathbf{X}^{\otimes 2}\mathbf{G}\mathbf{G}^+. \tag{26}$$

For nonsingular matrix $\mathbf{X}$ it holds that (see Lütkepohl, 2005, p. 664 or Magnus and Neudecker (2019), Theorem 3.13)

$$(\mathbf{G}^\top \mathbf{X}^{\otimes 2}\mathbf{G})^{-1} = \mathbf{G}^+\mathbf{X}^{-\otimes 2}(\mathbf{G}^+)^\top. \tag{27}$$

Lemma 6.1. *For scalar α we have*

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

Proof. Using Magnus and Neudecker (1980) Lemma 4.4 (i) and Lemma 4.7 (iv), we have

$$\begin{aligned}
(\mathbf{F}(\mathbf{\Sigma}^{\otimes 2} + \alpha\text{vec}^2(\mathbf{\Sigma})))^{-1} &= (\mathbf{G}^+\mathbf{G}\mathbf{F}(\mathbf{\Sigma}^{\otimes 2} + \alpha\text{vec}^2(\mathbf{\Sigma}))\mathbf{G})^{-1} \\
&= (\mathbf{G}^+\mathbf{G}\mathbf{F}(\mathbf{\Sigma}^{\otimes 2} + \alpha\text{vec}^2(\mathbf{\Sigma}))\mathbf{G})^{-1} \\
&= (\mathbf{G}^+(\mathbf{\Sigma}^{\otimes 2} + \alpha\text{vec}^2(\mathbf{\Sigma}))\mathbf{G})^{-1} \\
&= (\mathbf{G}^\top(\mathbf{\Sigma}^{\otimes 2} + \alpha\text{vec}^2(\mathbf{\Sigma}))\mathbf{G})^{-1}\mathbf{G}^\top\mathbf{G} \qquad (28) \\
&= \mathbf{F}\left(\mathbf{\Sigma}^{-\otimes 2} + \frac{\alpha}{1+\alpha p}\text{vec}^2\left(\mathbf{\Sigma}^{-1}\right)\right)\mathbf{G} \\
&= \mathbf{G}^+\left(\mathbf{\Sigma}^{-\otimes 2} + \frac{\alpha}{1+\alpha p}\text{vec}^2\left(\mathbf{\Sigma}^{-1}\right)\right)\mathbf{G}. \qquad (29)
\end{aligned}$$

Then

$$(28) = (29)$$

$$\Leftrightarrow$$

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

□

6.1.4 Matrix Derivatives

We first want to clarify exactly how we take the derivative of a (matrix) function of a matrix of variables since there is no inherently "right" way to arrange the individual univariate derivatives. Throughout, we follow Magnus (2010), who recommend for an $m \times p$ matrix function $\mathbf{F}$ of an $n \times q$ matrix of variables $\mathbf{X}$ to define the derivative as the $mp \times nq$ matrix

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

In this derivative matrix, for example, in its third row and second column, we have the derivative of the third element of $\text{vec}(\mathbf{F}(\mathbf{X}))$ with respect to the second element in $\text{vec}(\mathbf{X})$. Hence the column vector notation in the nominator ($\text{vec}(\mathbf{F}(\mathbf{X}))$) and the row vector notation in the denominator ($\text{vec}(\mathbf{X})^\top$).

For our scalar-valued log-likelihood function ($m = p = 1$), we have, in accordance with this definition, that its derivative with respect to $\boldsymbol{\Sigma}$ is a row vector,

$$\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}}|\mathbf{R})}{\partial \text{vec}(\boldsymbol{\Sigma})^\top} = \text{vec}(\triangle_{\mathcal{D}}^{\Sigma})^\top,$$

with

$$(\triangle_{\mathcal{D}}^{\Sigma})_{ij} = \frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}}|\mathbf{R})}{\partial (\boldsymbol{\Sigma})_{ij}}, \tag{30}$$

where we omit the time subscript t for readability in this and the following section. However, we still have to consider that $\boldsymbol{\Sigma}$ is a symmetric matrix. That is, for $i \neq j$, an infinitesimally small change in $\boldsymbol{\Sigma}_{(i,j)}$, changes $\boldsymbol{\Sigma}_{(j,i)}$ by the same amount. To

do so, we can use the chain rule,

$$\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}}|\mathbf{R})}{\partial \text{vech}(\boldsymbol{\Sigma})^\top} = \frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}}|\mathbf{R})}{\partial \text{vec}(\boldsymbol{\Sigma})^\top} \frac{\partial \text{vec}(\boldsymbol{\Sigma})}{\partial \text{vech}(\boldsymbol{\Sigma})^\top}.$$

For symmetric $\boldsymbol{\Sigma}$ we have according to Lemma 6.2 (see p. 42) that

$$\frac{\partial \text{vec}\left(\boldsymbol{\Sigma}\right)}{\partial \text{vech}\left(\boldsymbol{\Sigma}\right)^\top} = \mathbf{G},$$

where $\mathbf{G}$ is the duplication matrix. Finally, it is usually the case that the score is defined as a column vector, such that we write

$$\nabla_{\mathcal{D}}^{\Sigma} = \left(\frac{\partial \log p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}}|\mathbf{R})}{\partial \text{vech}(\boldsymbol{\Sigma})^\top}\right)^\top.$$

In summary, for a given distribution $\mathcal{D}$,

> $\nabla_{\mathcal{D}}^{\Sigma}(\nabla_{\mathcal{D}}^{\Omega})$ is the $p(p+1)/2 \times 1$ score *vector* w.r.t. $\boldsymbol{\Sigma}$ ($\boldsymbol{\Omega}$) taking symmetry into account, and
>
> $\triangle_{\mathcal{D}}^{\Sigma}(\triangle_{\mathcal{D}}^{\Omega})$ is the $p \times p$ score *matrix* w.r.t. $\boldsymbol{\Sigma}$ ($\boldsymbol{\Omega}$) *ignoring* symmetry (used in e.g. Theorem 3.1 and Table 1).

Lemma 6.2. *(Magnus and Neudecker, 1980, Lemma 3.8). Let $\mathbf{X}$ be a square matrix of variables. Then*

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

Lemma 6.3. *(Harville, 1997, p. 371). For non-singular symmetric matrix $\mathbf{X}$,*

$$\frac{\partial \text{vech}(\mathbf{X}^{-1})}{\partial \text{vech}(\mathbf{X})^\top} = -\mathbf{G}^+\mathbf{X}^{-\otimes 2}\mathbf{G}.$$

Lemma 6.4. *(Lütkepohl, 1989, Lemma 1 and Murray, 2016, equation (15).) For lower triangular matrix* $\mathbf{C}$ *we have*

$$\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{C}\mathbf{C}^\top)^\top} = \frac{1}{2}(\mathbf{G}^+(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^\top)^{-1} = \mathbf{F}(\mathbf{I}\otimes\mathbf{C})\mathbf{Z}\mathbf{C}^{-\otimes 2}\mathbf{G},$$

where $\mathbf{Z}$ *is a diagonal matrix defined such that for any square matrix* $\mathbf{A}$, $\mathbf{Z}\text{vec}(\mathbf{A}) = \text{vec}(\bar{\blacktriangle}(\mathbf{A}))$ *with* $\bar{\blacktriangle}(\cdot)$ *defined in Definition 6.2. In fact it can be shown that* $\mathbf{Z} = \frac{1}{2}(\mathbf{G}\mathbf{F})^\top(\mathbf{G}\mathbf{F})$.

Lemma 6.5. *Let* $\mathbf{X}$ *be a non-singular symmetric matrix,* $\mathbf{C}$ *its lower Cholesky factor, and* $\mathbf{Y}$ *be a diagonal matrix. Then*

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{X})^\top} &= \mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{X}^{-\otimes 2}\mathbf{G}\\ &= \mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^\top)^{-1},\end{aligned}$$

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{X}^{-1})^\top} &= -\mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{X}^{\otimes 2}\mathbf{G}\\ &= -\mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^\top)^{-1},\end{aligned}$$

$$\begin{aligned}&\frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{X}^{-1})^\top}\\ &\quad= \mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{X}^{\otimes 2}\mathbf{G}\\ &\quad= \mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^\top)^{-1},\end{aligned}$$

$$\begin{aligned}&\frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{X})^\top}\\ &\quad= -\mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{X}^{-\otimes 2}\mathbf{G}\\ &\quad= -\mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^\top)^{-1}.\end{aligned}$$

Proof. We have

$$\begin{aligned}\mathrm{dvec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top) &= \mathrm{vec}(\mathrm{d}\mathbf{C}\mathbf{Y}\mathbf{C}^\top) + \mathrm{vec}(\mathbf{C}\mathbf{Y}\mathrm{d}\mathbf{C}^\top)\\ &= (\mathbf{I}+\mathbf{K}_{pp})\mathrm{vec}(\mathrm{d}\mathbf{C}\mathbf{Y}\mathbf{C}^\top)\\ &= (\mathbf{I}+\mathbf{K}_{pp})(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathrm{vec}(\mathrm{d}\mathbf{C})\\ &= 2\mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top\mathrm{dvech}(\mathbf{C}),\end{aligned}$$

$$\begin{aligned}\mathrm{dvec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}) &= \mathrm{vec}(\mathrm{d}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}) + \mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathrm{d}\mathbf{C}^{-1})\\ &= -\mathbf{K}_{pp}\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}) - \mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})\\ &= -(\mathbf{I}+\mathbf{K}_{pp})\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})\\ &= -(\mathbf{I}+\mathbf{K}_{pp})(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathrm{vec}(\mathrm{d}\mathbf{C})\\ &= -2\mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top\mathrm{dvech}(\mathbf{C}),\end{aligned}$$

where we used (21), (12) and (22), such that

$$\frac{\partial\mathrm{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial\mathrm{vech}(\mathbf{C})^\top} = 2\mathbf{G}\mathbf{G}^+(\mathbf{C}\mathbf{Y}\otimes\mathbf{I})\mathbf{F}^\top \tag{31}$$

and

$$\frac{\partial\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial\mathrm{vech}(\mathbf{C})^\top} = -2\mathbf{G}\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})\mathbf{F}^\top. \tag{32}$$

Furthermore,

$$\begin{aligned}\mathrm{dvech}(\mathbf{X}^{-1}) &= \mathbf{G}^+\mathrm{dvec}(\mathbf{C}^{-\top}\mathbf{C}^{-1})\\ &= -\mathbf{G}^+\mathrm{vec}([\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}]^\top\mathbf{C}^\top) - \mathbf{G}^+\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})\\ &= -\mathbf{G}^+(\mathbf{I}+\mathbf{K}_{pp})\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})\\ &= -\mathbf{G}^+\mathbf{G}\mathbf{G}^+\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})\\ &= -2\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{C}^{-\top}\mathbf{C}^{-1})\mathrm{vec}(\mathrm{d}\mathbf{C})\\ &= -2\mathbf{G}^+(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^\top\mathrm{dvech}(\mathbf{C}).\end{aligned}$$

such that

$$\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X}^{-1})^\top} = -\frac{1}{2}(\mathbf{G}^+(\mathbf{C}^{-\top} \otimes \mathbf{X}^{-1})\mathbf{F}^\top)^{-1}.$$

Finally, recall Lemma 6.4, then the lemma follows by applying the chain rule,

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{X})^\top} &= \frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X}^{-1})^\top}\frac{\partial \text{vech}(\mathbf{X}^{-1})}{\partial \text{vech}(\mathbf{X})^\top} \\ &= \frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X})^\top},\end{aligned}$$

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{X}^{-1})^\top} &= \frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X})^\top}\frac{\partial \text{vech}(\mathbf{X})}{\partial \text{vech}(\mathbf{X}^{-1})^\top} \\ &= \frac{\partial \text{vec}(\mathbf{C}\mathbf{Y}\mathbf{C}^\top)}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X}^{-1})^\top},\end{aligned}$$

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{X}^{-1})^\top} &= \frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X})^\top}\frac{\partial \text{vech}(\mathbf{X})}{\partial \text{vech}(\mathbf{X}^{-1})^\top} \\ &= \frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X}^{-1})^\top},\end{aligned}$$

$$\begin{aligned}\frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{X})^\top} &= \frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X}^{-1})^\top}\frac{\partial \text{vech}(\mathbf{X}^{-1})}{\partial \text{vech}(\mathbf{X})^\top} \\ &= \frac{\partial \text{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})}{\partial \text{vech}(\mathbf{C})^\top}\frac{\partial \text{vech}(\mathbf{C})}{\partial \text{vech}(\mathbf{X})^\top}.\end{aligned}$$

There are two versions for each derivative. The longer versions immediately make obvious the nesting of the case where $\mathbf{Y} = c\mathbf{I}$. The equality between the two versions can be derived based on the two different chain rule applications or

by noticing that

$$
\begin{aligned}
&(\mathbf{G}^{+}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\mathbf{G}\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\mathbf{X}^{-\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}(\mathbf{G}^{+}\mathbf{X}^{\otimes 2}(\mathbf{G}^{+})^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{X}^{\otimes 2}(\mathbf{G}^{+})^{\top}\mathbf{G}^{\top}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{X}^{\otimes 2}(\mathbf{G}\mathbf{G}^{+})^{\top}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{G}\mathbf{G}^{+}\mathbf{X}^{\otimes 2}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{X}^{\otimes 2}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
&(\mathbf{G}^{+}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\mathbf{G}^{+}\mathbf{X}^{\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\mathbf{G}\mathbf{G}^{+}\mathbf{X}^{\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\mathbf{X}^{\otimes 2}\mathbf{G}\\
&\quad=(\mathbf{G}^{\top}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}(\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}(\mathbf{G}^{+})^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}(\mathbf{G}^{+})^{\top}\mathbf{G}^{\top}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}(\mathbf{G}\mathbf{G}^{+})^{\top}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}\mathbf{G}\mathbf{G}^{+}\mathbf{X}^{-\otimes 2}(\mathbf{C}\otimes\mathbf{I})\mathbf{F}^{\top})^{-1}\\
&\quad=(\mathbf{G}^{+}(\mathbf{C}^{-\top}\otimes\mathbf{X}^{-1})\mathbf{F}^{\top})^{-1},
\end{aligned}
$$

where we have used Theorem 3.13 (c) of Magnus and Neudecker (2019) for the third equality.

□

Lemma 6.6. *Let* $\mathbf{X}$ *be a non-singular symmetric matrix,* $\mathbf{C}$ *its lower Cholesky factor, and* $\mathbf{Y}$ *and* $\mathbf{Z}$ *be square matrices. Then*

$$\frac{\partial \mathrm{tr}(\mathbf{XY})}{\partial \mathrm{vech}(\mathbf{X})^\top} = \mathrm{vec}(\mathbf{Y})^\top \mathbf{G},$$

$$\frac{\partial \mathrm{tr}(\mathbf{X}^{-1}\mathbf{Y})}{\partial \mathrm{vech}(\mathbf{X})^\top} = -\mathrm{vec}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1})^\top \mathbf{G},$$

$$\frac{\partial \mathrm{tr}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Z})}{\partial \mathrm{vech}(\mathbf{C})^\top} = -2\mathrm{vec}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top})^\top \mathbf{F}^\top,$$

and

$$\frac{\partial^2 \mathrm{tr}(\mathbf{XY})}{\partial \mathrm{vech}(\mathbf{X}) \partial \mathrm{vech}(\mathbf{X})^\top} = \mathbf{0}.$$

Proof. We have

$$\begin{aligned} \mathrm{dtr}(\mathbf{XY}) &= \mathrm{tr}(\mathbf{Y}\mathrm{d}\mathbf{X}) \\ &= \mathrm{vec}(\mathbf{Y})^\top \mathrm{vec}(\mathrm{d}\mathbf{X}) \\ &= \mathrm{vec}(\mathbf{Y})^\top \mathbf{G}\mathrm{dvech}(\mathbf{X}), \end{aligned}$$

using (14) and (16),

$$\begin{aligned} \mathrm{dtr}(\mathbf{X}^{-1}\mathbf{Y}) &= -\mathrm{tr}(\mathbf{X}^{-1}\mathrm{d}\mathbf{X}\mathbf{X}^{-1}\mathbf{Y}) \\ &= -\mathrm{tr}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1}\mathrm{d}\mathbf{X}) \\ &= -\mathrm{vec}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1})^\top \mathrm{vec}(\mathrm{d}\mathbf{X}) \\ &= -\mathrm{vec}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1})^\top \mathbf{G}\mathrm{dvech}(\mathbf{X}), \end{aligned}$$

using (13), (14) and (16),

$$\mathrm{d}^2\mathrm{tr}(\mathbf{XY}) = \mathrm{dtr}(\mathbf{Y}\mathrm{d}\mathbf{X}) = 0.$$

Furthermore,

$$\begin{aligned}
\mathrm{d}\mathrm{tr}(\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Z}) &= -\mathrm{tr}((\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1})^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Z} + \mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}\mathbf{C}^{-1}\mathbf{Z}) \\
&= -\mathrm{tr}(\mathrm{d}\mathbf{C}^{\top}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top} + \mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}) \\
&= -2\mathrm{tr}(\mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1}\mathrm{d}\mathbf{C}) \\
&= -2\mathrm{vec}(\mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})^{\top}\mathrm{d}\mathrm{vec}(\mathbf{C}) \\
&= -2\mathrm{vec}(\mathbf{C}^{-1}\mathbf{Z}\mathbf{C}^{-\top}\mathbf{Y}\mathbf{C}^{-1})^{\top}\mathbf{F}^{\top}\mathrm{d}\mathrm{vech}(\mathbf{C}),
\end{aligned}$$

For properties of the differential "d", see Magnus and Neudecker (2019) pp. 163-169 and pp. 434-436. To convert differentials to derivatives, see Tables 9.2 and 10.1 in Magnus and Neudecker (2019).

□

Lemma 6.7. *Let* $\mathbf{\Sigma}$ *be a symmetric positive definite matrix and* $\mathbf{C}$ *its lower Cholesky factor. Then*

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

$$\begin{aligned}
&\frac{\partial^2 \log|\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}(\mathbf{\Sigma})\partial \mathrm{vech}(\mathbf{\Sigma})^{\top}} \\
&\quad = -\mathbf{G}^{\top}(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1})\mathbf{F}^{\top}(\mathbf{G}^{\top}(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\mathbf{\Sigma}^{-\otimes 2}\mathbf{G} \\
&\quad = -\mathbf{G}^{\top}(\mathbf{C}^{-\top} \otimes \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1},
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial^2 \log|\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}(\mathbf{\Sigma}^{-1})\partial \mathrm{vech}(\mathbf{\Sigma}^{-1})^{\top}} \\
&\quad = \mathbf{G}^{\top}(\mathbf{C}\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1}\mathbf{G}^{+}\mathbf{\Sigma}^{\otimes 2}\mathbf{G} \\
&\quad = \mathbf{G}^{\top}(\mathbf{C}\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}^{-\top} \otimes \mathbf{\Sigma}^{-1})\mathbf{F}^{\top})^{-1}.
\end{aligned}$$

Proof. Decompose $\mathbf{\Sigma} = \mathbf{T}\mathbf{D}\mathbf{T}^{\top}$, 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

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

such that

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

and

$$\mathrm{dvec}(\mathbf{\Sigma}) = \mathrm{dvec}(\mathbf{T}\mathbf{D}\mathbf{T}^\top) = (\mathbf{T}^{\otimes 2}) \mathrm{dvec}(\mathbf{D}),$$

such that

$$\frac{\partial \mathrm{vec}(\mathbf{D})}{\partial \mathrm{vec}(\mathbf{\Sigma})^\top} = (\mathbf{T}^{\otimes 2})^{-1}.$$

Then

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

where we used (12). Furthermore

$$\begin{aligned} \frac{\partial \log |\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}(\mathbf{\Sigma}^{-1})^\top} &= \frac{\partial \log |\mathbf{\Sigma}|_{\mathbf{n}}}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \frac{\partial \mathrm{vech}(\mathbf{\Sigma})}{\partial \mathrm{vech}(\mathbf{\Sigma}^{-1})^\top} \\ &= -\mathrm{vec}(\mathbf{C}^{-\top} \mathrm{dg}(\mathbf{n}) \mathbf{C}^{-1})^\top \mathbf{G} \mathbf{G}^{+} \mathbf{\Sigma}^{\otimes 2} \mathbf{G} \\ &= -\mathrm{vec}(\mathbf{\Sigma} \mathbf{C}^{-\top} \mathrm{dg}(\mathbf{n}) \mathbf{C}^{-1} \mathbf{\Sigma})^\top \mathbf{G} \\ &= -\mathrm{vec}(\mathbf{C} \mathrm{dg}(\mathbf{n}) \mathbf{C}^\top)^\top \mathbf{G}, \end{aligned}$$

where we used Lemma 6.3 and equation (12). Now applying Lemma 6.5 on

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

and using $\mathbf{G}^\top\mathbf{G}\mathbf{G}^+ = \mathbf{G}^\top$ directly gives the desired results.

□

Lemma 6.8. *Let* $\mathbf{X}$ *be a square matrix. Then*

$$\mathbf{F}^\top\mathbf{F}\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(◣(\mathbf{X}))$$

Proof. $\mathbf{F}$ eliminates those elements from $\mathrm{vec}(\mathbf{X})$ which are on $\mathbf{X}$'s upper triangular part. Then $\mathbf{F}^\top$ we know from equation (19) is the matrix which for a triangular matrix maps its half-vectorization $\mathrm{vech}(\cdot)$ to its vectorization $\mathrm{vec}(\cdot)$.

□

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 6.9. *Let* $\mathcal{D} \in (\mathcal{W}, i\mathcal{W}, t\mathcal{W}, it\mathcal{W}, F, \mathcal{R}, i\mathcal{R}, t\mathcal{R}, it\mathcal{R}, F\mathcal{R}, iF\mathcal{R})$ *be one of the distributions for RCs. Then*

$$\nabla^{\Sigma}_{\mathcal{D}} = \mathbf{G}^{\top}\mathrm{vec}(\triangle^{\Sigma}_{\mathcal{D}}),$$

with

$$\triangle^{\Sigma}_{\mathcal{D}} = \mathbf{C}^{-\top}\tilde{\blacksquare}(\mathbf{C}^{\top}\blacksquare(\triangle^{\Omega}_{\mathcal{D}}\mathbf{C}\mathbf{M}^{-1}_{\mathcal{D}}))\mathbf{C}^{-1},$$

where $\triangle^{\Omega}_{\mathcal{D}}$ *is symmetric with elements*

$$(\triangle^{\Omega}_{\mathcal{D}})_{ij} = \frac{\partial p_{\mathcal{D}}}{\partial(\mathbf{\Omega})_{ij}}, \tag{33}$$

$\mathbf{M}_{\mathcal{D}}$ *is the expectation of the stochastic representation kernel for the respective distribution (see Stollenwerk (2023)),* $\mathbf{C}$ *is the lower Cholesky factor of the expected value matrix* $\mathbf{\Sigma}$ *and* $\blacksquare(\cdot)$ *and* $\tilde{\blacksquare}(\cdot)$ *are defined in 6.1 and 6.3. For all Wishart-type distributions* $\mathbf{M}_{\mathcal{D}} = m_{\mathcal{D}}\mathbf{I}$*, with scalar* $m_{\mathcal{D}}$*, and the expression for* $\triangle^{\Sigma}_{\mathcal{D}}$ *reduces to*

$$\triangle^{\Sigma}_{\mathcal{D}} = \frac{1}{m_{\mathcal{D}}}\triangle^{\Omega}_{\mathcal{D}}.$$

Proof.

$$
\begin{aligned}
\frac{\partial p_{\mathcal{D}}}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} &= \frac{\partial p_{\mathcal{D}}}{\partial \mathrm{vech}(\mathbf{\Omega})^\top} \frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} \\
&= \mathrm{vec}(\triangle_{\mathcal{D}}^{\Omega})^\top \mathbf{G} \frac{\partial \mathrm{vech}(\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}^\top)}{\partial \mathrm{vech}(\mathbf{C}\mathbf{C}^\top)^\top} \\
&= \mathrm{vec}(\triangle_{\mathcal{D}}^{\Omega})^\top \mathbf{G}\mathbf{G}^{+}(\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1} \otimes \mathbf{I})\mathbf{F}^\top(\mathbf{G}^{+}(\mathbf{C} \otimes \mathbf{I})\mathbf{F}^\top)^{-1} \\
&= \mathrm{vec}(\triangle_{\mathcal{D}}^{\Omega})^\top (\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1} \otimes \mathbf{I})\mathbf{F}^\top(\mathbf{G}^{+}(\mathbf{C} \otimes \mathbf{I})\mathbf{F}^\top)^{-1} \\
&= \mathrm{vec}(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1})^\top \mathbf{F}^\top(\mathbf{G}^{+}(\mathbf{C} \otimes \mathbf{I})\mathbf{F}^\top)^{-1} \\
&= 2\mathrm{vec}(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1})^\top \mathbf{F}^\top\mathbf{F}(\mathbf{I} \otimes \mathbf{C})\mathbf{Z}\mathbf{C}^{-\otimes 2}\mathbf{G} \\
&= 2\mathrm{vec}(\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C})\mathbf{M}_{\mathcal{D}}^{-1})^\top (\mathbf{I} \otimes \mathbf{C})\mathbf{Z}\mathbf{C}^{-\otimes 2}\mathbf{G} \\
&= 2\mathrm{vec}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C})\mathbf{M}_{\mathcal{D}}^{-1})^\top \mathbf{Z}\mathbf{C}^{-\otimes 2}\mathbf{G} \\
&= 2\mathrm{vec}(\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C})\mathbf{M}_{\mathcal{D}}^{-1}))^\top \mathbf{C}^{-\otimes 2}\mathbf{G} \\
&= 2\mathrm{vec}(\mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C})\mathbf{M}_{\mathcal{D}}^{-1})\mathbf{C}^{-1})^\top \mathbf{G} \\
&= 2\mathrm{vec}(\mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1})^\top \mathbf{G},
\end{aligned}
$$

where we used Lemmas 6.5, 6.8, and 6.4, Definition 6.2 and equation (25). For $\bar{\blacktriangleright}(\cdot)$ see Definition 6.2. Note that, even though

$$
\mathrm{vec}(\triangle_{\mathcal{D}}^{\Sigma})^\top \mathbf{G} = \frac{\partial \log p_{\mathcal{D}}}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} = 2\mathrm{vec}(\mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1})^\top \mathbf{G},
$$

this does **not** imply that

$$
\triangle_{\mathcal{D}}^{\Sigma} = 2\mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1}
$$

since $\triangle_{\mathcal{D}}^{\Sigma}$ is symmetric, but the r.h.s. is not necessarily symmetric. However, since we know that $\triangle_{\mathcal{D}}^{\Sigma}$ is the **unique symmetric** matrix for which $\frac{\partial p_{\mathcal{D}}}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top} = \mathrm{vec}(\triangle_{\mathcal{D}}^{\Sigma})^\top \mathbf{G}$ holds, it must be true that

$$
\begin{aligned}
\triangle_{\mathcal{D}}^{\Sigma} &= \mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1} + (\mathbf{C}^{-\top}\bar{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1})^\top \\
&= \mathbf{C}^{-\top}\tilde{\blacktriangleright}(\mathbf{C}^\top\blacktriangleright(\triangle_{\mathcal{D}}^{\Omega}\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}))\mathbf{C}^{-1}.
\end{aligned}
$$

Furthermore, note that

$$\begin{aligned}\mathbf{C}^{-\top}\bar{\blacksquare}(\mathbf{C}^{\top}\blacksquare(\mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})))\mathbf{C}^{-1} &= \mathbf{C}^{-\top}\bar{\blacksquare}(\mathbf{C}^{\top}\mathrm{dg}(\mathbf{C}^{-\top})\mathrm{dg}(\mathbf{n}))\mathbf{C}^{-1}\\ &= \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{C}^{\top})\mathrm{dg}(\mathbf{C}^{-\top})\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}\\ &= \mathbf{C}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}^{-1}.\end{aligned}$$

□

Lemma 6.10. *Let* $\mathbf{n}$ *be a real column vector of length* p*, then*

$$\frac{\partial \log|\boldsymbol{\Sigma}|_{\mathbf{n}}}{\partial \mathbf{n}^{\top}} = 2\log \mathrm{vecd}(\mathbf{C})^{\top}.$$

Proof.

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

□

The following Lemma is just for rewriting of the scores in Table 3.

Lemma 6.11. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be symmetric* $p \times p$ *positive definite matrices with lower Cholesky factors* $\mathbf{C}_X$ *and* $\mathbf{C}_Y$*, respectively, and let* $\mathbf{n}$ *be a real vector of length* p*. Then*

$$\log \mathrm{vecd}(\mathbf{C}_X) - \log \mathrm{vecd}(\mathbf{C}_Y) = \log \mathrm{vecd}(\mathbf{C}_Y^{-1}\mathbf{C}_X)$$

Proof.

$$\begin{aligned}\log \mathrm{vecd}(\mathbf{C}_X) - \log \mathrm{vecd}(\mathbf{C}_Y) &= \frac{1}{2}\frac{\partial(\log|\mathbf{X}|_{\mathbf{n}} - \log|\mathbf{Y}|_{\mathbf{n}})}{\partial \mathbf{n}^{\top}}\\ &= \frac{1}{2}\frac{\partial \log(|\mathbf{X}|_{\mathbf{n}}|\mathbf{Y}|_{-\mathbf{n}})}{\partial \mathbf{n}^{\top}}\\ &= \frac{1}{2}\frac{\partial \log(|\mathbf{C}_Y^{-1}\mathbf{C}_X\mathbf{C}_X^{\top}\mathbf{C}_X^{-\top}|_{\mathbf{n}})}{\partial \mathbf{n}^{\top}} = \log \mathrm{vecd}(\mathbf{C}_Y^{-1}\mathbf{C}_X).\end{aligned}$$

□

Lemma 6.12. *Consider the multivariate gamma function. We have*

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

with $\boldsymbol{\psi}_i(\mathbf{n}) = \psi(n_i - \frac{1}{2}(i-1))$ *and* $\overleftarrow{\boldsymbol{\psi}}_i(\overleftarrow{\mathbf{n}}) = \psi(n_{p-i+1} - \frac{1}{2}(p-i))$*, where* $\psi(\cdot)$ *is defined in equation (5.2.2) of NIST Digital Library of Mathematical Functions. Furthermore*

$$\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n} \partial \mathbf{n}^\top} = \mathrm{dg}(\boldsymbol{\psi}'(\mathbf{n})), \text{ and}$$
$$\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n} \partial \mathbf{n}^\top} = \mathrm{dg}(\overleftarrow{\boldsymbol{\psi}}'(\overleftarrow{\mathbf{n}}))$$

with $\boldsymbol{\psi}'(\mathbf{n}) = \left[\psi'\left(n_1\right), \psi'\left(n_2 - \frac{1}{2}\right), \ldots, \psi'\left(n_p - \frac{1}{2}(p-1)\right)\right]^\top$ *and* $\overleftarrow{\boldsymbol{\psi}}'(\overleftarrow{\mathbf{n}})$ *similar to* $\overleftarrow{\boldsymbol{\psi}}_i(\overleftarrow{\mathbf{n}})$ *above. Finally,*

$$\frac{\partial \log \Gamma_p(n)}{\partial n} = \sum_{j=1}^{p} \psi\left(n - \frac{1}{2}(j-1)\right),$$

and

$$\frac{\partial^2 \log \Gamma_p(n)}{\partial n^2} = \sum_{j=1}^{p} \psi'\left(n - \frac{1}{2}(j-1)\right).$$

Proof.

$$\begin{aligned}
\frac{\partial \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n}^\top} &= \frac{\partial \log\left(\pi^{p(p-1)/4} \prod_{j=1}^{p} \Gamma\left(n_j - \frac{1}{2}(j-1)\right)\right)}{\partial \mathbf{n}^\top} \\
&= \frac{\sum_{j=1}^{p} \partial \log\left(\Gamma\left(n_j - \frac{1}{2}(j-1)\right)\right)}{\partial \mathbf{n}^\top} \\
&= \left[\psi\left(n_1\right), \psi\left(n_2 - \frac{1}{2}\right), \ldots, \psi\left(n_p - \frac{1}{2}(p-1)\right)\right] \\
&= \boldsymbol{\psi}(\mathbf{n})^\top,
\end{aligned}$$

with $\boldsymbol{\psi}_i(\mathbf{n}) = \psi\left(n_i - \frac{1}{2}(i-1)\right)$ as defined in equation (5.2.2) of NIST Digital

Library of Mathematical Functions, and

$$\begin{aligned}\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n} \partial \mathbf{n}^\top} &= \mathrm{dg}\left(\left[\psi'\left(n_1\right), \psi'\left(n_2 - \frac{1}{2}\right), \ldots, \psi'\left(n_p - \frac{1}{2}(p-1)\right)\right]\right) \\ &= \mathrm{dg}\left(\boldsymbol{\psi}'(\mathbf{n})\right)\end{aligned}$$

since for $i \neq j$, $\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial n_i \partial n_j} = 0$, with $\boldsymbol{\psi}_i'(\mathbf{n}) = \psi'\left(n_i - \frac{1}{2}(i-1)\right)$. Obviously, we have

$$\frac{\partial \log \Gamma_p(\overleftarrow{\mathbf{n}})}{\partial \overleftarrow{\mathbf{n}}^\top} = \boldsymbol{\psi}(\overleftarrow{\mathbf{n}})^\top,$$

such that

$$\frac{\partial \log \Gamma_p(\overleftarrow{\mathbf{n}})}{\partial \mathbf{n}^\top} = \frac{\overleftarrow{\partial \log \Gamma_p(\overleftarrow{\mathbf{n}})}}{\partial \overleftarrow{\mathbf{n}}^\top} =: \overleftarrow{\boldsymbol{\psi}}(\overleftarrow{\mathbf{n}})^\top.$$

Similarly for

$$\frac{\partial^2 \log \Gamma_p(\mathbf{n})}{\partial \mathbf{n} \partial \mathbf{n}^\top} = \mathrm{dg}(\overleftarrow{\boldsymbol{\psi}}'(\overleftarrow{\mathbf{n}})).$$

Finally,

$$\begin{aligned}\frac{\partial \log \Gamma_p(n)}{\partial n} &= \frac{\sum_{j=1}^p \partial \log\left(\Gamma\left(n - \frac{1}{2}(j-1)\right)\right)}{\partial n} \\ &= \sum_{j=1}^p \psi\left(n - \frac{1}{2}(j-1)\right),\end{aligned}$$

and

$$\frac{\partial^2 \log \Gamma_p(n)}{\partial n^2} = \sum_{j=1}^p \psi'\left(n - \frac{1}{2}(j-1)\right).$$

□

6.2 Scores and Fisher Information Matrices

6.2.1 Scores w.r.t. $\boldsymbol{\Omega}$

We want to derive the scores listed in Table 1. The scores w.r.t. $\boldsymbol{\Omega}$ of the Riesz, Inverse Riesz, t-Riesz, and Inverse t-Riesz distributions are easy to derive by taking

the logarithm of the p.d.f.s given in Stollenwerk, 2023 and straightforwardly applying Lemmas 6.7 and 6.6. For the F-Riesz and Inverse F-Riesz, we have to start from the p.d.f. representations in Stollenwerk, 2023 because they contain elements of the form $|\mathbf{I}+\mathbf{X}|_{\mathbf{n}}$, which are not necessarily positive definite and for which we can thus not apply Lemma 6.7. We show, as an example, the most complicated derivation, which is the one of the Inverse F-Riesz. We denote $\mathbf{J} = (\boldsymbol{\Omega}^{-1}+\mathbf{R}^{-1})^{-1}$ with lower Cholesky factor $\mathbf{C}_J$ to derive

$$
\begin{aligned}
\frac{\partial \log p_{iF\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} &= \frac{\partial \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} + \frac{\partial \log |\mathbf{J}|_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}}{\partial \mathrm{vech}(\mathbf{J})^\top}\frac{\partial \mathrm{vech}(\mathbf{J})}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} \\
&= -\frac{1}{2}\mathrm{vec}(\mathbf{C}_\Omega^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^{-1})^\top\mathbf{G} \\
&\quad + \frac{1}{2}\mathrm{vec}(\mathbf{C}_J^{-\top}\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_J^{-1})^\top\mathbf{G}\frac{\partial \mathrm{vech}(\mathbf{J})}{\partial \mathrm{vech}(\mathbf{J}^{-1})^\top}\frac{\partial \mathrm{vech}(\mathbf{J}^{-1})}{\partial \mathrm{vech}(\boldsymbol{\Omega})^\top} \\
&= -\frac{1}{2}\mathrm{vec}(\mathbf{C}_\Omega^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^{-1})^\top\mathbf{G} \\
&\quad + \frac{1}{2}\mathrm{vec}(\mathbf{C}_J^{-\top}\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_J^{-1})^\top\mathbf{G}\mathbf{G}^+(\mathbf{J}^{-1})^{-\otimes 2}\mathbf{G}\mathbf{G}^+\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G} \\
&= -\frac{1}{2}\mathrm{vec}(\mathbf{C}_\Omega^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^{-1})^\top\mathbf{G} \\
&\quad + \frac{1}{2}\mathrm{vec}(\mathbf{C}_J^{-\top}\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_J^{-1})^\top\mathbf{G}\mathbf{G}^+(\mathbf{J}\boldsymbol{\Omega}^{-1})^{\otimes 2}\mathbf{G} \\
&= -\frac{1}{2}\mathrm{vec}(\mathbf{C}_\Omega^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^{-1} + \boldsymbol{\Omega}^{-1}\mathbf{J}\mathbf{C}_J^{-\top}\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_J^{-1}\mathbf{J}\boldsymbol{\Omega}^{-1})^\top\mathbf{G} \\
&= -\frac{1}{2}\mathrm{vec}(\mathbf{C}_\Omega^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^{-1} + \boldsymbol{\Omega}^{-1}\mathbf{C}_J\mathrm{dg}(\mathbf{n}+\boldsymbol{\nu})\mathbf{C}_J^{\top}\boldsymbol{\Omega}^{-1})^\top\mathbf{G},
\end{aligned}
$$

where apart from the aforementioned Lemmas we used Lemma 6.3 and equation (26). The scores w.r.t. $\boldsymbol{\Omega}$ of the other Riesz-type distributions are, as mentioned above, much simpler, and the ones of the Wishart-type distributions then follow by setting all elements in the degree of freedom parameter vectors of the corresponding Riesz-type score equal to each other. Finally, to construct Table 1, recall equation (33).

6.2.2 Scores w.r.t. $\boldsymbol{\Sigma}$

The scores w.r.t. $\boldsymbol{\Sigma}$ are then straightforward to derive using Lemma 6.9 and using the distribution-specific maps from $\boldsymbol{\Omega}$ to $\boldsymbol{\Sigma}$ (see Stollenwerk, 2023)

6.2.3 Fisher Information Matrices w.r.t. Ω

For the Riesz (and Wishart) and t-Riesz (and t-Wishart) distribution we will use

$$\begin{aligned}\mathcal{I}_{\mathcal{D}}^{\Omega} &= \mathbb{E}\left[\left(\frac{\partial \log p_{\mathcal{D}}}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} \frac{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}}\right)^{\top} \frac{\partial \log p_{\mathcal{D}}}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} \frac{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}}\right] \\ &= \left(\frac{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}}\right)^{\top} \mathcal{I}_{\mathcal{D}}^{(\Omega^{-1})} \frac{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}\right)^{\top}}. \end{aligned} \tag{34}$$

Riesz

We have

$$\begin{aligned}\mathcal{I}_{\mathcal{R}}^{(\Omega^{-1})} &= -\mathbb{E}\left[\frac{\partial^2 \log p_{\mathcal{R}}}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right) \partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\right] \\ &= -\mathbb{E}\left[\frac{\partial^2 \log |\boldsymbol{\Omega}|_{-\frac{\mathbf{n}}{2}}}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right) \partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}} - \frac{\partial^2 \frac{\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}\mathbf{R}\right)}{2}}{\partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right) \partial \operatorname{vech}\left(\boldsymbol{\Omega}^{-1}\right)^{\top}}\right] \\ &= \frac{1}{2}\mathbf{G}^{\top}\left(\mathbf{C}_{\Omega}\operatorname{dg}(\mathbf{n}) \otimes \mathbf{I}\right)\mathbf{F}^{\top}\left(\mathbf{G}^{\top}\left(\mathbf{C}_{\Omega} \otimes \mathbf{I}\right)\mathbf{F}^{\top}\right)^{-1}\mathbf{G}^{\top}\boldsymbol{\Omega}^{\otimes 2}\mathbf{G}. \end{aligned}$$

such that using equation (34) and Lemma (6.3), we have

$$\begin{aligned}\mathcal{I}_{\mathcal{R}}^{\Omega} &= \mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}(\mathbf{G}^{+})^{\top}\frac{1}{2}\mathbf{G}^{\top}(\mathbf{C}_{\Omega}\operatorname{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{\top}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1}\mathbf{G}^{\top}\boldsymbol{\Omega}^{\otimes 2}\mathbf{G} \\ &\quad \times \mathbf{G}^{+}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G} \\ &= \frac{1}{2}\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G}\mathbf{G}^{+}(\mathbf{C}_{\Omega}\operatorname{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1} \\ &= \frac{1}{2}\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}(\mathbf{C}_{\Omega}\operatorname{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1} \\ &= \frac{1}{2}\mathbf{G}^{\top}(\mathbf{C}_{\Omega}^{-\top}\operatorname{dg}(\mathbf{n}) \otimes \boldsymbol{\Omega}^{-1})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1}. \end{aligned} \tag{35}$$

Wishart

Starting from equation (35) and setting $n_1, \ldots, n_p = n$ we have

$$\begin{aligned}\mathcal{I}_{\mathcal{W}}^{\Omega} &= \frac{n}{2}\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G}\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1} \\ &= \frac{n}{2}\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G}. \end{aligned}$$

Inverse Riesz

$$\begin{aligned}\mathcal{I}^{\Omega}_{i\mathcal{R}} &= -\mathbb{E}\left[\frac{\partial^2 \log|\mathbf{\Omega}|_{\frac{\nu}{2}}}{\partial \text{vech}\left(\mathbf{\Omega}\right)\partial \text{vech}\left(\mathbf{\Omega}\right)^\top} - \frac{\partial^2 \frac{\text{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)}{2}}{\partial \text{vech}\left(\mathbf{\Omega}\right)\partial \text{vech}\left(\mathbf{\Omega}\right)^\top}\right] \\ &= \frac{1}{2}\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top}\otimes\mathbf{C}_\Omega^{-\top}\text{dg}(\boldsymbol{\nu})\mathbf{C}_\Omega^{-1})\mathbf{F}^\top(\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top}\otimes\mathbf{\Omega}^{-1})\mathbf{F}^\top)^{-1}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G} \qquad (36)\\ &= \frac{1}{2}\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top}\otimes\mathbf{C}_\Omega^{-\top}\text{dg}(\boldsymbol{\nu})\mathbf{C}_\Omega^{-1})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}_\Omega\otimes\mathbf{I})\mathbf{F}^\top)^{-1}\end{aligned}$$

Inverse Wishart

Starting form equation (36) and setting $\nu_1,\ldots,\nu_p = \nu$ we have

$$\begin{aligned}\mathcal{I}^{\Omega}_{i\mathcal{W}} &= \frac{\nu}{2}\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top}\otimes\mathbf{C}_\Omega^{-\top}\mathbf{C}_\Omega^{-1})\mathbf{F}^\top(\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top}\otimes\mathbf{\Omega}^{-1})\mathbf{F}^\top)^{-1}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}\\ &= \frac{\nu}{2}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}\end{aligned}$$

t-Riesz

We will need the expectation of

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

For a better overview, denote the normalizing constant of the t-Riesz distribution by

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

Then the desired expectation obtains from

$$
\begin{aligned}
&\mathbb{E}[(1+\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))^{-2}\mathrm{vec}^2(\mathbf{R})] \\
&\quad= c(\nu,\mathbf{n},\mathbf{\Omega})\int_{\mathbf{R}>\mathbf{0}}\mathrm{vec}^2(\mathbf{R})|\mathbf{R}|^{\frac{\mathbf{n}-p-1}{2}}(1+\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R}))^{-\frac{\nu+p\bar{\mathbf{n}}+4}{2}}\mathrm{d}\mathbf{R} \\
&\quad= \frac{c(\nu,\mathbf{n},\mathbf{\Omega})}{c(\nu+4,\mathbf{n},\mathbf{\Omega})}\mathbb{E}[\mathrm{vec}^2(\mathbf{R})] \text{ for } \mathbf{R}\sim t\mathcal{R}(\mathbf{n},\nu+4,\mathbf{\Omega}) \\
&\quad= \frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}\frac{\Gamma_p(\mathbf{n}/2)\Gamma((\nu+4)/2)}{\Gamma(((\nu+4)+p\bar{\mathbf{n}})/2)}\mathbb{E}[\mathrm{vec}^2(\mathbf{R})] \text{ for } \mathbf{R}\sim t\mathcal{R}(\mathbf{n},\nu+4,\mathbf{\Omega}) \\
&\quad= \frac{(\nu+2)\nu}{(\nu+p\bar{\mathbf{n}})\,(\nu+p\bar{\mathbf{n}}+2)}\mathbb{E}[\mathrm{vec}^2(\mathbf{A})]\mathbb{E}[(\bar{b})^{-4}] \text{ for } \mathbf{A}\sim\mathcal{R}(\mathbf{\Omega},\mathbf{n}) \text{ and } \bar{b}\sim\chi_{\nu+4} \\
&\quad= \frac{1}{(\nu+p\bar{\mathbf{n}})\,(\nu+p\bar{\mathbf{n}}+2)}(\mathrm{vec}^2(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top)+2\mathbf{G}\mathbf{G}^+(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top\otimes\mathbf{\Omega})\mathbf{G}\mathbf{G}^+),
\end{aligned}
$$

where we used the stochastic representation of the t-Riesz distribution, equation (40) and $(\bar{b})^{-2}$ following an Inverse Chi-squared distribution with[20]

$$
\mathbb{E}[(\bar{b})^{-2}] = (\nu+2)^{-1} \text{ and } \mathbf{Var}((\bar{b})^{-2}) = \frac{2}{(\nu+2)^2\nu},
$$

such that

$$
\mathbb{E}[(\bar{b})^{-4}] = \mathbb{E}[(\bar{b})^{-2}]^2 + \mathbf{Var}((\bar{b})^{-2}) = ((\nu+2)\nu)^{-1}. \tag{37}
$$

Now, using the derived expectation, we have

$$
\begin{aligned}
\mathcal{I}_{t\mathcal{R}}^{(\Omega^{-1})} &= -\mathbb{E}\left[\frac{\partial^2\log|\mathbf{\Omega}|^{-\frac{\mathbf{n}}{2}}}{\partial\mathrm{vech}(\mathbf{\Omega}^{-1})\partial\mathrm{vech}(\mathbf{\Omega}^{-1})^\top}+\frac{\nu+p\bar{\mathbf{n}}}{2}\frac{\partial^2\log\left(1+\mathrm{tr}(\mathbf{\Omega}^{-1}\mathbf{R})\right)}{\partial\mathrm{vech}(\mathbf{\Omega}^{-1})\partial\mathrm{vech}(\mathbf{\Omega}^{-1})^\top}\right] \\
&= \frac{1}{2}\mathbf{G}^\top(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}_\Omega\otimes\mathbf{I})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{\Omega}^{\otimes 2}\mathbf{G} \\
&\quad- \frac{\nu+p\bar{\mathbf{n}}}{2(\nu+p\bar{\mathbf{n}})\,(\nu+p\bar{\mathbf{n}}+2)} \\
&\qquad\times\mathbf{G}^\top(\mathrm{vec}^2(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top)+2\mathbf{G}\mathbf{G}^+(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top\otimes\mathbf{\Omega})\mathbf{G}\mathbf{G}^+)\mathbf{G} \\
&= \frac{1}{2}\mathbf{G}^\top(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\otimes\mathbf{I})\mathbf{F}^\top(\mathbf{G}^+(\mathbf{C}_\Omega\otimes\mathbf{I})\mathbf{F}^\top)^{-1}\mathbf{G}^+\mathbf{\Omega}^{\otimes 2}\mathbf{G} \\
&\quad- \frac{1}{2(\nu+p\bar{\mathbf{n}}+2)}\mathbf{G}^\top(\mathrm{vec}^2(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top)+2(\mathbf{C}_\Omega\mathrm{dg}(\mathbf{n})\mathbf{C}_\Omega^\top\otimes\mathbf{\Omega}))\mathbf{G}.
\end{aligned} \tag{38}
$$

Finally, use equation (34) to arrive at $\mathcal{I}_{t\mathcal{R}}^{\Omega}$.

20. Compare, e.g. Wikipedia.

t-Wishart

Starting form equation (38) and setting $n_1, \dots, n_p = n$, we have

$$\begin{aligned}
\mathcal{I}_{t\mathcal{W}}^{(\Omega^{-1})} &= \frac{n}{2}\mathbf{G}^\top \boldsymbol{\Omega}^{\otimes 2}\mathbf{G} - \frac{n}{2(\nu+pn+2)}\mathbf{G}^\top(n\mathrm{vec}^2(\boldsymbol{\Omega}) + 2\boldsymbol{\Omega}^{\otimes 2})\mathbf{G} \\
&= \frac{n}{2}\mathbf{G}^\top\left(\boldsymbol{\Omega}^{\otimes 2} - \frac{2}{\nu+pn+2}\boldsymbol{\Omega}^{\otimes 2} - \frac{n}{\nu+pn+2}\mathrm{vec}^2(\boldsymbol{\Omega})\right)\mathbf{G} \\
&= \frac{n}{2}\mathbf{G}^\top\left(\frac{\nu+pn}{\nu+pn+2}\boldsymbol{\Omega}^{\otimes 2} - \frac{n}{\nu+pn+2}\mathrm{vec}^2(\boldsymbol{\Omega})\right)\mathbf{G}.
\end{aligned}$$

Now, using equations (34), (26) and (23) and Lemma 6.3 we have

$$\begin{aligned}
\mathcal{I}_{t\mathcal{W}}^{\Omega} &= \frac{n}{2}(\mathbf{G}\mathbf{G}^+\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G})^\top\left(\frac{\nu+pn}{\nu+pn+2}\boldsymbol{\Omega}^{\otimes 2} - \frac{n}{\nu+pn+2}\mathrm{vec}^2(\boldsymbol{\Omega})\right)\mathbf{G}\mathbf{G}^+\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G} \\
&= \frac{n}{2}\mathbf{G}^\top\left(\frac{\nu+pn}{\nu+pn+2}\boldsymbol{\Omega}^{-\otimes 2} - \frac{n}{\nu+pn+2}\mathrm{vec}^2(\boldsymbol{\Omega}^{-1})\right)\mathbf{G}.
\end{aligned}$$

Inverse t-Riesz

Not derived since we don't need it in the paper. The derivation should be very similar to the one of the Inverse t-Wishart below, but at some point, we would need $\mathbb{E}[\mathrm{vec}^2(\mathbf{C}_\Omega^{-\top}\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\mathbf{C}_\Omega^{-1})]$, which is probably very similar to derive as Lemma 6.13.

Inverse t-Wishart

This derivation is similar to the one of the t-Riesz distribution. We will need the expectation of

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

For a better overview, denote the normalizing constant of the Inverse t-Riesz distribution by

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

Then the desired expectation obtains from

$$\begin{aligned}
&\mathbb{E}[(1+\mathrm{tr}(\mathbf{\Omega R}^{-1}))^{-2}\mathrm{vec}^2(\mathbf{R}^{-1})] \\
&= c(n,\nu,\mathbf{\Omega})\int_{\mathbf{R}>\mathbf{0}} \mathrm{vec}^2(\mathbf{R}^{-1})|\mathbf{R}|^{-\frac{\nu+p+1}{2}}(1+\mathrm{tr}(\mathbf{\Omega R}^{-1}))^{-\frac{n+p\nu+4}{2}}\mathrm{d}\mathbf{R} \\
&= \frac{c(n,\nu,\mathbf{\Omega})}{c(n+4,\nu,\mathbf{\Omega})}\mathbb{E}[\mathrm{vec}^2(\mathbf{R}^{-1})],\ \text{for } \mathbf{R}\sim it\mathcal{W}(\mathbf{\Omega},n+4,\nu) \\
&= \frac{\Gamma((n+p\nu)/2)}{\Gamma((n+4+p\nu)/2)}\frac{\Gamma((n+4)/2)}{\Gamma(n/2)}\mathbb{E}[(\underline{b})^{-4}]\mathbb{E}[\mathrm{vec}^2(\mathbf{C}_\Omega^{-\top}\bar{\mathcal{B}}\bar{\mathcal{B}}^\top\mathbf{C}_\Omega^{-1})] \\
&= \frac{\nu}{(n+p\nu)(n+p\nu+2)}(2\mathbf{G}\mathbf{G}^+\mathbf{\Omega}^{-\otimes 2}(\mathbf{G}\mathbf{G}^+)^\top+\nu\mathrm{vec}^2(\mathbf{\Omega}^{-1}))
\end{aligned}$$

where we used equation (37) and the fact that $\mathbf{C}_\Omega^{-\top}\bar{\mathcal{B}}\bar{\mathcal{B}}^\top\mathbf{C}_\Omega^{-1}\sim\mathcal{W}(\mathbf{\Omega}^{-1},\nu)$[21] and thus we used equation (40) with $\mathbf{n}=(\nu,\ldots,\nu)$. Now, using the derived expectation we have, using Lemma 6.7 with $\mathbf{n}=(n,\ldots,n)$,

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

<u>F-Riesz and Inverse F-Riesz</u>

It is not derived since we do not need it in the paper.[22]

<u>Matrix-F</u>

$$\begin{aligned}
&\frac{\partial^2\log p_F}{\partial\mathrm{vech}(\mathbf{\Omega})\partial\mathrm{vech}(\mathbf{\Omega})^\top} = \frac{\nu}{2}\frac{\partial^2\log|\mathbf{\Omega}|}{\partial\mathrm{vech}(\mathbf{\Omega})\partial\mathrm{vech}(\mathbf{\Omega})^\top}-\frac{n+\nu}{2}\frac{\partial^2\log|\mathbf{\Omega}+\mathbf{R}|}{\partial\mathrm{vech}(\mathbf{\Omega})\partial\mathrm{vech}(\mathbf{\Omega})^\top} \\
&= -\frac{\nu}{2}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}+\frac{n+\nu}{2}\mathbf{G}^\top(\mathbf{\Omega}+\mathbf{R})^{-\otimes 2}\mathbf{G} \\
&= -\frac{\nu}{2}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}+\frac{n+\nu}{2}\mathbf{G}^\top\left(\mathbf{C}_\Omega^{-\otimes 2}\right)^\top(\mathbf{I}+\mathbf{C}_\Omega^{-1}\mathbf{R}\mathbf{C}_\Omega^{-\top})^{-\otimes 2}\mathbf{C}_\Omega^{-\otimes 2}\mathbf{G}.
\end{aligned}$$

21. See Gupta and Nagar (2000).
22. Also, it seems very difficult.

Now, $\mathbf{Z}_{Kollo} = \mathbf{C}_\Omega^{-1}\mathbf{R}\mathbf{C}_\Omega^{-\top}$ follows a matrix F distribution with scale matrix $\mathbf{I}$ as defined in Theorem 2.4.9 of Kollo and von Rosen (2005). They derive on p. 265,

$$\begin{aligned}\mathbb{E}\left[\left(\mathbf{I}+\mathbf{C}_\Omega^{-1}\mathbf{R}\mathbf{C}_\Omega^{-\top}\right)^{-\otimes 2}\right] &= \left(c_3\mathbf{I}+c_4\mathbf{K}_{pp}+c_4\mathrm{vec}^2\left(\mathbf{I}\right)\right)\\ &= \left((c_3-c_4)\mathbf{I}+c_4\left(\mathbf{I}+\mathbf{K}_{pp}\right)+c_4\mathrm{vec}^2\left(\mathbf{I}\right)\right)\\ &= \left((c_3-c_4)\mathbf{I}+2c_4\mathbf{G}\mathbf{G}^{+}+c_4\mathrm{vec}^2\left(\mathbf{I}\right)\right),\end{aligned}$$

with, according to their p. 263,

$$\begin{aligned}c_4 &= \frac{n-p-1}{(n+\nu-1)(n+\nu+2)}\left(\left(n-p-2+\frac{1}{n+\nu}\right)c_2-\left(1+\frac{n-p-1}{n+\nu}\right)c_1\right),\\ c_3 &= \frac{n-p-1}{n+\nu}\left(\left(n-p-2\right)c_2-c_1\right)-\left(n+\nu+1\right)c_4,\\ c_2 &= \frac{n(\nu-p-2)+n^2+n}{(\nu-p)(\nu-p-1)(\nu-p-3)},\\ c_1 &= \frac{n^2(\nu-p-2)+2n}{(\nu-p)(\nu-p-1)(\nu-p-3)}.\end{aligned}$$

Thus

$$\begin{aligned}\mathcal{I}_F^{\Omega} &= -\mathbb{E}\left[\frac{\partial^2 \log p_F}{\partial \mathrm{vech}(\mathbf{\Omega})\partial \mathrm{vech}(\mathbf{\Omega})^\top}\right]\\ &= \frac{\nu}{2}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}-\frac{n+\nu}{2}\mathbf{G}^\top(\mathbf{C}_\Omega^{-\top})^{\otimes 2}((c_3-c_4)\mathbf{I}+2c_4\mathbf{G}\mathbf{G}^{+}+c_4\mathrm{vec}^2(\mathbf{I}))\mathbf{C}_\Omega^{-\otimes 2}\mathbf{G}\\ &= \frac{\nu}{2}\mathbf{G}^\top\mathbf{\Omega}^{-\otimes 2}\mathbf{G}-\frac{n+\nu}{2}\mathbf{G}^\top((c_3-c_4)\mathbf{\Omega}^{-\otimes 2}+2c_4\mathbf{\Omega}^{-\otimes 2}+c_4\mathrm{vec}^2(\mathbf{\Omega}^{-1}))\mathbf{G}\\ &= \frac{1}{2}\mathbf{G}^\top((\nu-(n+\nu)(c_3+c_4))\mathbf{\Omega}^{-\otimes 2}-(n+\nu)c_4\mathrm{vec}^2(\mathbf{\Omega}^{-1}))\mathbf{G}.\end{aligned}$$

6.2.4 Fisher Information Matrices w.r.t. $\mathbf{\Sigma}$

Similar to equation (34), we can derive

$$\mathcal{I}_{\mathcal{D}}^{\Sigma} = \left(\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top}\right)^\top \mathcal{I}_{\mathcal{D}}^{\Omega}\frac{\partial \mathrm{vech}(\mathbf{\Omega})}{\partial \mathrm{vech}(\mathbf{\Sigma})^\top}, \tag{39}$$

to get the Fisher information matrices w.r.t. $\mathbf{\Sigma}$.

Recall, that for the Wishart-type distributions, i.e. $\mathcal{D} \in (\mathcal{W}, \mathcal{W}, t\mathcal{W}, it\mathcal{W}, F)$,

we have $\mathbf{\Omega} = m_{\mathcal{D}}^{-1}\mathbf{\Sigma}$, with distribution-specific scalar $m_{\mathcal{D}}$ and thus

$$\frac{\partial \text{vech}(\mathbf{\Omega})}{\partial \text{vech}(\mathbf{\Sigma})^\top} = \mathbf{I} m_{\mathcal{D}}^{-1}.$$

Since, for all Wishart-type distributions, $\mathbf{\Omega}^{-1}$ enters in quadratic form in $\mathcal{I}_{\mathcal{D}}^{\Omega}$, applying formula (39) with $\mathcal{I}_{\mathcal{D}}^{\Omega}$ rewritten in terms of $\mathbf{\Omega}$, amounts to just replacing $\mathbf{\Omega}$ with $\mathbf{\Sigma}$. For example,

$$\begin{aligned}
\mathcal{I}_{\mathcal{W}}^{\Sigma} &= m_{\mathcal{D}}^{-2}\frac{n}{2}\mathbf{G}^\top \mathbf{\Omega}^{-\otimes 2}\mathbf{G} \\
&= m_{\mathcal{D}}^{-2}\frac{n}{2}\mathbf{G}^\top (m_{\mathcal{D}}^{-1}\mathbf{\Sigma})^{-\otimes 2}\mathbf{G} \\
&= \frac{n}{2}\mathbf{G}^\top \mathbf{\Sigma}^{-\otimes 2}\mathbf{G}.
\end{aligned}$$

Thus, the expressions in Table 2 follow.

6.3 Proofs of Theorems in Paper

6.3.1 Proof of Theorem 3.1

Proof. We omit the subscripts t. Note (Table 2) that the $\mathcal{I}^{\Sigma}$ of all Wishart-type have the form

$$\mathcal{I}_{\mathcal{D}}^{\Sigma} = \mathbf{G}^\top(\alpha_{\mathcal{D}}\mathbf{\Sigma}^{\otimes 2} + c_{\mathcal{D}}\,\text{vec}^2(\mathbf{\Sigma}))\mathbf{G},$$

where the scalars $\alpha_{\mathcal{D}}$ and $c_{\mathcal{D}}$ only depend on the d.o.f. parameter(s) of the respective distribution. Thus, using Lemma 6.1 we have

$$\begin{aligned}
(\mathcal{I}^{\Sigma})^{-1} &= \alpha_{\mathcal{D}}(\mathbf{G}^\top(\mathbf{\Sigma}^{\otimes 2} + \alpha_{\mathcal{D}}^{-1}c_{\mathcal{D}}\text{vec}^2(\mathbf{\Sigma}))\mathbf{G})^{-1} \\
&= \alpha_{\mathcal{D}}\mathbf{G}^+\left(\mathbf{\Sigma}^{-\otimes 2} + \frac{\alpha_{\mathcal{D}}^{-1}c_{\mathcal{D}}}{1+\alpha_{\mathcal{D}}^{-1}c_{\mathcal{D}}p}\text{vec}^2(\mathbf{\Sigma}^{-1})\right)\mathbf{G}^+ \\
&= \mathbf{G}^+\left(\alpha_{\mathcal{D}}\mathbf{\Sigma}^{-\otimes 2} + \frac{\alpha_{\mathcal{D}}c_{\mathcal{D}}}{\alpha_{\mathcal{D}}+c_{\mathcal{D}}p}\text{vec}^2(\mathbf{\Sigma}^{-1})\right)\mathbf{G}^+.
\end{aligned}$$

Define $\beta_{\mathcal{D}} = \frac{\alpha_{\mathcal{D}} c_{\mathcal{D}}}{\alpha_{\mathcal{D}} + c_{\mathcal{D}} p}$,

$$
\begin{aligned}
&\mathbf{G}^{+}(\alpha_{\mathcal{D}}\boldsymbol{\Sigma}^{\otimes 2} + \beta_{\mathcal{D}}\mathrm{vec}^{2}(\boldsymbol{\Sigma}))(\mathbf{G}^{+})^{\top}\mathbf{G}^{\top}\mathrm{vec}(\triangle_{\mathcal{D}}^{\Sigma}) \\
&\qquad = \mathbf{G}^{+}(\alpha_{\mathcal{D}}\boldsymbol{\Sigma}^{\otimes 2} + \beta_{\mathcal{D}}\mathrm{vec}^{2}(\boldsymbol{\Sigma}))\mathrm{vec}(\triangle_{\mathcal{D}}^{\Sigma}) \\
&\qquad = \alpha_{\mathcal{D}}\mathbf{G}^{+}\mathrm{vec}(\boldsymbol{\Sigma}\triangle_{\mathcal{D}}^{\Sigma}\boldsymbol{\Sigma}) + \beta_{\mathcal{D}}\mathrm{tr}(\boldsymbol{\Sigma}\triangle_{\mathcal{D}}^{\Sigma})\mathbf{G}^{+}\mathrm{vec}(\boldsymbol{\Sigma}) \\
&\qquad = \alpha_{\mathcal{D}}\mathrm{vech}(\boldsymbol{\Sigma}\triangle_{\mathcal{D}}^{\Sigma}\boldsymbol{\Sigma}) + \beta_{\mathcal{D}}\mathrm{tr}(\boldsymbol{\Sigma}\triangle_{\mathcal{D}}^{\Sigma})\mathrm{vech}(\boldsymbol{\Sigma})
\end{aligned}
$$

where we used equations (25), (12), and (14). Now simply apply the *ivech* operator. □

6.3.2 Proof of Theorem 3.1

Proof. We omit the subscripts t in the proof. We have

$$
\begin{aligned}
(\mathcal{I}_{\mathcal{R}}^{\Sigma})^{-1}\nabla_{\mathcal{R}}^{\Sigma} &= \mathbb{E}\left[\left(\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{\top}\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right]^{-1} \\
&\quad \times \left(\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{\top} \\
&= \left(\left(\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{\top}\mathbb{E}\left[\left(\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\right)^{\top}\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\right]\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{-1} \\
&\quad \times \left(\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{\top}\left(\frac{\partial p_{\mathcal{R}}}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}\right)^{\top} \\
&= \left(\frac{\partial \mathrm{vech}(\boldsymbol{\Omega})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})^{\top}}\right)^{-1}(\mathcal{I}_{\mathcal{R}}^{\Omega})^{-1}\nabla_{\mathcal{R}}^{\Omega} = \frac{\partial \mathrm{vech}(\boldsymbol{\Sigma})}{\partial \mathrm{vech}(\boldsymbol{\Omega})^{\top}}(\mathcal{I}_{\mathcal{R}}^{\Omega})^{-1}\nabla_{\mathcal{R}}^{\Omega} \\
&= \mathbf{G}^{+}(\mathbf{C}_{\Omega}\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1} \\
&\quad \times (\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G}\mathbf{G}^{+}(\mathbf{C}_{\Omega}\,\mathrm{dg}(\mathbf{n}) \otimes \mathbf{I})\mathbf{F}^{\top}(\mathbf{G}^{+}(\mathbf{C}_{\Omega} \otimes \mathbf{I})\mathbf{F}^{\top})^{-1})^{-1} \\
&\quad \times \mathbf{G}^{\top}\mathrm{vec}(\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}) \\
&= (\mathbf{G}^{\top}\boldsymbol{\Omega}^{-\otimes 2}\mathbf{G})^{-1}\mathbf{G}^{\top}\mathrm{vec}(\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}) \\
&= \mathbf{G}^{+}\boldsymbol{\Omega}^{\otimes 2}(\mathbf{G}\mathbf{G}^{+})^{\top}\mathrm{vec}(\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}) \\
&= \mathbf{G}^{+}\boldsymbol{\Omega}^{\otimes 2}\mathrm{vec}(\boldsymbol{\Omega}^{-1}\mathbf{R}\boldsymbol{\Omega}^{-1} - \mathbf{C}_{\Omega}^{-\top}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{-1}) \\
&= \mathrm{vech}(\mathbf{R} - \mathbf{C}_{\Omega}\mathrm{dg}(\mathbf{n})\mathbf{C}_{\Omega}^{\top}) = \mathrm{vech}(\mathbf{R} - \boldsymbol{\Sigma}),
\end{aligned}
$$

where $\mathcal{I}_{\mathcal{R}}^{\Omega}$ is given in equation (35) and we used Lemma 6.5, equations (27) and (26). Thus, the theorem follows from

$$\mathbf{S}_{\mathcal{R}}^{\Sigma} = \mathbf{R} - \mathbf{\Sigma}.$$

□

6.3.3 Riesz Covariance

For the derivation of the Fisher information matrix of the t-Riesz distribution we need the covariance matrix of $\text{vech}(\mathbf{R})$, if $\mathbf{R} \sim \mathcal{R}(\mathbf{\Sigma}, \mathbf{n})$.

Lemma 6.13. *Let* $\mathbf{R}$ *follow a Riesz distribution,* $\mathbf{R} \sim \mathcal{R}(\mathbf{\Sigma}, \mathbf{n})$ *and let* $\boldsymbol{\mathcal{K}}_{\mathcal{R}}$ *be its stochastic representation kernel as defined in Stollenwerk (2023). Then*

$$
\begin{aligned}
\mathbb{E}[\text{vech}^2(\boldsymbol{\mathcal{K}}_{\mathcal{R}})] &= \text{vech}^2(\text{dg}(\mathbf{n})) + 2\mathbf{G}^+(\text{dg}(\mathbf{n}) \otimes \mathbf{I})(\mathbf{G}^+)^\top, \\
\mathbf{Cov}\left(\text{vech}\left(\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right)\right) &= 2\mathbf{G}^+\left(\text{dg}(\mathbf{n}) \otimes \mathbf{I}\right)\left(\mathbf{G}^+\right)^\top, \\
\mathbb{E}[\text{vech}^2(\mathbf{R})] &= \text{vech}^2(\mathbf{\Sigma}) + 2\mathbf{G}^+(\mathbf{\Sigma} \otimes \mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top)(\mathbf{G}^+)^\top, \text{ and} \\
\mathbf{Cov}\left(\text{vech}(\mathbf{R})\right) &= 2\mathbf{G}^+\left(\mathbf{\Sigma} \otimes \mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top\right)\left(\mathbf{G}^+\right)^\top.
\end{aligned}
$$

Proof. First, see §2.1.2 of Kollo and von 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 $\text{vec}(\mathbf{R})$, or we take it into account by getting the characteristic function of e.g. $\text{vech}\left(\mathbf{R}\right)$. 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 von Rosen (2005) do. The characteristic function of $\text{vech}\left(\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right)$ where, as defined in Stollenwerk (2023), $\boldsymbol{\mathcal{K}}_{\mathcal{R}} = \underline{\mathbf{B}}\underline{\mathbf{B}}^\top$ is given by

$$
\begin{aligned}
\phi(\mathbf{Z}) &= \mathbb{E}\left[e^{i\,\text{vech}(\mathbf{Z})^\top \text{vech}(\boldsymbol{\mathcal{K}}_{\mathcal{R}})}\right] \\
&= \mathbb{E}\left[\text{etr}\left(i\frac{1}{2}\left(\mathbf{Z}+\mathbf{Z}\right)\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right)\right] \text{ p. 244 Kollo and von Rosen (2005)} \\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\int_{\boldsymbol{\mathcal{K}}_{\mathcal{R}}>\mathbf{0}} |\boldsymbol{\mathcal{K}}_{\mathcal{R}}|_{\frac{\mathbf{n}-p-1}{2}}\,\text{etr}\left(i\frac{1}{2}\left(\mathbf{Z}+\mathbf{Z}\right)\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right)\text{etr}\left(-\frac{1}{2}\mathbf{I}\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right) \\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\int_{\boldsymbol{\mathcal{K}}_{\mathcal{R}}>\mathbf{0}} |\boldsymbol{\mathcal{K}}_{\mathcal{R}}|_{\frac{\mathbf{n}-p-1}{2}}\,\text{etr}\left(-\frac{1}{2}\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)\boldsymbol{\mathcal{K}}_{\mathcal{R}}\right) \\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\mathbf{n}/2\right)}\Gamma_p\left(\mathbf{n}/2\right)\left|\frac{1}{2}\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}} \\
&= \frac{1}{2^{p\bar{\mathbf{n}}/2}}2^{p\bar{\mathbf{n}}/2}\left|\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}} \\
&= \left|\left(\mathbf{I}-i\left(\mathbf{Z}+\mathbf{Z}\right)\right)^{-1}\right|_{\frac{\mathbf{n}}{2}},
\end{aligned}
$$

where $\mathbf{Z}$ is a diagonal matrix with elements $\text{dg}(\mathbf{Z})$. See also Díaz-García (2013),

Lemma 1.

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

$$\frac{\partial \text{vech}(\boldsymbol{\Xi})}{\partial \text{vech}(\mathbf{Z})^\top} = -i2(\mathbf{G}^\top\mathbf{G})^{-1} = -i2(\mathbf{G}^\top\mathbf{G})^{-\top}$$

Then

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

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

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

and

$$\begin{aligned}
\mathbb{E}[\text{vec}^2(\mathbf{R})] &= \mathbf{C}_\Omega^{\otimes 2}\mathbb{E}[\text{vec}^2(\boldsymbol{\mathcal{K}}_\mathcal{R})](\mathbf{C}_\Omega^{\otimes 2})^\top \\
&= \mathbf{C}_\Omega^{\otimes 2}\mathbf{G}\mathbb{E}[\text{vech}^2(\boldsymbol{\mathcal{K}}_\mathcal{R})]\mathbf{G}^\top(\mathbf{C}_\Omega^{\otimes 2})^\top \\
&= \mathbf{C}_\Omega^{\otimes 2}\text{vec}^2(\text{dg}(\mathbf{n}))(\mathbf{C}_\Omega^{\otimes 2})^\top + 2(\mathbf{C}_\Omega^{\otimes 2})\mathbf{G}\mathbf{G}^+(\text{dg}(\mathbf{n})\otimes\mathbf{I})(\mathbf{C}_\Omega^{\otimes 2}\mathbf{G}\mathbf{G}^+)^\top \\
&= \text{vec}^2(\mathbf{C}_\Omega\text{dg}(\mathbf{n})\mathbf{C}_\Omega^\top) + 2\mathbf{G}\mathbf{G}^+(\mathbf{C}_\Omega\text{dg}(\mathbf{n})\mathbf{C}_\Omega^\top\otimes\boldsymbol{\Omega})\mathbf{G}\mathbf{G}^+, \qquad (40)
\end{aligned}$$

and consequently

$$\begin{aligned}\mathbb{E}[\text{vech}^2(\mathbf{R})] &= \mathbf{G}^+\mathbb{E}[\text{vec}^2(\mathbf{R})](\mathbf{G}^+)^\top \\ &= \text{vech}^2(\mathbf{C}_\Omega \text{dg}(\mathbf{n})\mathbf{C}_\Omega^\top) + 2\mathbf{G}^+(\mathbf{C}_\Omega \text{dg}(\mathbf{n})\mathbf{C}_\Omega^\top \otimes \boldsymbol{\Omega})(\mathbf{G}^+)^\top \\ &= \text{vech}^2(\boldsymbol{\Sigma}) + 2\mathbf{G}^+(\boldsymbol{\Sigma} \otimes \mathbf{C}\text{dg}(\mathbf{n})^{-1}\mathbf{C}^\top)(\mathbf{G}^+)^\top.\end{aligned}$$

Thus

$$\mathbf{Cov}(\text{vech}(\boldsymbol{\mathcal{K}}_\mathcal{R})) = 2\mathbf{G}^+(\text{dg}(\mathbf{n}) \otimes \mathbf{I})(\mathbf{G}^+)^\top,$$

and

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

□

6.4 Additional Material

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			$\hat{a}_1 \times 100$			
Wishart	1.704	1.209	0.902	0.515	0.348	0.354
Riesz	1.193	0.824	0.601	0.325	0.239	0.227
Inv.Wishart	0.112	0.556	0.332	0.119	0.161	0.173
Inv.Riesz	0.110	0.478	0.301	0.112	0.143	0.155
t-Wishart	1.038	0.777	0.669	0.439	0.352	0.341
t-Riesz	0.838	0.660	0.535	0.286	0.249	0.241
Inv.t-Wishart	0.708	0.585	0.464	0.311	0.221	0.214
Inv.t-Riesz	0.635	0.522	0.396	0.255	0.191	0.185
F	0.890	0.713	0.538	0.337	0.251	0.239
F-Riesz	0.470	0.366	0.300	0.203	0.168	0.156
Inv.F-Riesz	0.546	0.412	0.332	0.215	0.168	0.165
			$\hat{a}_2 \times 100$			
Wishart	0.619	0.516	0.263	0.206	0.086	0.083
Riesz	0.818	0.631	0.356	0.256	0.104	0.101
Inv.Wishart	0.319	0.397	0.212	0.142	0.065	0.068
Inv.Riesz	0.357	0.435	0.238	0.149	0.074	0.076
t-Wishart	4.247	4.361	3.278	3.501	2.082	2.194
t-Riesz	4.287	4.505	3.497	4.163	2.261	2.367
Inv.t-Wishart	3.237	3.399	2.713	3.103	1.653	1.733
Inv.t-Riesz	3.412	3.394	2.857	3.343	1.673	1.793
F	0.594	0.454	0.256	0.160	0.083	0.080
F-Riesz	1.152	0.902	0.489	0.290	0.141	0.137
Inv.F-Riesz	1.098	0.871	0.464	0.272	0.131	0.128

Table 11: Estimated scaled score parameters $\hat{a}_1 \times 100$ and $\hat{a}_2 \times 100$ for our restricted GAS model where d.o.f. parameters are restricted to be constant. All estimates are highly significant. The median (smallest) t-statistic of $\hat{a}_1 \times 100$ is 240 (61) and of $\hat{a}_2 \times 100$ is 151 (39).

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	0.983	0.988	0.991	0.997	0.995	0.988
Riesz	0.989	0.992	0.996	0.999	0.997	0.993
Inv.Wishart	0.998	0.996	0.997	1.000	0.998	0.997
Inv.Riesz	0.998	0.996	0.998	1.000	0.999	0.997
t-Wishart	0.992	0.994	0.996	0.999	0.996	0.991
t-Riesz	0.994	0.995	0.996	0.999	0.997	0.994
Inv.t-Wishart	0.994	0.996	0.997	0.999	0.998	0.996
Inv.t-Riesz	0.995	0.996	0.998	1.000	0.998	0.997
F	0.993	0.995	0.996	0.999	0.998	0.996
F-Riesz	0.997	0.998	0.999	1.000	0.999	0.998
Inv.F-Riesz	0.997	0.997	0.999	1.000	0.999	0.998

Table 12: Persistence parameter $\hat{b}_1+\hat{b}_2+\hat{b}_3$ for our restricted GAS model where d.o.f. parameters are restricted to be constant. All estimated parameters are highly significant, with the median (minimum) t-statistic being 248 (72), 41 (10), and 76 (11), respectively.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	0.983	0.989	0.990	0.997	0.995	0.988
Riesz	0.989	0.993	0.996	0.999	0.997	0.994
Inv.Wishart	0.997	0.996	0.997	0.999	0.998	0.997
Inv.Riesz	0.998	0.996	0.998	1.000	0.999	0.997
t-Wishart	0.992	0.994	0.996	0.999	0.997	0.992
t-Riesz	0.994	0.995	0.997	0.999	0.998	0.996
Inv.t-Wishart	0.994	0.996	0.997	0.999	0.998	0.996
Inv.t-Riesz	0.996	0.996	0.998	1.000	0.999	0.997
F	0.994	0.995	0.997	0.999	0.998	0.997
F-Riesz	0.997	0.998	0.999	1.000	0.999	0.998
Inv.F-Riesz	0.997	0.997	0.999	1.000	0.999	0.998

Table 13: Persistence parameter $\hat{b}_1 + \hat{b}_2 + \hat{b}_3$. All estimated parameters are highly significant, with the median (minimum) t-statistic being 142 (75), 45 (10), and 75 (11), respectively.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			$\bar{\boldsymbol{\xi}}^{\mathbf{n}}$, ξ^n			
Wishart	18	21	26	34	46	47
Riesz	17	20	24	31	41	42
Inv.Wishart	-	-	-	-	-	-
Inv.Riesz	-	-	-	-	-	-
t-Wishart	27	32	35	44	53	53
t-Riesz	24	29	31	39	46	46
Inv.t-Wishart	33	32	39	31	48	49
Inv.t-Riesz	33	33	37	30	50	49
F	100	110	156	194	175	256
F-Riesz	80	86	108	125	144	155
Inv.F-Riesz	157	2480	1756	796	976	2539
			$\bar{\boldsymbol{\xi}}^{\boldsymbol{\nu}}$, ξ^{ν}			
Wishart	-	-	-	-	-	-
Riesz	-	-	-	-	-	-
Inv.Wishart	31	28	35	52	59	59
Inv.Riesz	21	26	32	41	53	53
t-Wishart	23	24	25	21	31	28
t-Riesz	24	23	27	22	34	31
Inv.t-Wishart	29	35	40	51	61	62
Inv.t-Riesz	27	33	37	47	56	57
F	30	36	44	57	86	81
F-Riesz	22	26	31	41	56	55
Inv.F-Riesz	32	39	45	54	66	67

Table 14: (Mean of) estimated intercept of d.o.f. parameters specification, $\bar{\boldsymbol{\xi}}^{\mathbf{n}}$ and $\bar{\boldsymbol{\xi}}^{\boldsymbol{\nu}}$ (ξ^n and ξ^{ν} for Wishart-type distributions), that is, (mean of) estimated unconditional mean of d.o.f parameters. All are significant at the 1 % significance level.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
			$\widehat{b^{\mathbf{n}}}$, $\widehat{b^{n}}$			
Wishart	0.998	0.996	0.998	0.997	0.998	0.998
Riesz	0.997	0.992	0.996	0.996	0.998	0.998
Inv.Wishart	-	-	-	-	-	-
Inv.Riesz	-	-	-	-	-	-
t-Wishart	0.983	0.976	0.979	0.817	0.983	0.978
t-Riesz	0.996	0.993	0.998	0.990	0.998	0.998
Inv.t-Wishart	0.966	0.953	0.939	0.958	0.968	0.953
Inv.t-Riesz	0.913	0.948	0.944	0.959	0.965	0.952
F	0.671	1.000	0.545	0.743	1.000	0.851
F-Riesz	0.940	0.946	0.934	0.943	0.962	0.980
Inv.F-Riesz	0.991	0.961	0.998	0.713	0.964	0.977
			$\widehat{b^{\boldsymbol{\nu}}}$, $\widehat{b^{\nu}}$			
Wishart	-	-	-	-	-	-
Riesz	-	-	-	-	-	-
Inv.Wishart	1.000	0.939	0.908	0.999	0.997	0.980
Inv.Riesz	0.997	0.695	0.993	0.996	0.997	0.998
t-Wishart	0.947	0.953	0.969	0.902	0.953	0.936
t-Riesz	0.952	0.969	0.964	0.951	0.976	0.960
Inv.t-Wishart	0.963	0.923	0.930	0.791	0.945	0.939
Inv.t-Riesz	0.982	0.951	0.978	0.990	0.997	0.999
F	0.706	0.864	0.756	0.830	0.999	1.000
F-Riesz	0.002	0.225	0.972	0.924	0.993	0.988
Inv.F-Riesz	0.903	0.798	0.524	0.925	0.969	0.973

Table 15: Estimated GARCH parameters $\widehat{b^{\mathbf{n}}}$ and $\widehat{b^{\boldsymbol{\nu}}}$ ($\widehat{b^{n}}$ and $\widehat{b^{\nu}}$ for Wishart-type distributions). No background color indicates significance at the 1% level, light gray and mid-light gray indicate significance at the 5% and 10% level, dark gray indicates insignificance.