Probability Distributions for Realized Covariance Measures

Michael Stollenwerk

Heidelberg University

October 5, 2022

Abstract

Realized covariance measures (RCs) are an essential input to assess the risks involved in different investment allocations and it is thus useful to model and forecast them. Thus, a realistic distributional assumption is essential. We compare all probability distributions hitherto applied to time series of RCs in the literature. We derive them in an intuitive and unified framework based on their stochastic representations in terms of random lower and upper triangular (Barlett) matrices. Furthermore, we derive a novel family of probability distributions, which has a property called "tail homogeneity". That is, in times of crisis periods, i.e. large RCs, this family assumes high dependence between the individual entries in the RCs. Finally, we show rigorously, how the considered distributions are related to each other. Empirically, we confirm in an in-sample fit experiment previous results that "fat-tailed" distributions outperform others and show that the novel distribution family achieves a very good fit. Out-of-sample forecasting comparisons further corroborate the excellent performance of the novel distribution family.

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

1 Introduction

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

In this paper, we derive all hitherto used probability distributions for realized covariance measures in an intuitive and unified framework based on their stochastic representations in terms of random lower and upper triangular (Barlett) matrices.[2] We show explicitly how the different distributions are related to each other and derive missing distributions (the t-Riesz distribution family and the Inverse F-Riesz distribution) to complete the picture of distribution relationships. The novel t-Riesz distribution family features tail homogeneity in each of its coordinates in contrast to the tail heterogeneity of the recently proposed F-Riesz distribution by Blasques et al. (2021). We show that, especially in times of high market volatility and for financial assets of the same industry sector, tail homogeneity is a more realistic assumption to the realized covariance (RC) time series data. Furthermore, the novel t-Riesz distribution family can be rooted in a realistic low-level assumption on the intraday return vectors (from which the realized covariance measures are constructed) and is easy to handle analytically and numerically. Finally, we compare all distributions in terms of their fit to different data sets of

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

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

time series of RCs and assess their forecasting performance.

The previously proposed probability distributions for realized covariance measures are the (Non-Central) Wishart (Golosnoy, Gribisch, and Liesenfeld 2012, Gorgi et al. 2019, Yu, Li, and Ng 2017), Inverse Wishart (Gourieroux, Jasiak, and Sufana 2009, Asai and So 2013), Matrix-F (Opschoor et al. 2018, Zhou et al. 2019), Riesz (Gribisch and Hartkopf 2022), Inverse Riesz and F-Riesz (both Blasques et al. 2021) distributions. All distributions in this paper except for the Riesz and Wishart distributions can, in some sense, accommodate the well-known stylized fact that realized covariance measures are "fat tailed". It is, however, often argued that assuming simple non-fat-tailed conditional distributions with time-varying parameters yields fat-tailed unconditional distributions. Thus, they say, assuming more complex fat-tailed distributions is pointless. This paper will make it evident that this is false and that there is merit in assuming more complex fat-tailed distributions. Also, to this point, Bai and Chen (2008) test for the distributional assumption in multivariate GARCH models for return vectors and find that even with time-varying covariance matrix (i.e. GARCH dynamics), the normal distribution is rejected, whereas the t-distribution is not. Furthermore, in the age of modern computers, it certainly cannot hurt to assume a tractable, more realistic distribution on the data prior to adding time-varying parameters.

The rest of this paper is structured as follows. The next section derives and discusses all probability distributions, Section 3 takes a closer look at the newly derived t-Riesz distribution family and how it compares to the F-Riesz distribution. Section 4 introduces a time variation in the expected value parameters of the distributions, 5 discusses estimation, 6 performs the in-sample and out-of-sample empirical analysis and 7 concludes.

2 Probability Distributions

All hitherto in the literature considered probability distributions for RCs[3] can be generated from the $p \times p$ random triangular matrices

$$\underline{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{n_1-1+1}} & & 0 \\ & \sqrt{\chi^2_{n_2-2+1}} & \\ & & \ddots \\ \mathcal{N}(0,1) & & \sqrt{\chi^2_{n_p-p+1}} \end{bmatrix}$$

and/or

$$\bar{\mathbf{B}} = \begin{bmatrix} \sqrt{\chi^2_{\nu_1-p+1}} & & \mathcal{N}(0,1) \\ & \sqrt{\chi^2_{\nu_2-p+2}} & \\ & & \ddots \\ 0 & & \sqrt{\chi^2_{\nu_p-p+p}}, \end{bmatrix},$$

that is

$$(\underline{\mathbf{B}})_{ij} \sim \begin{cases} \mathcal{N}(0,1) & \text{for } i < j, \\ \chi_{n_i-i+1} & \text{for } i = j \end{cases} \tag{1}$$

and/or

$$(\bar{\mathbf{B}})_{ij} \sim \begin{cases} \chi_{\nu_i-p+i} & \text{for } i = j, \\ \mathcal{N}(0,1) & \text{for } i > j, \end{cases} \tag{2}$$

where all random variables inside the matrices are independent of each other.[4] We refer to these random matrices as Barlett matrices in reference to the well-known Barlett decomposition of Wishart random matrices and refer to the n_i and ν_i as the degree of freedom (dof) parameters. We collect the dof parameters in the $p \times 1$ vectors $\mathbf{n} = (n_1, \ldots, n_p)^\top$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^\top$ and denote the special cases of the Barlett matrices, where for all i, $n_i = n$ and $\nu_i = \nu$ as $\underline{\mathcal{B}}$ and $\bar{\mathcal{B}}$, respectively.

3. I.e. distributions with support on symmetric positive (semi)definite matrices.
4. The χ_n distribution is given in e.g. Walck (2007), Section 8.14.

Note that for $p = 1$, the random matrices reduce to the random variables χ_n and χ_ν, respectively. For the matrix distributions to exist it is necessary and sufficient that we restrict $n_i > i - 1$ and $\nu_i > p - i$, since otherwise the χ distributions on the main diagonals would not exist.[5]

Let $\mathcal{D} \in (\mathcal{W}, i\mathcal{W}, t\mathcal{W}, it\mathcal{W}, F, \mathcal{R}, i\mathcal{R}, t\mathcal{R}, it\mathcal{R}, F\mathcal{R}, iF\mathcal{R})$ denote the different probability distributions, see the Introduction or any of the Tables below for the full names of the distributions. The ones in green are novel distributions derived in this paper. The stochastic representation of any $\mathcal{D}$ can be written as

$$\mathbf{R}_{\mathcal{D}} = \mathbf{C}_{\mathbf{\Omega}} \boldsymbol{\mathcal{K}}_{\mathcal{D}} \mathbf{C}_{\mathbf{\Omega}}^{\top}, \tag{3}$$

where $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ is a distribution specific function of one or both of the Barlett matrices or their special cases given in Table 1 and $\mathbf{C}_{\mathbf{\Omega}}$ denotes the lower Cholesky factor of symmetric positive definite parameter matrix $\mathbf{\Omega}_{\mathcal{D}}$. The distribution parameters are thus given by the degree of freedom parameter(s) in $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ (one or two of the set $(n, \nu, \mathbf{n}, \boldsymbol{\nu})$), which we collect in the vector $\boldsymbol{\theta}_{\mathcal{D}}$,[6] and the parameter matrix $\mathbf{\Omega}_{\mathcal{D}}$. Finally we can write for $p \times p$ realized covariance matrix $\mathbf{R}$,

$$\mathbf{R} \sim \mathcal{D}(\mathbf{\Omega}, \boldsymbol{\theta}),$$

where from now on, we will omit the subscripts $\mathcal{D}$ for the parameters $\mathbf{\Omega}$ and $\boldsymbol{\theta}$ and for the $\mathbf{R}$, they are always to be understood as specific to the chosen distribution.

We now standardize the distributions. That is, we characterize them in terms of their $p \times p$ symmetric positive definite expected value matrix

$$\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^{\top} := \mathbb{E}\left[\mathbf{R}\right], \tag{4}$$

where $\mathbf{C}$ is the lower Cholesky factor, instead of in terms of $\mathbf{\Omega}$. This allows for a simple two-step estimation strategy, where the $\mathcal{O}(p^2)$ parameter matrix $\mathbf{\Sigma}$ is

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

6. The exact composition of $\boldsymbol{\theta}$ for the different distributions is given in Table 10 in Section 8.2 of the appendix. It is also easy to see from e.g. Table 1, which degree of freedom parameter(s) characterize which distribution.

Distribution	$\boldsymbol{\mathcal{K}}_{\mathcal{D}}$	Distribution	$\boldsymbol{\mathcal{K}}_{\mathcal{D}}$
Wishart	$\underline{\boldsymbol{\mathcal{B}}}\underline{\boldsymbol{\mathcal{B}}}^{\top}$	Riesz	$\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}$
Inv.Wishart	$\bar{\boldsymbol{\mathcal{B}}}^{-\top}\bar{\boldsymbol{\mathcal{B}}}^{-1}$	Inv.Riesz	$\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}$
t-Wishart	$(\bar{b})^{-2}\underline{\boldsymbol{\mathcal{B}}}\underline{\boldsymbol{\mathcal{B}}}^{\top}$	t-Riesz	$(\bar{b})^{-2}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}$
Inv.t-Wishart	$(\underline{b})^{2}\bar{\boldsymbol{\mathcal{B}}}^{-\top}\bar{\boldsymbol{\mathcal{B}}}^{-1}$	Inv.t-Riesz	$(\underline{b})^{2}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}$
F	$\bar{\boldsymbol{\mathcal{B}}}^{-\top}\underline{\boldsymbol{\mathcal{B}}}\underline{\boldsymbol{\mathcal{B}}}^{\top}\bar{\boldsymbol{\mathcal{B}}}^{-1}$	F-Riesz	$\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}$
F	$\underline{\boldsymbol{\mathcal{B}}}\bar{\boldsymbol{\mathcal{B}}}^{-\top}\bar{\boldsymbol{\mathcal{B}}}^{-1}\underline{\boldsymbol{\mathcal{B}}}^{\top}$	Inv.F-Riesz	$\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}$

Table 1: Stochastic representations of all considered probability distributions. $\underline{b}$ and $\bar{b}$ are χ_n and χ_ν distributed random variables, thus one-dimensional $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$, respectively. The distributions in green are novel distributions derived in this paper.

estimated by its obvious method of moments estimator in the first step. Define

$$\mathbf{M}_{\mathcal{D}} := \mathbb{E}\left[\boldsymbol{\mathcal{K}}_{\mathcal{D}}\right], \tag{5}$$

then the stochastic representation of the standardized distributions has to be

$$\mathbf{R} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top}, \tag{6}$$

since

$$\mathbb{E}\left[\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top}\right] = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbb{E}\left[\boldsymbol{\mathcal{K}}_{\mathcal{D}}\right]\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top} = \boldsymbol{\Sigma}. \tag{7}$$

This implies that

$$\mathbf{C}_{\boldsymbol{\Omega}} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}} \Leftrightarrow \mathbf{C} = \mathbf{C}_{\boldsymbol{\Omega}}\mathbf{M}_{\mathcal{D}}^{\frac{1}{2}} \text{ and } \boldsymbol{\Omega} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}^{\top} \Leftrightarrow \mathbf{C} = \mathbf{C}_{\boldsymbol{\Omega}}\mathbf{M}_{\mathcal{D}}\mathbf{C}_{\boldsymbol{\Omega}}.$$

Theorem 2.1. *Let* $\underline{\mathbf{B}}$ *and* $\bar{\mathbf{B}}$ *be defined as above. Then*

$$\begin{aligned}
\mathbb{E}\left[\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\right] &= \mathrm{dg}\left(\mathbf{n}\right), \\
\mathbb{E}\left[\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}\right)^{-1}\right] &= \mathrm{dg}\left(\overset{iR''}{\mathbf{m}}\right), \\
\mathbb{E}\left[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}\right] &= \mathrm{dg}\left(\overset{FR'}{\mathbf{m}}\right) \quad \textit{and} \\
\mathbb{E}\left[\underline{\mathbf{B}}\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}\right)^{-1}\underline{\mathbf{B}}^{\top}\right] &= \mathrm{dg}\left(\overset{iFR''}{\mathbf{m}}\right),
\end{aligned}$$

Distribution	$\mathbf{M}_{\mathcal{D}}$	Distribution	$\mathbf{M}_{\mathcal{D}}$
Wishart	$\mathbf{I}n$	Riesz	$\mathrm{dg}(\mathbf{n})$
Inv.Wishart	$\mathbf{I}\frac{1}{\nu-p-1}$	Inv.Riesz	$\mathrm{dg}(\overset{iR''}{\mathbf{m}})$
t-Wishart	$\mathbf{I}\frac{n}{\nu-2}$	t-Riesz	$\mathrm{dg}(\mathbf{n})\frac{1}{\nu-2}$
Inv.t-Wishart	$\mathbf{I}\frac{n}{\nu-p-1}$	Inv.t-Riesz	$\mathrm{dg}(\overset{iR''}{\mathbf{m}})n$
F	$\mathbf{I}\frac{n}{\nu-p-1}$	F-Riesz	$\mathrm{dg}(\overset{FR'}{\mathbf{m}})$
F	$\mathbf{I}\frac{n}{\nu-p-1}$	Inv.F-Riesz	$\mathrm{dg}(\overset{iFR''}{\mathbf{m}})$

Table 2: Expectations of $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$.

with the $p \times 1$ *vectors*

$$
\begin{aligned}
\mathbf{n} &= (n_1, n_2, \ldots, n_p)^\top, \\
\overset{iR''}{\mathbf{m}} &= (\overset{iR''}{m}_1, \overset{iR''}{m}_2, \ldots, \overset{iR''}{m}_p)^\top, \\
\overset{iR''}{m}_i &= \begin{cases} \frac{1}{\nu_i - p - 1}, & \text{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(1 + \sum_{j=1}^{j-1} \overset{iR''}{m}_j\right) & \text{for } i = 2, \ldots, p, \end{cases} \\
\overset{FR'}{\mathbf{m}} &= (\overset{FR'}{m}_1, \overset{FR'}{m}_2, \ldots, \overset{FR'}{m}_p)^\top, \\
\overset{FR'}{m}_i &= \begin{cases} \frac{n_1}{\nu_1 - p - 1}, & \text{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(n_i + \sum_{j=1}^{i-1} \overset{FR'}{m}_j\right) & \text{for } i = 2, \ldots, p \end{cases} \quad \text{and} \qquad (8) \\
\overset{iFR''}{\mathbf{m}} &= (\overset{iFR''}{m}_1, \overset{iFR''}{m}_2, \ldots, \overset{iFR''}{m}_p)^\top \qquad (9) \\
\overset{iFR''}{m}_i &= \begin{cases} n_1 \overset{iR''}{m}_1, & \text{for } i = 1 \\ \sum_{j=1}^{i-1} \overset{iR''}{m}_j + (n_i - i + 1) \overset{iR''}{m}_i, & \text{for } i = 2, \ldots, p. \end{cases}
\end{aligned}
$$

Note that if $\forall i$, $n_i = n$, $\nu_i = \nu$, *then* $\forall i$, $\overset{iR''}{m}_i = \frac{1}{\nu-p-1}$ *and* $\overset{FR'}{m}_i = \overset{iFR''}{m}_i = \frac{n}{\nu-p-1}$. *Furthermore, if* $p = 1$ *we directly get the expectations of a* χ^2_n *and inverse* χ^2_ν *distribution, as well as of* χ^2_n/χ^2_ν.

Proof in Appendix.

The expectations $\mathbf{M}_{\mathcal{D}}$ are straightforward to derive by applying Theorem 2.1. They are listed in Table 2. Notice that they all are diagonal matrices, so if all its diagonal elements are non-negative, $\mathbf{M}_{\mathcal{D}}^{-1/2}$ is uniquely defined. If an element is negative and the conditions for existence of the distributions are fulfilled, then the

expected value $\boldsymbol{\Sigma}$ does not exist. In this paper, we assume that the expectations always exist, and we can thus equivalently characterize the distribution with $\boldsymbol{\Sigma}$ and write[7]

$$\mathbf{R} \sim \mathcal{D}(\boldsymbol{\Sigma}, \boldsymbol{\theta}). \tag{10}$$

7. Whether we mean the standardized distribution or the one in the original form will always be obvious by whether we parameterize it with $\boldsymbol{\Sigma}$ or $\boldsymbol{\Omega}$

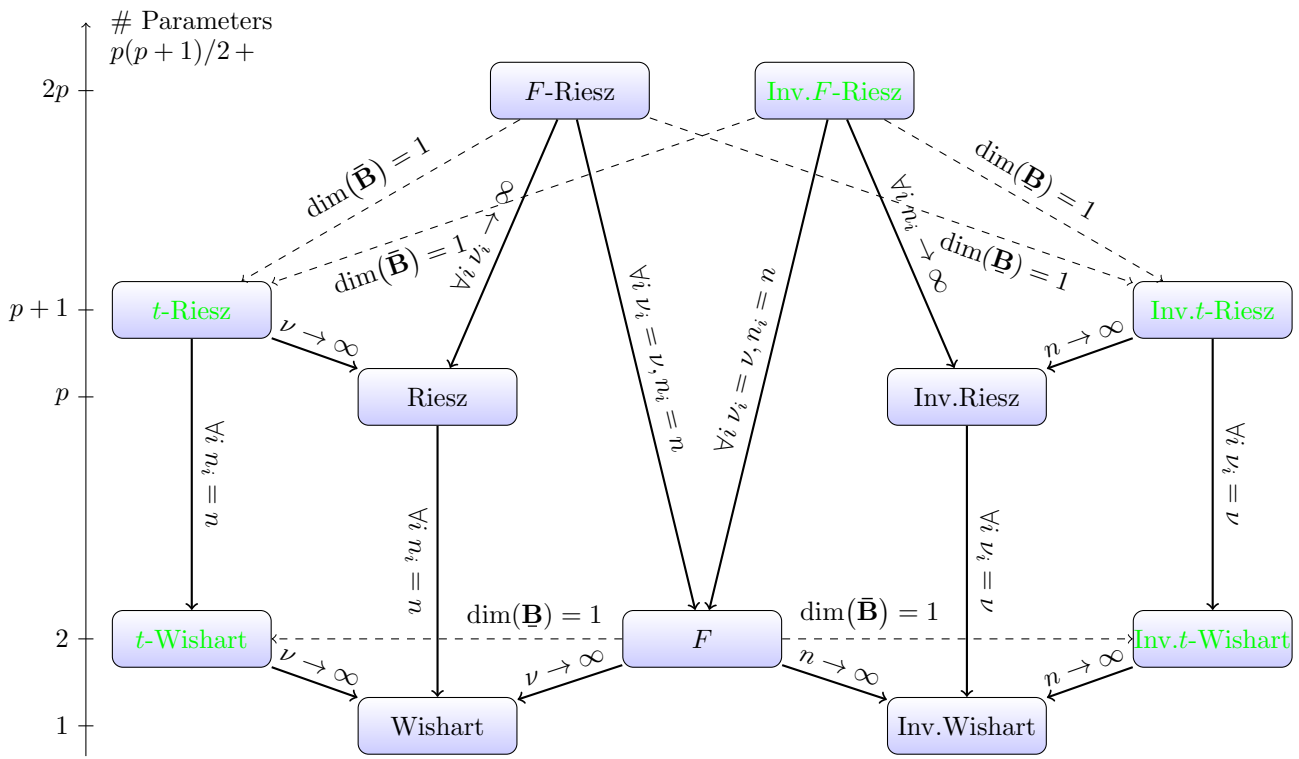

Figure 1: Relationships between the standardized probability distributions with pdfs in Table 3 and stochastic representations given by equation (6) in conjunction with Tables 1 and 2. Above every arrow, we indicate how the distributions are related. Dashed arrows indicate a more general relationship, not a nesting. On the vertical axis we can see the number of parameters of the respective distributions. All distributions have $p(p+1)/2$ distinct parameters in the symmetric positive definite matrix $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) plus the number of dof parameters.

Figure 1 shows how the hitherto considered distributions relate to each other. Every Wishart-type distribution is a special case of its Riesz-type counterpart and is obtained by setting the entries in the dof parameter vector all equal to each other. This is easy to see from the stochastic representations. The proofs for the dashed arrows are also immediately evident since it is easy to see from the stochastic representations in Table 1 that in the case of $\bar{\mathbf{B}}$ ($\underline{\mathbf{B}}$) being one-dimensional, the stochastic representation of the (Inverse) F-Riesz distribution reduces to the stochastic representation of the (Inverse) t-Riesz distribution. We derive the remaining relationships in the following theorem.

Theorem 2.2.

$$\begin{aligned}
\mathbf{C}\mathbf{M}_{F\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{F\mathcal{R}}\mathbf{M}_{F\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} &\xrightarrow[\forall i:\nu_i\to\infty]{d} \mathbf{C}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{R}}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \\
\mathbf{C}\mathbf{M}_{iF\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{iF\mathcal{R}}\mathbf{M}_{iF\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} &\xrightarrow[\forall i:n_i\to\infty]{d} \mathbf{C}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{i\mathcal{R}}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \\
\mathbf{C}\mathbf{M}_{t\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{t\mathcal{R}}\mathbf{M}_{t\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} &\xrightarrow[\nu\to\infty]{d} \mathbf{C}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{R}}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \\
\mathbf{C}\mathbf{M}_{it\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{it\mathcal{R}}\mathbf{M}_{it\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} &\xrightarrow[n\to\infty]{d} \mathbf{C}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{i\mathcal{R}}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top}.
\end{aligned}$$

Proof in Appendix.

Theorem 2.2 holds for the corresponding Wishart-type distributions as well since they are just special cases of the Riesz-type distributions with the degree of freedom parameter vector entries being all equal to each other. Finally, note that an inverse F distribution is again an F distribution[8], but an inverse F-Riesz distribution is not again an F-Riesz distribution.

Note that for all Riesz-type distributions, a different ordering of the assets in the RCs yields a different "version" of the respective probability distribution. That is, if we assume $\mathbf{R} \sim \mathcal{D}(\boldsymbol{\Omega}, \boldsymbol{\theta})$, with stochastic representation $\mathbf{R} = \mathbf{C}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}(\mathbf{C}_{\boldsymbol{\Omega}})^{\top}$, then any other ordering the assets, $\mathbf{P}\mathbf{R}\mathbf{P}^{\top}$, where $\mathbf{P}$ denotes an arbitrary permutation matrix, has stochastic representation $\mathbf{P}\mathbf{C}_{\boldsymbol{\Omega}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}(\mathbf{C}_{\boldsymbol{\Omega}})^{\top}\mathbf{P}^{\top}$, which yields for Riesz-type distributions a different and generally unknown probability distribution, which we call $\mathcal{D}_{\mathbf{P}}(\boldsymbol{\Omega}, \boldsymbol{\theta})$. In practice, we are given a randomly ordered RC

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

$\mathbf{PRP}^\top$ for which the model $\mathcal{D}(\boldsymbol{\Omega}, \boldsymbol{\theta})$ is only correctly specified if, by chance, the random ordering corresponds to the true one ($\mathbf{P} = \mathbf{I}$). To recover the true ordering we can treat the ordering as a parameter to optimize over. Blasques et al. (2021) have proposed an efficient algorithm to optimize the likelihood for many different orderings[9] and then choosing the ordering with the highest estimated likelihood value. In a simulation experiment, they find that their algorithm's estimated likelihood values come close to the likelihood value of the true data generating process and that the ordering of the assets gets close to the true ordering.

Finally, we note that for the special case that $\mathbf{P}$ is the exchange matrix, that is a matrix with ones on the diagonal from the upper right-hand corner to the lower left-hand corner and zeros elsewhere,

$$\mathbf{P}_e = \begin{bmatrix} 0 & & 1 \\ & & \\ 1 & & 0 \end{bmatrix},$$

the reordering corresponds to simply reversing the asset order, and for this case, the pdfs, $\mathcal{D}_{\mathbf{P}_e}(\boldsymbol{\Omega}, \boldsymbol{\theta})$, have been derived. The two versions in Blasques et al. (2021) for each Riesz-type distribution correspond to the original and the reverse ordered RCs.[10] For example, if we have $\mathbf{R} \sim F\mathcal{R}(\boldsymbol{\Omega}, \mathbf{n}, \boldsymbol{\nu})$, then

$$\begin{aligned} \mathbf{P}_e\mathbf{R}\mathbf{P}_e &= \mathbf{P}_e\mathbf{C}_{\boldsymbol{\Omega}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\mathbf{P}_e \\ &= \mathbf{P}_e\mathbf{C}_{\boldsymbol{\Omega}}\mathbf{P}_e\mathbf{P}_e\bar{\mathbf{B}}^{-\top}\mathbf{P}_e\mathbf{P}_e\underline{\mathbf{B}}\mathbf{P}_e\mathbf{P}_e\underline{\mathbf{B}}^\top\mathbf{P}_e\mathbf{P}_e\bar{\mathbf{B}}^{-1}\mathbf{P}_e \\ &= \mathbf{P}_e\mathbf{C}_{\boldsymbol{\Omega}}\mathbf{P}_e(\mathbf{P}_e\bar{\mathbf{B}}\mathbf{P}_e)^{-\top}\mathbf{P}_e\underline{\mathbf{B}}\mathbf{P}_e\mathbf{P}_e\underline{\mathbf{B}}^\top\mathbf{P}_e(\mathbf{P}_e\bar{\mathbf{B}}\mathbf{P}_e)^{-1}, \end{aligned}$$

which is the stochastic representation of their $F\mathcal{R}^{II}(\mathbf{P}_e\boldsymbol{\Omega}\mathbf{P}_e, \boldsymbol{\nu}, \mathbf{n})$ distributions, because $\mathbf{P}_e\mathbf{C}_{\boldsymbol{\Omega}}\mathbf{P}_e$ is the upper Cholesky factor of the reverse order $\boldsymbol{\Omega}$, $\mathbf{P}_e\boldsymbol{\Omega}\mathbf{P}_e$, $\mathbf{P}_e\bar{\mathbf{B}}\mathbf{P}_e$ is equal to $\underline{\mathbf{B}}$ but with degrees of freedom $\boldsymbol{\nu}$ instead of $\mathbf{n}$ and $\mathbf{P}_e\underline{\mathbf{B}}\mathbf{P}_e$ is equal to $\bar{\mathbf{B}}$ but with $\mathbf{n}$ instead of $\boldsymbol{\nu}$.

9. As the number of possible orderings explodes with increasing p, trying all possible orderings is infeasible.

10. They call them either type-I and type-II distributions but do not derive the reverse-order relationship between the two.

2.1 Probability Density Functions

To derive the probability density functions, we first need to introduce some special functions.

Definition 2.1 (Generalized Power Function)**.** *Let* $\mathbf{X}$ *be a real* $p \times p$ *matrix and let* $\mathbf{X}_{[i]}$ *denote the square submatrix created by taking the first* i *rows and columns of* $\mathbf{X}$*. Then the generalized power function (a.k.a. highest weight vector), denoted by* $|\mathbf{X}|_{\mathbf{n}}$ *is defined as*

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

The generalized power function is defined in e.g. Faraut and Korányi (1994). The determinant-with-subscript notation is borrowed from Blasques et al. (2021). It makes immediately visible the close relation of the generalized power function to the determinant, as it is easily seen that for $n_1 = n_2 = \dots = n_p = n$ we have $|\mathbf{X}|_{\mathbf{n}} = |\mathbf{X}|^n$. Next, we define the (multivariate) gamma function as in e.g. equations (5.2.1), (35.3.4) and (35.3.5) of the NIST Digital Library of Mathematical Functions.

Definition 2.2 (Multivariate Gamma Function)**.** *Let* $\mathbf{n}$ *be a real vector of length* p *and let* n *be a scalar. Then the vector-valued multivariate gamma function is defined as*

$$\Gamma_p(\mathbf{n}) = \pi^{p(p-1)/2}\prod_{i=1}^{p}\Gamma\left(n_i - \frac{i-1}{2}\right)$$

and the scalar-valued multivariate gamma function is defined as

$$\Gamma_p(n) = \pi^{p(p-1)/2}\prod_{i=1}^{p}\Gamma\left(n - \frac{i-1}{2}\right).$$

Obviously if $n_1 = n_2 = \dots = n_p = n$, then

$$\Gamma_p(\mathbf{n}) = \Gamma_p(n).$$

The next Lemma shows that in case of positive definite $\mathbf{X}$ the generalized power function can be written as function of the diagonal elements of the lower Cholesky

decomposition of $\mathbf{X}$. In that case $|\mathbf{X}|_{\mathbf{n}}$ is equal to the power weighted determinant defined in Blasques et al. (2021).

Lemma 2.1 (Power Weighted Determinant)**.** *Let* $\mathbf{\Sigma}$ *be positive definite and* $\mathbf{\Sigma} = \mathbf{TDT}^\top$ *be the unique decomposition into lower triangular square matrix with ones on the main diagonal,* $\mathbf{T}$ *and diagonal matrix with positive entries on the diagonal* $\mathbf{D}$*. Then we can rewrite*

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

Proof in Appendix.

Lemma 2.2. *Let the upper generalized multivariate gamma function,* $\bar{\Gamma}_U(\cdot)$*, be defined as in Blasques et al. (2021) and denote a vector with its elements in reverse order by a superscript left arrow, e.g.* $\overleftarrow{\mathbf{n}} = (n_p, n_{p-1}, \ldots, n_1)^\top$ *, then*

$$\Gamma_p\left(\overleftarrow{\mathbf{n}}\right) = \bar{\Gamma}_U(\mathbf{n}). \tag{11}$$

Proof. We have $(\overleftarrow{\mathbf{n}})_i = n_{p-i+1}$, such that

$$\begin{aligned}\Gamma_p\left(\overleftarrow{\mathbf{n}}\right) &= \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma\left(n_{p-i+1} - \frac{i-1}{2}\right) \\ &= \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma\left(n_i - \frac{p-1}{2}\right) = \bar{\Gamma}_U(\mathbf{n}).\end{aligned}$$

□

The next lemma corresponds to Lemma 3 in Blasques et al. (2021). It is used extensively in this paper, which is why we copy it here.[11]

Lemma 2.3 (Lemma 3, Blasques et al. 2021)**.** *Given a scalar* n*, a vector* $\mathbf{n}$ *of length* p*, a vector of ones* $\mathbf{1}$ *of length* p *and a positive definite matrix* $\mathbf{R}$*, the following identities hold.*

(i) If $\mathbf{n} = n \cdot \mathbf{1}$*, then* $|\mathbf{R}|_{n\cdot\mathbf{1}} = |\mathbf{R}|^n$

11. Part (v) below, where ${}_U|\mathbf{R}|_{\mathbf{n}}$ denotes their upper power weighted determinant, is only included to see the equivalence between our and their pdf representations.

(ii) *Let* $\mathbf{n}_1, \mathbf{n}_2$ *be two vectors of length* p*, then we have* $|\mathbf{R}|_{\mathbf{n}_1} \cdot |\mathbf{R}|_{\mathbf{n}_2} = |\mathbf{R}|_{\mathbf{n}_1+\mathbf{n}_2}$.

(iii) $(|\mathbf{R}|_{\mathbf{n}})^{-1} = |\mathbf{R}|_{-\mathbf{n}}$.

(iv) *If* $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^\top$*, where* $\mathbf{\Sigma}$ *is positive definite with lower Cholesky factor* $\mathbf{C}$*, then* $|\mathbf{R}|_{\mathbf{n}} \cdot |\mathbf{\Sigma}|_{-\mathbf{n}} = |\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}|_{\mathbf{n}}$*. As a special case we have* $|\mathbf{C}\mathrm{dg}(\mathbf{n})\mathbf{C}^\top|_{\boldsymbol{\nu}} = |\mathbf{\Sigma}|_{\boldsymbol{\nu}} \prod_{i=1}^{p} n_i^{\nu_i}$.

(v) ${}_U|\mathbf{R}|_{\mathbf{n}} = |\mathbf{R}^{-1}|_{-\mathbf{n}}$.

We list the probability density functions (pdfs) for all considered probability distributions in table 3. We derive the ones of the (Inverse) t-Riesz distribution and the Inverse F-Riesz distribution in Theorem 2.3, the ones of the (Inverse) Riesz and F-Riesz distributions are given in Theorems 4, 7 and 8 of Blasques et al. (2021).[12]

Theorem 2.3 (Probability Density Functions)**.** *The probability density functions of* $\mathbf{C}_{\mathbf{\Omega}}\mathcal{K}_{\mathcal{D}}\mathbf{C}_{\mathbf{\Omega}}$ *for* $\mathcal{D} \in (t\mathcal{R}, it\mathcal{R}, iF\mathcal{R})$ *obtain as*

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

$$p_{it\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, n, \boldsymbol{\nu}) = \frac{\Gamma\left((n + p\bar{\boldsymbol{\nu}})/2\right)}{\Gamma\left(n/2\right)\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} |\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} \left(1 + \mathrm{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}},$$

$$p_{iF\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, \mathbf{n}, \boldsymbol{\nu}) = \frac{\Gamma_p((\boldsymbol{\nu} + \mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p\left(\mathbf{n}/2\right)} |\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} \left|\left(\mathbf{I} + \mathbf{C}_{\mathbf{\Omega}}^\top \mathbf{R}^{-1}\mathbf{C}_{\mathbf{\Omega}}\right)^{-1}\right|_{\frac{\nu+\mathbf{n}}{2}}.$$

Proof in Appendix.

The pdfs of the Wishart-type distributions follow by simply setting all entries in the degree of freedom vectors equal. The pdfs of the standardized probability distributions, which we simply indicate by writing $p_{\mathcal{D}}(\mathbf{\Sigma}, \boldsymbol{\theta})$ instead of $p_{\mathcal{D}}(\mathbf{\Omega}, \boldsymbol{\theta})$, are given in the appendix in Table 11. They are easily derived by replace $\mathbf{\Omega}$ with $\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}^\top$ in Table 3.

12. To convert them to our notation consider Lemma 2.2.

Distribution	Probability Density Function, $p_{\mathcal{D}}(\mathbf{R}\vert\boldsymbol{\Omega},\boldsymbol{\theta})$
Wishart	$\frac{1}{2^{np/2}\Gamma_p(n/2)}\vert\boldsymbol{\Omega}\vert^{-\frac{n}{2}}\vert\mathbf{R}\vert^{\frac{n-p-1}{2}}\operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}^{-1}\mathbf{R}\right)$
Riesz	$\frac{1}{2^{p\bar{\mathbf{n}}/2}\Gamma_p(\mathbf{n}/2)}\vert\boldsymbol{\Omega}\vert_{-\frac{\mathbf{n}}{2}}\vert\mathbf{R}\vert_{\frac{\mathbf{n}-p-1}{2}}\operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}^{-1}\mathbf{R}\right)$
Inv.Wishart	$\frac{1}{2^{\nu p/2}\Gamma_p(\nu/2)}\vert\boldsymbol{\Omega}\vert^{\frac{\nu}{2}}\vert\mathbf{R}\vert^{-\frac{\nu+p+1}{2}}\operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}\mathbf{R}^{-1}\right)$
Inv.Riesz	$\frac{1}{2^{p\bar{\boldsymbol{\nu}}/2}\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}\vert\boldsymbol{\Omega}\vert_{\frac{\boldsymbol{\nu}}{2}}\vert\mathbf{R}\vert_{-\frac{\boldsymbol{\nu}+p+1}{2}}\operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Omega}\mathbf{R}^{-1}\right)$
t-Wishart	$\frac{\Gamma((\nu+pn)/2)}{\Gamma_p(n/2)\Gamma(\nu/2)}\vert\boldsymbol{\Omega}\vert^{-\frac{n}{2}}\vert\mathbf{R}\vert^{\frac{n-p-1}{2}}\left(1+\operatorname{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})\right)^{-\frac{\nu+pn}{2}}$
t-Riesz	$\frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}\vert\boldsymbol{\Omega}\vert_{-\frac{\mathbf{n}}{2}}\vert\mathbf{R}\vert_{\frac{\mathbf{n}-p-1}{2}}\left(1+\operatorname{tr}(\boldsymbol{\Omega}^{-1}\mathbf{R})\right)^{-\frac{\nu+pn}{2}}$
Inv.t-Wishart	$\frac{\Gamma((n+p\nu)/2)}{\Gamma_p(\nu/2)\Gamma(n/2)}\vert\boldsymbol{\Omega}\vert^{\frac{\nu}{2}}\vert\mathbf{R}\vert^{-\frac{\nu+p+1}{2}}\left(1+\operatorname{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-\frac{n+p\nu}{2}}$
Inv.t-Riesz	$\frac{\Gamma((n+p\bar{\boldsymbol{\nu}})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma(n/2)}\vert\boldsymbol{\Omega}\vert_{\frac{\boldsymbol{\nu}}{2}}\vert\mathbf{R}\vert_{-\frac{\boldsymbol{\nu}-p-1}{2}}\left(1+\operatorname{tr}(\boldsymbol{\Omega}\mathbf{R}^{-1})\right)^{-\frac{n+p\nu}{2}}$
F	$\frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)}\vert\boldsymbol{\Omega}\vert^{-\frac{n}{2}}\vert\mathbf{R}\vert^{\frac{n-p-1}{2}}\left\vert\mathbf{I}_p+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right\vert^{-\frac{\nu+n}{2}}$
F-Riesz	$\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\vert\boldsymbol{\Omega}\vert_{-\frac{\mathbf{n}}{2}}\vert\mathbf{R}\vert_{\frac{\mathbf{n}-p-1}{2}}\left\vert\mathbf{I}_p+\mathbf{C}_{\boldsymbol{\Omega}}^{-1}\mathbf{R}\mathbf{C}_{\boldsymbol{\Omega}}^{-\top}\right\vert_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$
Inv.F-Riesz	$\frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)}\vert\boldsymbol{\Omega}\vert_{-\frac{\mathbf{n}}{2}}\vert\mathbf{R}\vert_{\frac{\mathbf{n}-p-1}{2}}\left\vert\left(\mathbf{I}+\mathbf{C}_{\mathbf{R}}^{\top}\boldsymbol{\Omega}^{-1}\mathbf{C}_{\mathbf{R}}\right)^{-1}\right\vert_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}$
Inv.F-Riesz	$\frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)}\vert\boldsymbol{\Omega}\vert_{\frac{\boldsymbol{\nu}}{2}}\vert\mathbf{R}\vert_{-\frac{\boldsymbol{\nu}+p+1}{2}}\left\vert\left(\mathbf{I}+\mathbf{C}_{\boldsymbol{\Omega}}^{\top}\mathbf{R}^{-1}\mathbf{C}_{\boldsymbol{\Omega}}\right)^{-1}\right\vert_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}$

Table 3: Probability density functions of all considered distributions. To get to the Wishart-types from the Riesz-types use Definition 2.2 and Lemma 2.1 (i), for the (Inverse) Riesz and F-Riesz distributions use Theorems 4, 7 and 8 in Blasques et al. (2021) together with Lemmas 2.2 and 2.1, for the rest see Theorem 2.3.

3 The t-Riesz Distribution Family

This section considers the novel t-Riesz distribution family more closely. In particular, we show that it can be grounded in a distributional assumption on the intraday return vectors that generate a realized covariance matrix. Specifically, following theorems establish connections between an assumed distribution on the vector of all intraday returns $\tilde{\mathbf{r}} = (\mathbf{r}_1^\top, \mathbf{r}_2^\top, \ldots, \mathbf{r}_n^\top)^\top$ (with mean zero and dispersion matrix $\tilde{\mathbf{\Omega}} = \mathbf{I}_n \otimes \mathbf{\Omega}$) and the distribution of the RC as defined in equation (**??**).

Theorem 3.1. *Let liquidity be measured as the number of return observations over the trading hours of a trading day, and let there be a return observation on all intraday intervals n for at least one asset. Furthermore, for the j'th most liquid asset, let there be a return observation on a subset of the intervals where the $(j+1)$'th most liquid asset has an observation and let the assets be sorted from least to most liquid in $\mathbf{r}_j$. Now, denote by $\dot{\mathbf{r}}$ and $\dot{\mathbf{\Omega}}$ the sub-vector of $\tilde{\mathbf{r}}$ and sub-matrix of $\tilde{\mathbf{\Omega}}$ that are obtained by striking out the rows (and columns) with missing observations.*

Then, if $\dot{\mathbf{r}}$ follows a multivariate t-distribution $\dot{\mathbf{r}} \sim mvt(\mathbf{0}, \dot{\mathbf{\Omega}}, \nu)$, the realized covariance matrix $\mathbf{R} = \sum_{j=1}^{n} \mathbf{r}_j \mathbf{r}_j'$ follows a t-Riesz distribution with scale matrix $\mathbf{\Omega}$, degree of freedom vector $\mathbf{n} = (n_1, n_2, \ldots, n_p)$, where n_i denotes the number of return observations of asset i and degree of freedom parameter ν, denoted by

$$\mathbf{R} \sim t\mathcal{R}(\mathbf{\Omega}, \mathbf{n}, \nu).$$

Proof in Appendix.

Theorem 3.1 is a generalization (and slight reformulation) of a finding in Gribisch and Hartkopf (2022) (see also Hassairi, Ktari, and Zine 2022 and Veleva 2009). They show that the Riesz distribution can be generated by assuming a normal distribution on the intraday returns with heterogeneous liquidity. It is, however, common knowledge that financial return data strongly rejects the assumption of normality. Furthermore, the normality assumption also implicitly assumes that the intraday return vectors $\mathbf{r}_j$ are independent of each other, which is a very

strong assumption. A multivariate t-distribution constitutes a much more realistic assumption, as it accommodates the fat tails observed in financial return data, and although due to the block-diagonal structure of $\tilde{\mathbf{\Omega}}$ there is no correlation between the $\mathbf{r}_j$ it does imply dependence between them. These more realistic assumptions are mirrored in the superior performance of the t-Riesz compared to the Riesz distributions.

In case of no missing observations, we can make an even more general statement on the distribution of RC.

Theorem 3.2. *Let* $\tilde{\mathbf{r}}$ *follow a multivariate elliptically contoured distribution (as defined in Gupta, Varga, and Bodnar 2013 Definition 2.1) with zero mean vector, dispersion matrix* $\tilde{\mathbf{\Omega}}$ *and pdf*

$$f(\tilde{\mathbf{r}}) = |\mathbf{\Omega}|^{-n/2} h(\tilde{\mathbf{r}}^\top \tilde{\mathbf{\Omega}} \tilde{\mathbf{r}}). \tag{12}$$

Then the pdf of $\mathbf{R} = \sum_{j=1}^{n} \mathbf{r}_j \mathbf{r}_j'$ *obtains as*

$$\frac{\pi^{\frac{pn}{2}}}{\Gamma_p\left(\frac{n}{2}\right)} |\mathbf{R}|^{\frac{n-p-1}{2}} f(tr(\mathbf{R}\mathbf{\Omega}^{-1})). \tag{13}$$

Proof in Appendix.

Specifically, assuming $\tilde{\mathbf{r}} \sim mvt(\mathbf{0}, \tilde{\mathbf{\Omega}}, \nu)$, yields the t-Wishart distribution[13], which is equal to a t-Riesz distribution with all degrees of freedom parameters equal to each other.[14] Assuming a normal distribution on the intraday returns yields the Wishart distribution[15], but again the more realistic assumption of fat tails and inter-j dependence is mirrored in the superior performance of the t-Wishart distribution over the Wishart. We could also develop a test for the different Wishart-type distributions using Corollary 3.3.13.1 in Gupta and Nagar (2000).

13. A version of which was first introduced by Sutradhar and Ali (1989)

14. I also tried the hyperbolic-Wishart model, which nests the t-Wishart and Laplace-Wishart models. Interestingly, in both a low-dimensional and a high-dimensional estimation, the solver converges to almost exactly the t-Wishart. So I gave up the generalization attempts for the t-Wishart model.

15. Which is a Riesz distribution with all degree of freedom parameters equal to each other.

Next, we compare the properties of our newly generated distribution family to the recently proposed F-Riesz distributions of Blasques et al. (2021). They show a considerable improvement of the F-Riesz on the Riesz distribution proposed for RCs by Gribisch and Hartkopf (2022). Recall $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ for $\mathcal{D} \in (F\mathcal{R}, iF\mathcal{R})$,

$$\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1} \text{ and } \underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}$$

and for $\mathcal{D} \in (t\mathcal{R}, it\mathcal{R})$,

$$(\bar{b})^{-2}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top} \text{ and } (\underline{b})^{2}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1},$$

where $\underline{b}$ and $\bar{b}$ are χ_n and χ_ν distributed random variables, thus one-dimensional $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$, respectively. The scalar random variables $\underline{b}$ and $\bar{b}$ in the t-Riesz distribution family create a much larger dependence among the elements of the random matrices than in the F-Riesz distributions. Furthermore, given $\underline{\mathbf{B}}$ ($\bar{\mathbf{B}}$), a tail realization of $(\bar{b})^{-2}$ ($(\underline{b})^{2}$) yields a tail realization of the t-Riesz (inverse t-Riesz) distribution. This in stark contrast to the (inverse) F-Riesz which allows tail-heterogeneity via $\mathbf{n}$ and $\boldsymbol{\nu}$ (see Blasques et al. 2021). This contrast makes us call the t-Riesz distribution family "tail homogeneous". The natural conjecture would be that the t-Riesz distribution family might work better in market-wide crises. In contrast, the F-Riesz distribution might have advantages when one asset or subsections of the market are in distress.

Further advantages of our novel distributions over the F(-Riesz) distributions are that their evaluation is numerically more stable as their pdfs depend on the trace rather than the (power weighted) determinant and that they have slightly fewer parameters.

4 Time-Varying Mean

Lets add subscripts for the days in our sample, $t = 1, \ldots, T$, to represent the time series of RCs as $\mathbf{R}_t$. In the literature, it is standard to assume time variation in the mean, i.e., in the $\boldsymbol{\Sigma}$ matrix of the underlying distribution, while leaving the degree of freedom parameters fixed over time (see e.g. Golosnoy, Gribisch, and

Liesenfeld 2012, Opschoor et al. 2018, Blasques et al. 2021), that is

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

where $\mathcal{F}_{t-1} = \{\mathbf{R}_{t-1}, \mathbf{R}_{t-2}, \dots\}$. A standard updating mechanism for $\mathbf{\Sigma}_t$ is a scalar BEKK[16] recursion given by

$$\mathbf{\Sigma}_t = (1 - a - b)\mathbf{\Xi} + a\mathbf{R}_{t-1} + b\mathbf{\Sigma}_{t-1}, \tag{14}$$

where the intercept parameter matrix $\mathbf{\Xi}$ is symmetric positive definite of dimension $p \times p$ and a and b are scalar parameters, sometimes called ARCH and GARCH parameter, respectively. A necessary condition for stationarity is $a, b > 0 \wedge (a+b) < 1$ under which we have that the unconditional mean

$$\mathbb{E}\left[\mathbf{R}_t\right] = \mathbf{\Xi}.$$

In this paper, we assume stationarity. Note that, if we restrict $a = b = 0$ in equation (14), we end up with a static distribution, that is $\mathbf{\Sigma} = \mathbf{\Xi}$ and

$$\mathbf{R}_t \overset{iid}{\sim} \mathcal{D}(\mathbf{\Sigma}, \boldsymbol{\theta}). \tag{15}$$

5 Estimation

The total number of parameters is dominated by the order $\mathcal{O}(p^2)$ matrix $\mathbf{\Xi}$, which has $p(p+1)/2$ unique elements. This makes one-step numerical maximum likelihood estimation for, say, $p > 5$ very computationally expensive and, for, say, $p > 10$ infeasible. To alleviate this so-called curse of dimensionality we estimate $\mathbf{\Xi}$ (or $\mathbf{\Sigma}$ in the static model) with its from equation (15) obvious method-of-moments estimator

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

16. Named after Yoshi Baba, Robert Engle, Dennis Kraft, and Ken Kroner who wrote an earlier version of the paper Engle and Kroner (1995) in which the BEKK recursion is proposed.

and to estimate the remaining parameters ($\boldsymbol{\theta}$, a, and b) conditional on $\widehat{\boldsymbol{\Xi}}$ via standard numerical maximum likelihood estimation. This reduces to the size of the numerical optimization problem to the order of at most $\mathcal{O}(p)$, in case of the Wishart-type distributions even to at most 4[17] parameters. The consistency of this type of two-step estimation procedure is well established in the literature. In the empirical section of this paper, we always use this two-step estimation procedure.

A final complication is that, as mentioned above, the ordering of the assets matters for the Riesz-type distributions. We follow the algorithm proposed in Blasques et al. (2021) to optimize over the asset order.[18]

6 Empirical Analysis

6.1 Data

Our original data are one-minute close prices from all trading days from 1 January 1998 to 05 February 2021 for every stock that was a constituent of the S&P 500 index during the sample period. A close price is defined as the latest observed trade price of the respective one-minute interval and as such is we have previous tick interpolation on a fixed one-minute grid. We acquired the data from Quantquote[19], who combine, clean and process data directly obtained different exchanges, where the biggest are NYSE, NASDAQ and AMEX[20]. The data include observations from official trading hours as well as before- and after hour trading observations.

The aim is to produce the longest possible time series of accurately estimated daily integrated covariance estimators. We exclude dates before 1 January 2002, because the NYSE fully implemented decimal pricing in 2001[21] and there are numerous other trading irregularities during this year[22]. This leaves 4808 trading days. We then exclude stocks that have not been traded on one of the remaining

17. n, and/or ν, a and b.
18. The seed for the random generation of permutations to try initially is the same for all Riesz-type distributions.
19. The company is recommended by the Caltech Quantitative Finance Group, see http://quant.caltech.edu/historical-stock-data.html.
20. AMEX was bought by NYSE in 2008, and handled only 10% of trades at its height
21. On 29 January 2001 to be precise.
22. e.g. the days surrounding the terrorist attacks on 11 September 2001 and "computer systems connectivity problems" on 8 June 2001.

days in the sample, which leaves 465 of the 983 stocks in our sample. To be consistent across trading days we only keep observations from official trading hours.[23] We then choose the 100 stocks with the highest number of 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 this means that on average less then five close prices are missing per day. Note that since these missing observations are likely to be spread out over the trading day, this translates into much less than five artificial zero five minute returns on average per day. Thus the Epps effect (i.e. the fact that if previous tick interpolation is used and there are no trades in a time interval, an artificial zero-return is recorded), which leads to a downward bias of realized (co)variance(s) (Zhang (2011)), is no concern for us.

Excluding illiquid stocks is common practice in creating time series of integrated covariance estimators (see e.g. Lunde, Shephard, and Sheppard (2016)). While this procedure biases the sample towards stocks which where very liquid over the entire sample period[24] it does ensure that for those stocks included the integrated covariance estimates are accurate.

We follow Opschoor et al. (2018) and Blasques et al. (2021) and construct RCs of the 100 assets using five minute returns. In particular, we average for each trading day the five distinct RCs, obtained from constructing five minute log return vectors for each of the five distinct five minute grids over the trading day.[25] This estimator is known as the subsampling realized covariance estimator and has been introduced by Zhang, Mykland, and Aït-Sahalia (2005).[26] It has the advantage of being more efficient than the simple RC estimator, since it makes use of all the 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.[27]

23. Noureldin, Shephard, and Sheppard (2012) exclude the first and last 15 minutes of trading to control for overnight effects, which they model separately. Since I do not model them separately and I think that all available minutes should be used I do not follow that practice.
24. Relatively young firms (e.g. Facebook or Tesla) are excluded.
25. Starting (on a typical trading day) at 09:00, 09:01, 09:02, 09:03 and 09:04, respectively.
26. See also Sheppard (2012)
27. On a typical trading day we have for a given five minute grid 390/5=70 intraday return vectors, which would allow only for a maximum 70 assets if we require positive definite RCs. With subsampling, however RCs are based on sums of outer products of 385 five minute intraday return vectors, which implies that for up to 385 assets the resulting RCs are positive definite.

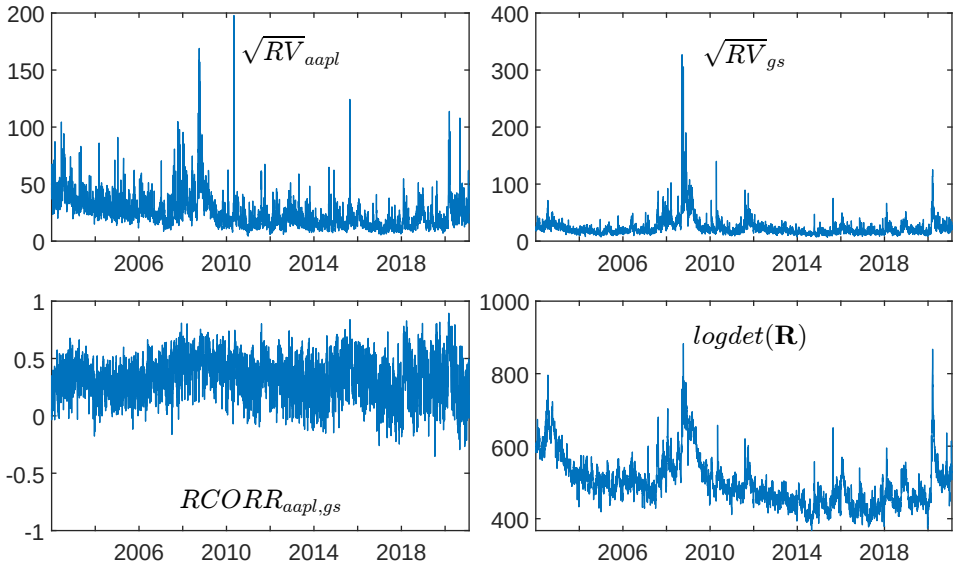

Figure 2: Top row: Annualized realized volatilities of Apple (aapl) and Goldman Sachs (gs), i.e. the square root of two of the elements on the main diagonal of the 100 assset $\mathbf{R}_t$ for the complete sample (1 January 2002 to 05 February 2021). Bottom row: Realized correlation between Apple and 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 asset $\mathbf{R}_t$ over time.

Finally, we randomly choose a 5, 10 and 25-dimensional principal submatrix, as well as the four principal submatrices corresponding to companies with SIC codes of the

1. Financial, Insurance and Real Estate,
2. Mining,
3. Trade and
4. High End Manufacturing division,

for a total of seven datasets.

For a view of the data see figure 2, which shows the annualized realized volatility for Cisco (csco) and Goldman Sachs (gs), as well as their realized correlation and the log-determinant of the 100-asset RCs. We see that the spikes in volatility are of similar magnitude in the recent COVID-19 induced market turmoil for both assets, while the global financial crises of 2008/2009 caused volatility to spike much higher for Goldman Sachs than for Apple. The dot-com crisis on the other

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	-67794	-84692	-168435	-527223	-537407	-381884
Riesz	-55534	-70180	-108027	-345020	-180751	-124508
Inv.Wishart	-50721	-53737	-111648	-255628	-162663	-78492
Inv.Riesz	-45490	-46495	-77633	-180138	36136	45223
t-Wishart	-24687	-30461	-38384	-25481	86847	233668
t-Riesz	-17543	-23263	-1941	49699	295409	355966
Inv.t-Wishart	-24761	-27528	-40690	-11924	191946	387973
Inv.t-Riesz	-20829	-19584	-8057	57397	376512	445333
F	-48758	-51941	-97299	-221813	-55002	14633
F-Riesz	-26402	-29242	-13809	-30051	307936	333553
Inv.F-Riesz	-29622	-29930	-22537	-70305	276602	308721

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

hand causes more volatility for Apple. We see that correlations are mainly positive and more stable around crises periods. Finally we see that the log-determinant of $\mathbf{R}_t$, as a measure of the size of the RCs, does indeed spike in the aforementioned market turmoil periods (dot-com, COVID, global financial crisis).

6.2 In-Sample

6.2.1 Static Distributions

As a first empirical exercise, we fit the different static distributions (i.e. $\mathbf{\Sigma} = \mathbf{\Xi}$) to the data. Table 4 shows the corresponding log-likelihood values. The distribution rankings are robust across the different cross-sectional dimensions p. We see a clear pattern in favor of the t-Riesz distribution family. This is especially clear in the industry-specific datasets (Mining, Finance and High-End Manufacturing). The F-Riesz and Inverse F-Riesz distributions are strong competitors, especially in the randomly drawn datasets. The Riesz and its special case, the Wishart, are the worst fitting static distributions.

Next, we look closely at the differences in fit between the tail-homogeneous Inverse t-Riesz and the tail-heterogeneous F-Riesz distribution. Figure 3 shows the difference in log-likelihood contributions between the two distributions depending

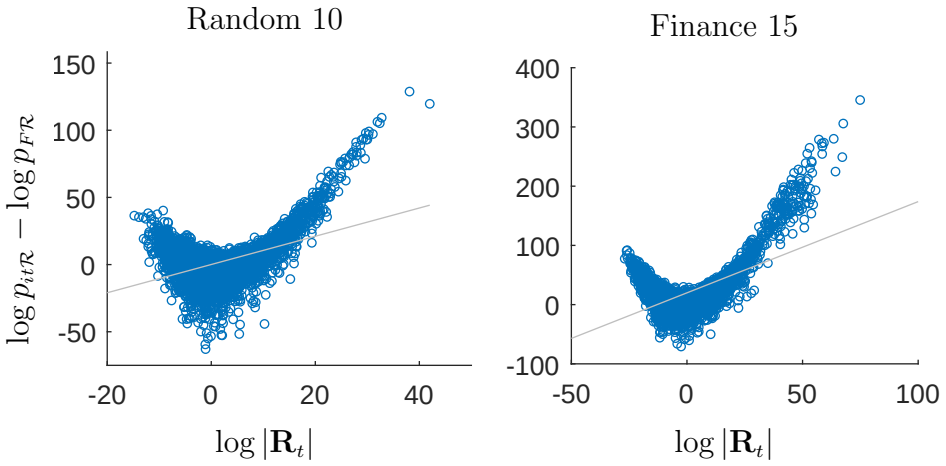

Figure 3: $\log p_{it\mathcal{R}}\left(\mathbf{R}_t|\widehat{\mathbf{\Sigma}},\hat{\theta}\right) - \log p_{F\mathcal{R}}\left(\mathbf{R}_t|\widehat{\mathbf{\Sigma}},\hat{\theta}\right)$

on the log determinant of the RCs for the random 10-dimensional and the Finance 15-dimensional dataset. We see, that the inverse t-Riesz distribution gains its advantage in static fit mainly from the "larger" RCs. This is in line with our expectation that tail heterogeneity is disadvantageous for crisis periods. It also fits better for very small RCs, which can be rationalized by the fact that in times of a very calm market the probability that one of the χ distributions on the main diagonal of $\underline{\mathbf{B}}$ lies in the tail (which is disadvantageous for the fit in market wide calm periods) is higher than only the probability that the $\underline{b}$ distribution lies in the tail.

6.2.2 Time-Varying Mean

Tables 5 and 6 contain the full sample estimated ARCH parameters ($\hat{a}$) and estimated persistence ($\hat{a} + \hat{b}$). They are all highly significant. We see several clear patterns. First, the estimated ARCH parameters become smaller with increasing cross-sectional dimension p for all distributions. This pattern has been documented by Pakel et al. (2021) to be estimation bias. It is larger the larger the dimension p and is caused by the method-of-moments estimator $\widehat{\mathbf{\Xi}}$, since it effectively estimates $\mathcal{O}(p^2)$ parameters with $\mathcal{O}(pT)$ data points. They show that composite likelihood estimation can mitigate the bias. Unfortunately, this estimation method is impossible for the Riesz-type distributions since their dof parameter vector(s) cannot be estimated. Another solution might be to use a shrinkage esti-

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	0.364	0.306	0.284	0.299	0.189	0.188
Riesz	0.339	0.286	0.259	0.275	0.161	0.160
Inv.Wishart	0.238	0.211	0.183	0.177	0.115	0.098
Inv.Riesz	0.242	0.206	0.181	0.168	0.108	0.094
t-Wishart	0.196	0.150	0.127	0.090	0.080	0.066
t-Riesz	0.186	0.132	0.117	0.072	0.070	0.052
Inv.t-Wishart	0.153	0.127	0.101	0.074	0.065	0.053
Inv.t-Riesz	0.154	0.122	0.097	0.067	0.059	0.050
F	0.257	0.231	0.198	0.192	0.126	0.110
F-Riesz	0.200	0.166	0.145	0.136	0.091	0.078
Inv.F-Riesz	0.215	0.179	0.156	0.147	0.095	0.084

Table 5: Estimated ARCH parameter $\hat{a}$.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	0.976	0.981	0.982	0.991	0.988	0.981
Riesz	0.978	0.983	0.983	0.991	0.991	0.984
Inv.Wishart	0.989	0.989	0.992	0.996	0.994	0.995
Inv.Riesz	0.990	0.991	0.993	0.995	0.994	0.996
t-Wishart	0.988	0.993	0.995	0.999	0.997	0.994
t-Riesz	0.987	0.994	0.997	0.999	0.998	0.996
Inv.t-Wishart	0.992	0.994	0.997	0.999	0.997	0.996
Inv.t-Riesz	0.992	0.994	0.997	0.998	0.997	0.997
F	0.989	0.989	0.992	0.996	0.995	0.995
F-Riesz	0.990	0.993	0.993	0.997	0.996	0.997
Inv.F-Riesz	0.990	0.993	0.993	0.997	0.996	0.997

Table 6: Estimated persistence parameter $\hat{a} + \hat{b}$.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	-15599	-11314	-1397	55874	333549	330537
Riesz	-11866	-6289	13107	88881	423237	423510
Inv.Wishart	-7668	1693	32727	132311	541567	547719
Inv.Riesz	-5813	4154	39879	148338	574711	583946
t-Wishart	-5298	5910	33588	149496	469265	467948
t-Riesz	-2650	8572	44108	178339	539295	541439
Inv.t-Wishart	-1422	11277	55196	195532	639630	655575
Inv.t-Riesz	440	13286	62134	213268	666312	684209
F	-6908	2468	35627	142413	571495	572881
F-Riesz	721	12621	62436	190194	669628	674391
Inv.F-Riesz	-122	11619	59471	185342	662615	665143

Table 7: Log-likelihood values for the estimated dynamic distributions and different datasets. The background shades are to be read column-wise, with the lowest log-likelihood value shaded black and the highest one being shaded white, with linear scaling in between.

mator as in Engle, Ledoit, and Wolf (2019). In this paper, however, we focus on differences between assumed probability distributions for RCs. We do not expect the relative ranking results to change if we use one of the abovementioned methods to estimate the intercept matrix $\mathbf{\Xi}$.

Second, the estimated persistence $\hat{a} + \hat{b}$ is stable across distributions and dimensions p and has typical magnitude for volatility time series. Third, the estimated a are smallest for the t-Riesz distribution family across all dimensions, followed by the F-Riesz distributions and largest for the Riesz distributions and the F distribution. That is, the t-Riesz distribution family reacts least to the previous realizations $\mathbf{R}_{t-1}$ to update the mean $\mathbf{\Sigma}_t$, which we interpret as an indication of the excellent (unconditional) fit of these distributions. In contrast, the Wishart distribution reacts most to the previous $\mathbf{R}_{t-1}$, indicating a worse fit of the distributional assumption. We note here that in terms of fit and forecasting performance, a large mean-shifting reaction to previous RCs is actually beneficial in crisis periods, where RCs suddenly spike in size and stay large for a short time. A good overall distributional fit of other distributions causes them to react more slowly to those volatility bursts.

Table 7 contains the log-likelihood values for the estimated distributions with time-varying mean. As for the static distributions, the ranking across distri-

butions is relatively stable over the cross-sectional dimension p. However, it is steeper across distributions, with the Inverse t-Riesz and the F-Riesz distributions emerging as the clear winners. In particular, the Inverse t-Riesz and F-Riesz distributions exhibit the largest likelihood values. They are very close, with the former winning the three industry-specific datasets and the latter winning the three random datasets. This is again in line with our economic intuition that tail homogeneity is advantageous for RCs of homogeneous assets. The Riesz (as proposed by Gribisch and Hartkopf 2022) and its special case, the Wishart distribution, are unambiguously the worst fitting distributions. In general, Inverse distributions fit better than non-inverted ones. This is not surprising as fitting Inverse distributions to $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_T\}$ is equivalent to fitting non inverted ones to $\{\mathbf{R}_1^{-1}, \mathbf{R}_2^{-1}, \ldots, \mathbf{R}_T^{-1}\}$. The inverted RCs, also known as precision or concentration matrices, exhibit much smaller tails, hence the good fit of non-inverted distributions. Through this reasoning, we can call all Inverse distributions fat-tailed distributions. Furthermore, every Wishart-type distribution has, by construction, a lower estimated likelihood value than its Riesz-type counterpart. However, it is noteworthy that the difference in likelihood values is particularly large between the (Inverse) F-Riesz distribution and the F distribution. Observe that the F ranks between the Inverse Wishart and Inverse Riesz distribution for all datasets. This indicates that from the already fat-tailed Inverse Wishart distribution, adding asset heterogeneity via different dof parameters is more advantageous than adding fat-tailedness by adding an inverse Barlett matrix to the stochastic representation. Figure 4 compares the log-likelihood contributions of the Inverse t-Riesz and F-Riesz distributions similar to Figure 3 but for the time-varying mean specification. As in the static case, higher log-likelihood contributions for the tail homogeneous Inverse t-Riesz distributions are associated with larger RCs, in line with our economic intuition that more volatile trading days exhibit more dependence among financial assets.

6.3 Out-of-Sample Forecasting Performance

For out-of-sample comparisons between different assumed probability distributions, a natural loss function is the log score, also known as the log posterior

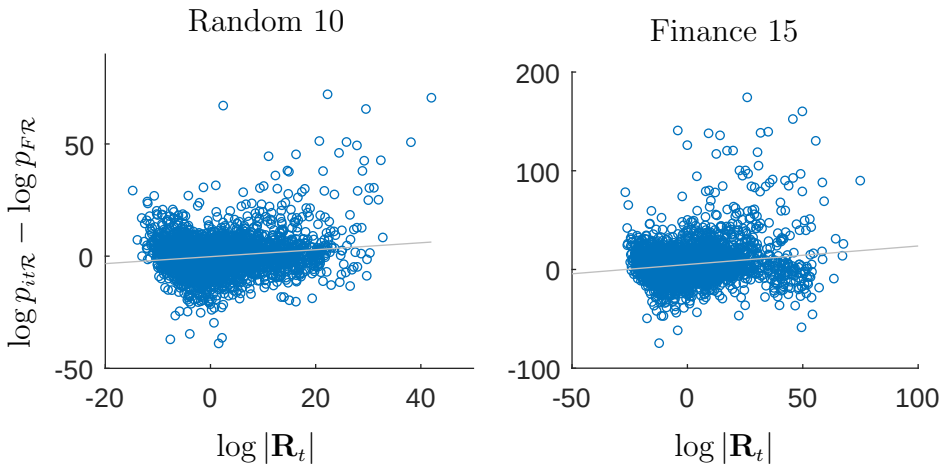

Figure 4: $\log p_{it\mathcal{R}}\left(\mathbf{R}_t|\widehat{\mathbf{\Sigma}}_t, \hat{\theta}\right) - \log p_{F\mathcal{R}}\left(\mathbf{R}_t|\widehat{\mathbf{\Sigma}}_t, \hat{\theta}\right)$

predictive likelihood, since it indicates how much probability mass the predictive distribution assigns to the observed outcome (compare e.g. Hautsch and Voigt 2019 and Blasques et al. 2021). The log score can also be justified as the consistent choice of loss function for maximum likelihood estimation in the following sense. It evaluates the out-of-sample data with the same loss function used to estimate the models in-sample. This is in line with Hansen and Dumitrescu (2022) who show that coherency between the estimation criterion and the actual objective is essential.

Since we are interested in overall distribution fit, it is important to not only the $t+1$ forecasting performance of the different distributions. To this end, Table 8 contains the log score losses over a one month forecasting period (22 trading days) for the entire forecasting window. The 90% model confidence sets (MCS, see Hansen, Lunde, and Nason 2011) are shaded in gray. We see that the Inverse t-Riesz distribution emerges as the clear winner in this comparison as it is the only member of the MCS for all datasets. Across datasets, the likelihood functions of the entire t-Riesz distribution family fit very well out-of-sample, slightly better than the F-Riesz, except for the 25-dimensional datasets. If we take a volatile period (2007 - 2011) and a calm period (2012 - 2019) forecasting window (see Tables 12 and 13 in the appendix), the Inverse t-Riesz distribution remains the sole member of the MCS except for the random 25-dimensional dataset in the calm period, where also the F-Riesz distribution is in the MCS. The above observations

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	319	616	541	1565	1870	900
Riesz	282	559	413	1061	1066	103
Inv.Wishart	224	425	174	-32	-101	-999
Inv.Riesz	214	404	126	-226	-423	-1271
t-Wishart	153	327	15	-685	-274	-972
t-Riesz	139	307	-44	-869	-662	-1432
Inv.t-Wishart	135	304	-94	-912	-1137	-1978
Inv.t-Riesz	128	292	-128	-1026	-1302	-2162
F	210	414	136	-230	-396	-1172
F-Riesz	147	319	-89	-715	-1173	-1959
Inv.F-Riesz	155	332	-59	-651	-1112	-1876

Table 8: Log score loss over a one month forecasting period (22 trading days), $-\sum_{j=1}^{22} p_{\mathcal{D}}(\mathbf{R}_{t+j}|\widehat{\boldsymbol{\Sigma}}_{j+1}, \hat{\boldsymbol{\theta}}_{j+1})$, for the entire forecasting window, where each model is re-estimated every 10 trading days. 90% model confidence sets in red.

confirm our intuition that tail homogeneity is a more realistic assumption than tail heterogeneity, indicating that the F-Riesz distribution might be overfitting the data.

In full disclosure, we also report the one-step-ahead log score loss results for the entire sample. Here the Inverse t-Riesz distribution is the sole member of the MCS

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	7.22	15.97	3.51	-18.61	-0.79	-29.75
Riesz	6.53	15.08	0.60	-25.22	-19.40	-49.70
Inv.Wishart	5.35	12.94	-4.84	-35.97	-49.62	-83.52
Inv.Riesz	5.04	12.52	-6.14	-39.49	-56.69	-91.34
t-Wishart	4.85	12.23	-4.79	-40.61	-32.63	-61.30
t-Riesz	4.38	11.67	-6.95	-46.42	-46.68	-76.80
Inv.t-Wishart	4.01	11.02	-9.59	-50.55	-70.32	-104.08
Inv.t-Riesz	3.70	10.67	-10.86	-53.90	-75.18	-110.57
F	5.18	12.85	-5.44	-38.59	-56.51	-87.99
F-Riesz	3.75	10.85	-10.81	-48.56	-76.43	-109.33
Inv.F-Riesz	3.93	11.09	-10.16	-47.36	-74.87	-107.25

Table 9: Log score loss, $-p_{\mathcal{D}}(\mathbf{R}_{t+1}|\widehat{\boldsymbol{\Sigma}}_{t+1}, \hat{\boldsymbol{\theta}}_{t+1})$, for the entire forecasting window, where each model is re-estimated every 10 trading days. 90% model confidence sets in red.

for the industry-specific Mining and Finance datasets, while for the other datasets, the F-Riesz distribution is also in the MCS. So even if we are only interested in one-step-ahead distributional predictions, the Inverse t-Riesz distribution is the best and the only choice in the case of homogeneous assets.

7 Conclusion

In this paper, we derived all hitherto used probability distributions for realized covariance measures and proposed several new ones. We showed that all distributions can be based on lower and/or upper triangular random Barlett matrices, and we theoretically showed how they relate to each other. Furthermore, we derived their probability density functions, expectations and standardized versions and investigated other peculiarities. Empirically we showed that the newly proposed t-Riesz distribution family generally provides the best fit for the data of RCs. We theoretically showed that this new distribution family can be derived from a more realistic (than the Riesz distribution) low-level assumption on the high-frequency return vectors from which realized covariance measures are constructed. Furthermore, we rationalized that Inverse distributions fit better than non-inverted ones and that the data support the Riesz generalizations of Wishart-type distributions.

References

Journal Articles

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.

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

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

Bai, Jushan, and Zhihong Chen. 2008. "Testing Multivariate Distributions in GARCH Models." *Journal of Econometrics* 143 (1): 19–36.

Bellman, Richard. 1956. "A Generalization of some Integral Identities due to Ingham and Siegel." *Duke Mathematical Journal* 23 (4): 571–577.

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

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

———. 2014. "On Riesz Distribution." *Metrika* 77 (4): 469–481.

Engle, Robert F., and Kenneth F. Kroner. 1995. "Multivariate Simultaneous Generalized ARCH." *Econometric Theory* 11 (1): 122–150.

Engle, Robert F., Olivier Ledoit, and Michael Wolf. 2019. "Large Dynamic Covariance Matrices." *Journal of Business & Economic Statistics* 37 (2): 363–375. eprint: https://doi.org/10.1080/07350015.2017.1345683.

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.

Gourieroux, Christian, Joann Jasiak, and Razvan Sufana. 2009. "The Wishart Autoregressive Process of Multivariate Stochastic Volatility." *Journal of Econometrics* 150 (2): 167–181.

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.*

Gribisch, Bastian, and Michael Stollenwerk. 2020. "Dynamic Principal Component CAW Models for High-Dimensional Realized Covariance Matrices." *Quantitative Finance* 20 (5): 799–821.

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

Hansen, Peter Reinhard, and Elena-Ivona Dumitrescu. 2022. "How should parameter estimation be tailored to the objective?" *Journal of Econometrics* 230 (2): 535–558.

Hassairi, Abdelhamid, Fatma Ktari, and Raoudha Zine. 2022. "On the Gaussian Representation of the Riesz Probability Distribution on Symmetric Matrices." *AStA Advances in Statistical Analysis.*

Hautsch, Nikolaus, and Stefan Voigt. 2019. "Large-Scale Portfolio Allocation under Transaction Costs and Model Uncertainty." Big Data in Dynamic Predictive Econometric Modeling, *Journal of Econometrics* 212 (1): 221–240.

Louati, Mahdi, and Afif Masmoudi. 2015. "Moment for the inverse Riesz distributions." *Statistics & Probability Letters* 102:30–37.

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.

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

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

Olkin, Ingram. 1959. "A Class of Integral Identities with Matrix Argument." *Duke Mathematical Journal* 26 (2): 207–213.

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.

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.

Sutradhar, Brajendra C., and Mir M. Ali. 1989. "A Generalization of the Wishart Distribution for the Elliptical Model and its Moments for the Multivariate t Model." *Journal of Multivariate Analysis* 29 (1): 155–162.

Veleva, Evelina. 2009. "Testing a Normal Covariance Matrix for Small Samples with Monotone Missing Data." *Applied Mathematical Sciences* 3 (54): 2671–2679.

Yu, Philip L. H., Wai K. Li, and F. C. Ng. 2017. "The Generalized Conditional Autoregressive Wishart Model for Multivariate Realized Volatility." *Journal of Business and Economic Statistics* 35 (4): 513–527.

Zhang, Lan. 2011. "Estimating covariation: Epps effect, microstructure noise." Realized Volatility, *Journal of Econometrics* 160 (1): 33–47.

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.

Books

Faraut, Jacques, and Adam Korányi. 1994. *Analysis on Symmetric Cones.* Oxford University Press.

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

Gupta, Arjun K., Tamas Varga, and Taras Bodnar. 2013. *Elliptically Contoured Models in Statistics and Portfolio Theory.*

Maaß, Hans. 1971. *Siegel's Modular Forms and Dirichlet Series.* 1st ed. Springer.

Mittelhammer, Ron C. 2013. *Mathematical Statistics for Economics and Business.* Springer.

Walck, Christian. 2007. *Hand-book on statistical distributions for experimentalists.* University of Stockholm.

Book Chapters

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.

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.

Working Papers

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.

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.

Other

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.

8 Appendix

8.1 Proofs

8.1.1 Proof of Theorem 2.1

Proof.

Proof of $\mathbb{E}\left[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right]$: This result has been proven in Díaz-García (2013). However, our proof is more straightforward as it directly uses the stochastic representations in terms of the Barlett matrices. We have

$$\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right)_{ij} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}\left(\underline{\mathbf{B}}^\top\right)_{kj} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}. \tag{16}$$

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

$$\mathbb{E}\left[\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right)_{ij}\right] = \sum_{k=1}^{p} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}\right] = \sum_{k=1}^{p} \mathbb{E}\left[\underline{\mathbf{B}}_{ik}\right]\mathbb{E}\left[\underline{\mathbf{B}}_{jk}\right] = 0, \tag{17}$$

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

$$\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right)_{ii} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}^2 = \sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^2, \tag{18}$$

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

$$\sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^2 \sim \chi^2_{n_i} \tag{19}$$

with expectation n_i. Thus

$$\mathbb{E}\left[\left(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\right)_{ii}\right] = n_i. \tag{20}$$

Proof of $\mathbb{E}\left[\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\right)^{-1}\right]$: See Louati and Masmoudi (2015).

Proof of $\mathbb{E}\left[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}\right]$: See Theorem 10 in Blasques et al. (2021).

Proof of $\mathbb{E}\left[\underline{\mathbf{B}}\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^{\top}\right)^{-1}\underline{\mathbf{B}}^{\top}\right]$: Due to independence we have

$$\mathbb{E}\left[\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\right] = \mathbb{E}\left[\underline{\mathbf{B}}\text{dg}(\overset{i\mathbb{R}''}{\mathbf{m}})\underline{\mathbf{B}}^{\top}\right],$$

where $\overset{i\mathbb{R}''}{\mathbf{m}}$ is given in (**??**). Denote

$$\mathbf{T} = \underline{\mathbf{B}}\left(\overset{i\mathbb{R}''}{\mathbf{m}}\right)^{1/2},$$

with elements $\mathbf{T}_{ij} = \underline{\mathbf{B}}_{ij}\sqrt{\overset{i\mathbb{R}''}{m}_j}$. The (i,j) element of $\mathbf{R} = \mathbf{T}\mathbf{T}^{\top}$ is

$$\mathbf{R}_{i,j} = \sum_{k=1}^{p}\mathbf{T}_{ik}\left(\mathbf{T}\right)_{kj}^{\top} = \sum_{k=1}^{p}\mathbf{T}_{ik}\mathbf{T}_{jk} = \sum_{k=1}^{p}\overset{i\mathbb{R}''}{m}_k\,\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk},$$

which for $i \neq j$ we have

$$\mathbb{E}\left[\mathbf{R}_{i,j}\right] = \sum_{k=1}^{p}\overset{i\mathbb{R}''}{m}_k\,\mathbb{E}\left[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}\right] = \sum_{k=1}^{p}\overset{i\mathbb{R}''}{m}_k\,\mathbb{E}\left[\underline{\mathbf{B}}_{ik}\right]\mathbb{E}\left[\underline{\mathbf{B}}_{jk}\right] = 0,$$

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

$$\mathbb{E}\left[\mathbf{R}_{i,i}\right] = \sum_{k=1}^{p}\overset{i\mathbb{R}''}{m}_k\,\mathbb{E}\left[\underline{\mathbf{B}}_{ik}^{2}\right] = \sum_{k\leq i}\overset{i\mathbb{R}''}{m}_k\,\mathbb{E}\left[\underline{\mathbf{B}}_{ik}^{2}\right],$$

with

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

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

$$\begin{aligned}
\mathbb{E}[\mathbf{R}_{1,1}] &= (n_1 - 1 + 1)\ \overset{i\!R^{ii}}{m}_1, \\
\mathbb{E}[\mathbf{R}_{2,2}] &= \overset{i\!R^{ii}}{m}_1 + (n_2 - 2 + 1)\ \overset{i\!R^{ii}}{m}_2, \\
\mathbb{E}[\mathbf{R}_{2,2}] &= \overset{i\!R^{ii}}{m}_1 + \overset{i\!R^{ii}}{m}_2 + (n_3 - 3 + 1)\ \overset{i\!R^{ii}}{m}_3, \\
&\vdots
\end{aligned}$$

or

$$\overset{i\!R^{ii}}{m}_i = \sum_{j=1}^{i-1} \overset{i\!R^{ii}}{m}_j + (n_i - i + 1)\ \overset{i\!R^{ii}}{m}_i$$

or more precisely

$$\overset{i\!R^{ii}}{m}_i = \begin{cases} n_1\ \overset{i\!R^{ii}}{m}_1, & \text{for } i = 1 \\ \sum_{j=1}^{i-1} \overset{i\!R^{ii}}{m}_j + (n_i + i - 1)\ \overset{i\!R^{ii}}{m}_i, & \text{for } i > 1, \end{cases} \tag{21}$$

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

8.1.2 Proof of Lemma 2.1

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

$$\mathbf{\Sigma} = \mathbf{T}\mathbf{D}\mathbf{T}^\top = \mathbf{C}\mathbf{C}^\top,$$

then

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

So

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

and thus

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

Finally

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

□

8.1.3 Proof of Theorem 2.3

Proof. All proofs start from the stochastic representations given in Table 1. The two integrals in the following Lemma are important for the derivation of the pdfs of the Riesz-type[28] distributions.

Lemma 8.1. *For* $\mathbf{n}$ *with* $n_i > i - 1$ *we have,*

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|^{\frac{\mathbf{n}-p-1}{2}} \text{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = 2^{p\bar{n}/2} \Gamma_p\left(\frac{\mathbf{n}}{2}\right) |\mathbf{B}^{-1}|^{\frac{\mathbf{n}}{2}} \tag{22}$$

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

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}^{-1}|^{\frac{\mathbf{n}+p+1}{2}} \text{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = \frac{1}{2^{p\bar{n}/2}} \Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right) |\mathbf{B}|^{\frac{\mathbf{n}}{2}}. \tag{23}$$

Proof. The proofs can be found in Faraut and Korányi (1994) paper VII.[29]. Throughout, according to their table on p. 97, for the cone of symmetric positive definite matrices we have the dimension $n = p(p+1)/2$, the rank $r = p$ and $d = 1$.[30] Furthermore, throughout their book they use the Euclidean measure on a Euclidean space, which translated into our notation is $dx = \prod_{i=1}^{p} a_{ii} 2^{p(p-1)/4} \prod_{i<j} a_{ij} = 2^{p(p-1)/4} d\mathbf{A}$ and leads to a slightly different multivariate gamma function.[31] In

28. See paper **??**.
29. Further references are Díaz-García (2014), Maaß (1971) p. 76, Gupta and Nagar (2000), Theorem 1.4.7, which is based on Olkin (1959), which in turn is based on the generalized Ingham formula in Bellman (1956).
30. For the notation see their Example 2 on p. 8 and p. 9.
31. I thank Jacques Faraut for pointing this out to me.

particular from their Theorem VII.1.1.

$$\Gamma_\Omega(\mathbf{n}) = 2^{p(p-1)/4}\Gamma_p(\mathbf{n}), \tag{24}$$

with $\Gamma_p(\mathbf{n})$ as in Definition 2.2. Their Proposition VII.1.2., with $x = \mathbf{A}$, $y = \frac{1}{2}\mathbf{B}$ and $\mathbf{s} = \frac{\mathbf{n}}{2}$ translates to

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) 2^{p(p-1)/4} d\mathbf{A} = 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right) |2\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}} \tag{25}$$

$$= 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right) 2^{p\bar{n}/2} |\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}}. \tag{26}$$

Their last equation on page 129 together with Proposition VII.1.5 (ii) and $x = \mathbf{A}$, $y = \frac{1}{2}\mathbf{B}$ and $\mathbf{s} = \frac{\mathbf{n}}{2}$ translates to

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}^{-1}|_{\frac{\mathbf{n}+p+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) 2^{p(p-1)/4} d\mathbf{A} = 2^{p(p-1)/4}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right) \left|\frac{1}{2}\mathbf{B}\right|_{\frac{\mathbf{n}}{2}} \tag{27}$$

$$= 2^{p(p-1)/4}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right) \frac{1}{2^{p\bar{n}/2}} |\mathbf{B}|_{\frac{\mathbf{n}}{2}}. \tag{28}$$

□

<u>t-Riesz distribution:</u> The stochastic representation is $\mathbf{R} = \mathbf{C}_{\mathbf{\Omega}}(\bar{b})^{-2}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathbf{C}_{\mathbf{\Omega}}^\top$, which can be written as $\mathbf{R} = w^{-1}\mathbf{A}$, with $\mathbf{A} \sim \mathcal{R}(\mathbf{\Omega}, \mathbf{n})$ independent of $w \sim \chi^2_\nu$. The joint pdf of w and $\mathbf{A}$ is given by

$$\frac{1}{\Gamma(\nu/2)\,2^{\nu/2}} w^{\frac{\nu}{2}-1} \exp\left(-\frac{w}{2}\right) \frac{|\mathbf{A}|_{\frac{\mathbf{n}-p-1}{2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{A}\right)\right)}{|\mathbf{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_p(\mathbf{n}/2)\,2^{p\bar{\mathbf{n}}/2}}.$$

Transforming $\mathbf{R} = w^{-1}\mathbf{A}$, with Jacobian $J(w, \mathbf{A} \to w, \mathbf{R}) = w^{p(p+1)/2}$ (see e.g. Gupta and Nagar 2000, equation 1.3.5.), we get the joint density of w and $\mathbf{R}$ as

$$\frac{1}{\Gamma(\nu/2)\,2^{\nu/2}} w^{\frac{\nu}{2}-1} \exp\left(-\frac{w}{2}\right) \frac{|w\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} \exp\left(-\frac{w}{2}\operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)}{|\mathbf{\Omega}|_{\frac{\mathbf{n}}{2}}\Gamma_p(\mathbf{n}/2)\,2^{p\bar{\mathbf{n}}/2}} w^{\frac{p(p+1)}{2}}$$

$$= \frac{|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} |\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}}{\Gamma(\nu/2)\,\Gamma_p(\mathbf{n}/2)\,2^{(\nu+p\bar{\mathbf{n}})/2}} w^{\frac{\nu+p\bar{\mathbf{n}}}{2}-1} \exp\left(-\frac{w}{2}\left(1 + \operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)\right),$$

where using Lemma **??**,

$$|w\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} = |\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} \prod_{i=1}^{p} w^{\frac{n_i-p-1}{2}} = |\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} w^{p\frac{\bar{\mathbf{n}}-(p+1)}{2}}.$$

Now integrating out w we get the probability density function of $\mathbf{R}$ as

$$\begin{aligned} p_{t\mathcal{R}}(\mathbf{R}|\mathbf{\Omega},\mathbf{n},\nu) &= \frac{|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}}{\Gamma(\nu/2)\,\Gamma_p(\mathbf{n}/2)\,2^{(\nu+p\bar{\mathbf{n}})/2}} \\ &\quad \times \int_0^\infty w^{\frac{\nu+p\bar{\mathbf{n}}}{2}-1} \exp\left(-\frac{w}{2}\left(1+\operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)\right) dw \\ &= \frac{|\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}}}{\Gamma(\nu/2)\,\Gamma_p(\mathbf{n}/2)\,2^{(\nu+p\bar{\mathbf{n}})/2}} \Gamma\left((\nu+p\bar{\mathbf{n}})/2\right) \left[\frac{1}{2}\left(1+\operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)\right]^{-(\nu+p\bar{\mathbf{n}})/2} \\ &= \frac{\Gamma\left((\nu+p\bar{\mathbf{n}})/2\right)}{\Gamma(\nu/2)\Gamma_p(\mathbf{n}/2)} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}}|\mathbf{R}|_{\frac{\mathbf{n}-p-1}{2}} \left(1+\operatorname{tr}\left(\mathbf{\Omega}^{-1}\mathbf{R}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}, \end{aligned}$$

where we used equation (5.9.1) of the NIST Digital Library of Mathematical Functions.

Inverse t-Riesz distribution: We have $\mathbf{R} = \mathbf{C}_{\mathbf{\Omega}}(\underline{b})^2 \left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\right)^{-1} \mathbf{C}_{\mathbf{\Omega}}^\top$, which can be written as $\mathbf{R} = w\mathbf{A}$, with $\mathbf{A} \sim i\mathcal{R}(\mathbf{\Omega},\boldsymbol{\nu})$ independent of $w \sim \chi^2_n$. The joint pdf of w and $\mathbf{A}$ is given by

$$\frac{1}{\Gamma(n/2)\,2^{n/2}} w^{\frac{n}{2}-1} \exp\left(-\frac{w}{2}\right) \frac{|\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{A}|_{-\frac{\nu+p+1}{2}}}{\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)2^{p\bar{\boldsymbol{\nu}}/2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(\mathbf{\Omega}\mathbf{A}^{-1}\right)\right).$$

Transforming $\mathbf{R} = w\mathbf{A}$, with Jacobian $J(w,\mathbf{A}\to w,\mathbf{R}) = w^{-p(p+1)/2}$ (see e.g. Gupta and Nagar 2000, equation 1.3.5.), we get the joint density of w and $\mathbf{R}$ as

$$\begin{aligned} &\frac{w^{n/2-1}}{\Gamma(n/2)\,2^{n/2}} \exp\left(-\frac{w}{2}\right) \frac{|\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}|w^{-1}\mathbf{R}|_{-\frac{\nu+p+1}{2}}}{\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)2^{p\bar{\boldsymbol{\nu}}/2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(w\mathbf{\Omega}\mathbf{R}^{-1}\right)\right) w^{-\frac{p(p+1)}{2}} \\ =&\frac{|\mathbf{\Omega}|_{\frac{\boldsymbol{\nu}}{2}}|\mathbf{R}|_{-\frac{\nu+p+1}{2}}}{\Gamma(n/2)\,\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)2^{(n+p\bar{\boldsymbol{\nu}})/2}} w^{\frac{n+p\bar{\boldsymbol{\nu}}}{2}-1} \exp\left(-\frac{w}{2}\left(1+\operatorname{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)\right)\right), \end{aligned}$$

where using Lemma **??**,

$$|w\mathbf{R}|_{-\frac{\nu+p+1}{2}} = |\mathbf{R}|_{-\frac{\nu+p+1}{2}} \prod_{i=1}^{p} w^{\frac{\nu_i+p+1}{2}} = |\mathbf{R}|_{-\frac{\nu+p+1}{2}} w^{p\frac{\bar{\boldsymbol{\nu}}+(p+1)}{2}}.$$

Now integrating out w we get the probability density function of $\mathbf{R}$ as

$$\begin{aligned}
p_{it\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, n, \boldsymbol{\nu}) &= \frac{|\mathbf{\Omega}|_{\frac{\nu}{2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}}}{\Gamma(n/2)\,\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right) 2^{(n+p\bar{\boldsymbol{\nu}})/2}} \\
&\quad \times \int_0^\infty w^{\frac{n+p\bar{\boldsymbol{\nu}}}{2}-1} \exp\left(-\frac{w}{2}\left(1+\operatorname{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)\right)\right) dw \\
&= \frac{|\mathbf{\Omega}|_{\frac{\nu}{2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}}}{\Gamma(n/2)\,\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right) 2^{(n+p\bar{\boldsymbol{\nu}})/2}} \Gamma\left((n+p\bar{\boldsymbol{\nu}})/2\right) \left(\frac{1}{2}\left(1+\operatorname{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}} \\
&= \frac{\Gamma\left((n+p\bar{\boldsymbol{\nu}})/2\right)}{\Gamma(n/2)\,\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)} |\mathbf{\Omega}|_{\frac{\nu}{2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} \left(\left(1+\operatorname{tr}\left(\mathbf{\Omega}\mathbf{R}^{-1}\right)\right)\right)^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}},
\end{aligned}$$

where we used equation (5.9.1) of the NIST Digital Library of Mathematical Functions.

Inverse F-Riesz distribution: The stochastic representation an F-Riesz distribution of type II with scale matrix $\mathbf{\Omega}^{-1}$, and degree of freedom parameter vectors $\boldsymbol{\nu}$ and $\mathbf{n}$ is $\mathbf{U}_{\mathbf{\Omega}^{-1}} \underline{\mathbf{B}}^{-1} \bar{\mathbf{B}} \bar{\mathbf{B}}^\top \underline{\mathbf{B}}^{-1} \mathbf{U}_{\mathbf{\Omega}^{-1}}^\top$, where $\mathbf{U}_{\mathbf{\Omega}^{-1}}$ is the upper Cholesky factor of $\mathbf{\Omega}^{-1}$.[32] Thus the stochastic representation of the inverse F-Riesz distribution of type II is given by

$$\mathbf{R} = \mathbf{C}_{\mathbf{\Omega}} \underline{\mathbf{B}} \bar{\mathbf{B}}^{-\top} \bar{\mathbf{B}}^{-1} \underline{\mathbf{B}}^\top \mathbf{C}_{\mathbf{\Omega}}^\top, \tag{29}$$

which translate to $\mathbf{R} \sim i\mathcal{R}^{II}(\mathbf{Y}, \boldsymbol{\nu})$, $\mathbf{Y} \sim \mathcal{R}^{I}(\mathbf{\Omega}, \mathbf{n})$.[33] For the probability density

32. See Blasques et al. (2021).
33. Recall that $\mathbf{U}_{\mathbf{\Omega}^{-1}}^{-\top} = \mathbf{C}_{\mathbf{\Omega}}$.

function we can consequently use

$$
\begin{aligned}
p_{\iota_F R^{II}}(\mathbf{R}|\mathbf{\Omega},\mathbf{n},\boldsymbol{\nu}) &= \int_{\mathbf{Y}>\mathbf{0}} p_{\iota_R{}^{II}}(\mathbf{R}|\mathbf{Y},\boldsymbol{\nu}) p_{R^I}(\mathbf{Y}|\mathbf{\Omega},\mathbf{n}) d\mathbf{Y} \\
&= \int_{\mathbf{Y}>\mathbf{0}} \left(|\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{Y}|_{\frac{\nu}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}\mathbf{R}^{-1}\right) \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2) 2^{p\bar{\nu}/2}} \right. \\
&\qquad \left. \times |\mathbf{Y}|_{\frac{\mathbf{n}-p-1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Omega}^{-1}\mathbf{Y}\right) \frac{1}{\Gamma_p(\mathbf{n}/2)\, 2^{p\bar{\mathbf{n}}/2}} \right) d\mathbf{Y} \\
&= \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)\, 2^{p(\bar{\nu}+\bar{\mathbf{n}})/2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \\
&\quad \times \int_{\mathbf{Y}>\mathbf{0}} |\mathbf{Y}|_{\frac{\mathbf{n}+\boldsymbol{\nu}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1})\right) d\mathbf{Y} \\
&= \frac{2^{p(\bar{\nu}+\bar{\mathbf{n}})/2}\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)\, 2^{p(\bar{\nu}+\bar{\mathbf{n}})/2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \left| \left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1} \right|_{\frac{\nu+\mathbf{n}}{2}} \\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \left| \left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1} \right|_{\frac{\nu+\mathbf{n}}{2}},
\end{aligned}
$$

where we used Theorem 8.1. Now rewrite using Lemma 3 in Blasques et al. (2021).

□

8.1.4 Proof of Theorem 2.2

Proof. We will make use of properties of probability limits of products of (inverse) random matrices and of Slutzky's Theorems for random matrices (see e.g. Theorems 5.6, 5.9 and 5.10 in Mittelhammer (2013)).

The non-zero off-diagonal elements of the lower triangular matrix $\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}$, $i<j$ are given by

$$
\left(\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\right)_{ij} = \frac{(\underline{\mathbf{B}})_{ij}}{\sqrt{n_i}} \xrightarrow{p} 0, \text{ as } n_i \to \infty,
$$

since $(\underline{\mathbf{B}})_{ij} \sim \mathcal{N}(0,1)$ for $i<j$.

Furthermore, note that for the *squared* diagonal ($i=j$) elements we have for $n_i \to \infty$,

$$
\begin{aligned}
\mathbb{E}\left[\frac{\left((\underline{\mathbf{B}})_{ii}\right)^2}{n_i}\right] &= \frac{n_i-i+1}{n_i} \to 1 \quad \text{and} \\
\mathbf{Var}\left(\frac{\left((\underline{\mathbf{B}})_{ii}\right)^2}{n_i}\right) &= 2\frac{n_i-i+1}{n_i^2} \to 0,
\end{aligned}
$$

since $\left(\left(\underline{\mathbf{B}}\right)_{ii}\right)^2$ is $\chi^2_{n_i-i+1}$ distributed, and thus as $n_i \to \infty$

$$\frac{\left(\left(\underline{\mathbf{B}}\right)_{ii}\right)^2}{n_i} \xrightarrow{p} 1 \Leftrightarrow \frac{\left(\underline{\mathbf{B}}\right)_{ii}}{\sqrt{n_i}} \xrightarrow{p} 1,$$

where the equivalence follows from the Continuous Mapping Theorem. Finally, we can conclude that as $n_i \to \infty$ for all i,

$$\operatorname*{plim}_{\mathbf{n}\to\infty} \left(\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\right) = \mathbf{I}, \tag{30}$$

where $\mathbf{n} \to \infty$ means that all elements in $\mathbf{n}$ converge to infinity and the plim operator on a matrix is to be understood element-wise. By similar arguments we get that as $\nu_i \to \infty$ for all i,

$$\operatorname*{plim}_{\boldsymbol{\nu}\to\infty} \left(\bar{\mathbf{B}}\mathrm{dg}(\boldsymbol{\nu})^{-\frac{1}{2}}\right) = \mathbf{I}$$

and consequently

$$\operatorname*{plim}_{\boldsymbol{\nu}\to\infty} \left(\left(\bar{\mathbf{B}}\mathrm{dg}(\boldsymbol{\nu})^{-\frac{1}{2}}\right)^{-1}\right) = \mathbf{I}. \tag{31}$$

Now it is easy to see from (8) that

$$\mathrm{dg}(\boldsymbol{\nu})^{-1}\mathrm{dg}(\mathbf{n}) \leq \mathrm{dg}\left(\overset{\mathcal{F}\mathcal{R}'}{\mathbf{m}}\right),$$

where the $\leq$-relation between two equal sized matrices is meant to be element-wise and thus since both matrices of the inequality are diagonal matrices,

$$\mathrm{dg}\left(\overset{\mathcal{F}\mathcal{R}'}{\mathbf{m}}\right)^{-\frac{1}{2}} \leq \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\mathrm{dg}(\boldsymbol{\nu})^{\frac{1}{2}}$$

and thus

$$\mathrm{dg}\left(\overset{\mathcal{F}\mathcal{R}'}{\mathbf{m}}\right)^{-\frac{1}{2}} \bar{\mathbf{B}}^{-\top} \leq \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\mathrm{dg}(\boldsymbol{\nu})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}. \tag{32}$$

By similar reasoning using equation 9 we get

$$\mathrm{dg}\left(\overset{i\mathcal{F}\mathcal{R}''}{\mathbf{m}}\right)^{-\frac{1}{2}} \underline{\mathbf{B}} \leq \mathrm{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}. \tag{33}$$

Now equations (30) and (32) imply

$$\operatorname*{plim}_{\boldsymbol{\nu}\to\infty}\left(\operatorname{dg}\left(\overset{\mathcal{FR}'}{\mathbf{m}}\right)^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\right)=\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}},$$

which using Slutzky implies

$$\operatorname{dg}\left(\overset{\mathcal{FR}'}{\mathbf{m}}\right)^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\xrightarrow[\boldsymbol{\nu}\to\infty]{d}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}, \tag{34}$$

and equations (31) and (33) imply

$$\operatorname*{plim}_{\mathbf{n}\to\infty}\left(\operatorname{dg}\left(\overset{i\mathcal{FR}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\underline{\mathbf{B}}\right)=\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}},$$

which implies

$$\operatorname{dg}\left(\overset{i\mathcal{FR}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\xrightarrow[\mathbf{n}\to\infty]{d}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}. \tag{35}$$

Finally equations (34) and (35) imply

$$\begin{aligned}&\operatorname{dg}\left(\overset{\mathcal{FR}'}{\mathbf{m}}\right)^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}\operatorname{dg}\left(\overset{\mathcal{FR}'}{\mathbf{m}}\right)^{-\frac{1}{2}}\xrightarrow[\boldsymbol{\nu}\to\infty]{d}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\quad\text{and}\\&\operatorname{dg}\left(\overset{i\mathcal{FR}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\operatorname{dg}\left(\overset{i\mathcal{FR}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\xrightarrow[\mathbf{n}\to\infty]{d}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}},\end{aligned}$$

which are the stochastic representations of the standardized Riesz I and inverse Riesz II ($\mathbf{M}_{\mathcal{D}}^{-1/2}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{-1/2}$), respectively.

The proofs for

$$\begin{aligned}(\nu-2)\left(\chi^2_\nu\right)^{-1}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}&\xrightarrow[\nu\to\infty]{d}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\operatorname{dg}(\mathbf{n})^{-\frac{1}{2}}\quad\text{and}\\\frac{\chi^2_n}{n}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}&\xrightarrow[n\to\infty]{d}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\operatorname{dg}(\overset{i\mathcal{R}''}{\mathbf{m}})^{\frac{1}{2}}\end{aligned}$$

are very easy, noticing that $(\nu-2)\left(\chi^2_\nu\right)^{-1}$ and χ^2_n/n converge in probability to 1. □

8.1.5 Proof of Theorem 3.1

Proof. It is well-known that $\dot{\mathbf{r}}$ has stochastic representation $\dot{\mathbf{r}}=\sqrt{y}\mathbf{C}_{\dot{\boldsymbol{\Omega}}}\dot{\mathbf{z}}$ with $\dot{z}_i\overset{iid}{\sim}\mathcal{N}(0,1)$ and $y\sim\Gamma(\nu/2,2/\nu)$. Equivalently, we can fix all entries with missing

observations in $\tilde{\mathbf{r}}$ equal to zero and write $\tilde{\mathbf{r}} = \sqrt{y}\mathbf{C}_{\tilde{\mathbf{\Omega}}}\tilde{\mathbf{z}}$, with

$$\tilde{z}_i = \begin{cases} 0 \text{ if there is a missing observation,} \\ \overset{iid}{\sim} \mathcal{N}(0,1) \text{ else.} \end{cases}$$

The RC can be written as

$$\mathbf{R} \;=\; \sum_{j=1}^{n} \mathbf{r}_j \mathbf{r}_j^\top = y\mathbf{C}_{\mathbf{\Omega}}\mathbf{Z}\mathbf{Z}^\top\mathbf{C}_{\mathbf{\Omega}}^\top,$$

where the $p \times n$ matrix $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots \mathbf{z}_n)$ with $\mathbf{z}_j = \mathbf{C}_{\mathbf{\Omega}}^{-1}\mathbf{r}_j$. Hassairi, Ktari, and Zine (2022) show that if the assets in $\mathbf{Z}$ are sorted according to their liquidity with the least liquid asset in the first row, then $\mathbf{Z}\mathbf{Z}^\top$ follows a Riesz distribution with parameter matrix parameter matrix $\Omega = \mathbf{I}$ and degree of freedom parameter vector $\mathbf{n}$, which implies that $\mathbf{C}_{\mathbf{\Omega}}\mathbf{Z}\mathbf{Z}^\top\mathbf{C}_{\mathbf{\Omega}}^\top \sim \mathcal{R}(\mathbf{\Omega}, \mathbf{n})$ follows a Riesz distribution with parameter matrix $\mathbf{\Omega}$ and degree of freedom parameter vector $\mathbf{n}$. Then according to Theorem 2.3 $\mathbf{R} \sim t\mathcal{R}(\mathbf{\Omega}, \mathbf{n}, \nu)$. □

8.1.6 Proof of Theorem 3.2

Proof. This Theorem is closely based on Gupta, Varga, and Bodnar (2013). If $\tilde{\mathbf{r}}$ follows an elliptically contoured distribution, $\tilde{\mathbf{r}} \sim E_{np}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Omega}, \psi)$, then according to their Theorem 2.1 $\mathbf{X}^\top \sim E_{n,p}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{\Omega}, \psi)$ and then according to their Theorem 2.3 $\mathbf{X} \sim E_{p,n}(\mathbf{0}, \mathbf{\Omega} \otimes \mathbf{I}_n, \psi)$. Then our Theorem follows from their Corollary 5.1 and noticing that

$$\begin{aligned} \operatorname{tr}\left(\mathbf{X}^\top\mathbf{\Omega}^{-1}\mathbf{X}\right) &\overset{(??)}{=} \operatorname{vec}(\mathbf{X})^\top \operatorname{vec}(\mathbf{\Omega}^{-1}\mathbf{X}) \overset{(??)}{=} \operatorname{vec}(\mathbf{X})^\top(\mathbf{I} \otimes \mathbf{\Omega}^{-1})\operatorname{vec}(\mathbf{X}) \\ &= \tilde{\mathbf{r}}^\top(\mathbf{I} \otimes \mathbf{\Omega}^{-1})\tilde{\mathbf{r}} \text{ and} \end{aligned}$$

$$\mathbf{R} \;=\; \mathbf{X}\mathbf{X}^\top = \sum_{j=1}^{n} \mathbf{r}_j \mathbf{r}_j^\top.$$

□

8.2 Additional Material

Distribution	$\boldsymbol{\theta}_{\mathcal{D}}$	Distribution	$\boldsymbol{\theta}_{\mathcal{D}}$
Wishart	n	Riesz	$\mathbf{n}$
Inv.Wishart	ν	Inv.Riesz	$\boldsymbol{\nu}$
t-Wishart	$(n, \nu)^\top$	t-Riesz	$(\mathbf{n}^\top, \nu)^\top$
Inv.t-Wishart	$(n, \nu)^\top$	Inv.t-Riesz	$(n, \boldsymbol{\nu}^\top)^\top$
F	$(n, \nu)^\top$	F-Riesz	$(\mathbf{n}^\top, \boldsymbol{\nu}^\top)^\top$
F	$(n, \nu)^\top$	Inv.F-Riesz	$(\mathbf{n}^\top, \boldsymbol{\nu}^\top)^\top$

Table 10: Degree of freedom parameters.

Distribution	Probability Density Function, $p_{\mathcal{D}}(\mathbf{R}\vert\boldsymbol{\Sigma},\boldsymbol{\theta})$
Wishart	$\frac{n^{np/2}}{2^{np/2}}\frac{1}{\Gamma_p(n/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert^{\frac{n}{2}}\ \operatorname{etr}\left(-\frac{1}{2}n\mathbf{Z}\right)$
Riesz	$\frac{\prod_{i=1}^p n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}}\frac{1}{\Gamma_p(\mathbf{n}/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{\frac{\mathbf{n}}{2}}\ \operatorname{etr}\left(-\frac{1}{2}\operatorname{dg}(\mathbf{n})\mathbf{Z}\right)$
Inv.Wishart	$\frac{(\nu-p-1)^{\nu p/2}}{2^{\nu p/2}}\frac{1}{\Gamma_p(\nu/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert^{-\frac{\nu}{2}}\ \operatorname{etr}\left(-\frac{1}{2}(\nu-p-1)\mathbf{Z}^{-1}\right)$
Inv.Riesz	$\frac{\prod_{i=1}^p m_i^{-\nu_i/2}}{2^{p\overleftarrow{\boldsymbol{\nu}}/2}}\frac{1}{\Gamma_p\left(\overleftarrow{\boldsymbol{\nu}}/2\right)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{-\frac{\boldsymbol{\nu}}{2}}\ \operatorname{etr}\left(-\frac{1}{2}\operatorname{dg}(\mathbf{m})^{-1}\mathbf{Z}^{-1}\right)$
t-Wishart	$\left(\frac{n}{\nu-2}\right)^{pn/2}\frac{\Gamma((\nu+pn)/2)}{\Gamma_p(n/2)\Gamma(\nu/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert^{\frac{n}{2}}\ \left(1+\frac{n}{\nu-2}\operatorname{tr}(\mathbf{Z})\right)^{-\frac{\nu+pn}{2}}$
t-Riesz	$\frac{\prod_{i=1}^p n_i^{n_i/2}}{(\nu-2)^{p\bar{\mathbf{n}}/2}}\frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{\frac{\mathbf{n}}{2}}\ \left(1+\frac{1}{\nu-2}\operatorname{tr}\left(\operatorname{dg}(\mathbf{n})\mathbf{Z}\right)\right)^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}$
Inv.t-Wishart	$\left(\frac{\nu-p-1}{n}\right)^{\nu p/2}\frac{\Gamma((n+p\nu)/2)}{\Gamma_p(\nu/2)\Gamma(n/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert^{-\frac{\nu}{2}}\ \left(1+\frac{\nu-p-1}{n}\operatorname{tr}(\mathbf{Z}^{-1})\right)^{-\frac{n+p\nu}{2}}$
Inv.t-Riesz	$\frac{\prod_{i=1}^p(\overset{i\mathbb{R}''}{m}_i)^{-\nu_i/2}}{n^{p\overleftarrow{\boldsymbol{\nu}}/2}}\frac{\Gamma((n+p\overleftarrow{\boldsymbol{\nu}})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma(n/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{-\frac{\boldsymbol{\nu}}{2}}\ \left(1+\frac{1}{n}\operatorname{tr}\left(\operatorname{dg}(\overset{i\mathbb{R}''}{\mathbf{m}})^{-1}\mathbf{Z}^{-1}\right)\right)^{-\frac{n+p\overleftarrow{\boldsymbol{\nu}}}{2}}$
F	$\left(\frac{n}{\nu-p-1}\right)^{np/2}\frac{\Gamma_p((\nu+n)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert^{\frac{n}{2}}\ \left\vert\mathbf{I}+\frac{n}{\nu-p-1}\mathbf{Z}\right\vert^{-\frac{\nu+n}{2}}$
F-Riesz	$\prod_{i=1}^p(\overset{\mathbb{F}\mathbb{R}'}{m}_i)^{n_i/2}\frac{\Gamma_p\left((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2\right)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{\frac{\mathbf{n}}{2}}\ \left\vert\mathbf{I}+\operatorname{dg}(\overset{\mathbb{F}\mathbb{R}'}{\mathbf{m}})^{1/2}\mathbf{Z}\operatorname{dg}(\overset{\mathbb{F}\mathbb{R}'}{\mathbf{m}})^{1/2}\right\vert_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$
Inv.F-Riesz	$\prod_{i=1}^p\left(\overset{i\mathbb{R}''}{m}_i\right)^{-\nu_i/2}\frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)}\vert\mathbf{R}\vert^{-\frac{p+1}{2}}\vert\mathbf{Z}\vert_{-\frac{\boldsymbol{\nu}}{2}}\ \left\vert\left(\mathbf{I}+\operatorname{dg}\left(\overset{i\mathbb{R}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\mathbf{Z}^{-1}\operatorname{dg}\left(\overset{i\mathbb{R}''}{\mathbf{m}}\right)^{-\frac{1}{2}}\right)^{-1}\right\vert_{\frac{\boldsymbol{\nu}+\mathbf{n}}{2}}$

Table 11: Probability density functions of all considered distributions. We define $\mathbf{Z}=\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}$, where $\mathbf{C}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}$. For the definition of $\overset{i\mathbb{R}''}{\mathbf{m}}$, $\overset{\mathbb{F}\mathbb{R}'}{\mathbf{m}}$ and $\overset{\mathbb{F}\mathbb{R}''}{\mathbf{m}}$ Theorem 2.1. To derive these representations from the ones in Table 3 use Lemma **??** and 2.3.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	485	600	1121	5104	4829	3769
Riesz	449	552	984	4305	4066	2986
Inv.Wishart	420	497	775	2526	3052	2201
Inv.Riesz	407	475	730	2269	2630	1891
t-Wishart	344	364	548	1697	2423	1723
t-Riesz	330	347	510	1491	2104	1344
Inv.t-Wishart	335	346	468	1538	1809	1018
Inv.t-Riesz	325	335	445	1357	1610	848
F	410	484	738	2341	2656	1979
F-Riesz	346	389	494	1765	1833	1164
Inv.F-Riesz	355	400	533	1847	1885	1237

Table 12: Log score loss over a one month forecasting period (22 trading days), $-\sum_{j=1}^{22} p_{\mathcal{D}}(\mathbf{R}_{t+j}|\widehat{\boldsymbol{\Sigma}}_{j+1}, \hat{\boldsymbol{\theta}}_{j+1})$, for the forecasting window from 1 January 2007 to 31 December 2011, where each model is re-estimated every 10 trading days. 90% model confidence sets in red.

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	117	357	-87	-1467	-1502	-1940
Riesz	89	324	-197	-1695	-2160	-2611
Inv.Wishart	76	290	-295	-1786	-2823	-3534
Inv.Riesz	67	270	-349	-1948	-3075	-3792
t-Wishart	15	227	-393	-2302	-2548	-3037
t-Riesz	0	203	-468	-2483	-2994	-3561
Inv.t-Wishart	-14	192	-528	-2586	-3584	-4297
Inv.t-Riesz	-21	180	-571	-2667	-3729	-4490
F	57	280	-340	-2026	-3058	-3655
F-Riesz	-5	189	-554	-2455	-3764	-4428
Inv.F-Riesz	-1	198	-537	-2414	-3711	-4343

Table 13: Log score loss over a one month forecasting period (22 trading days), $-\sum_{j=1}^{22} p_{\mathcal{D}}(\mathbf{R}_{t+j}|\widehat{\boldsymbol{\Sigma}}_{j+1}, \hat{\boldsymbol{\theta}}_{j+1})$, for the forecasting window from 1 January 2012 to 31 December 2019, where each model is re-estimated every 10 trading days. 90% model confidence sets in red.