Probability Distributions for Realized Covariance Measures

Michael Stollenwerk

Heidelberg University

November 18, 2023

Abstract

Realized covariance measures (RCs) are an essential input for assessing the risks of investment allocations. Thus useful to model and forecast them. To this end, a realistic distributional assumption is needed. In this paper, we compare all probability distributions that have so far been applied in the academic literature to time-series of RCs. We derive them in a unified framework based on their stochastic representations in terms of random lower and upper triangular (Barlett) matrices. These matrices are composed of standard normal distributions in the off-diagonal elements and χ-distributions on the diagonals. Furthermore, we derive a novel family of probability distributions, which has a property called *tail-homogeneity*. This property means that in crisis periods, i.e. large RCs, our distribution family assumes high dependence between the individual entries of the RCs ("homogeneous tails"). We show theoretically how the distributions differ from each other in terms of their tail-behaviour, and in terms of the implied marginal distributions of their log-determinants (a measure of the size of the RCs), realized variances, and covariances between the realized variances. Finally, we show rigorously how the distributions are related to each other. In the empirical part, we connect the previously derived theoretical differences to differences in fit and forecasting performance. We show that our novel distribution family achieves the best fit. Out-of-sample forecasting comparisons further corroborate the excellent performance of our novel distribution family.

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

1 Introduction

The covariance matrix of financial asset returns is an essential object in the financial econometrics literature because it has direct implications for efficient portfolio allocation and can be used in risk management and derivative pricing. A *realized covariance matrix* (RC) is an accurate and consistent ex-post estimate of the integrated covariance matrix of financial asset returns over a trading day. It is constructed from high-frequency data and can be interpreted as making the daily covariance matrix of the underlying financial asset returns "effectively observable". As such, it is advocated to model time-series of RCs directly (see e.g. Andersen et al. 2001, Andersen et al. 2006, McAleer and Medeiros 2008, Chiriac and Voev 2011). This direct modelling is in contrast to traditional multivariate volatility models (pioneered by Engle and Kroner 1995 and Engle 2002), which treat the covariance matrix as latent, and are based on low-frequency daily return data. In recent years, the increasing availability of high-frequency data has led to an enormous growth of models designed for RCs. These are often *observation-driven models* in which RCs are treated as random matrices with time-varying conditional distributions. While many different models that feature different probability distributions have been proposed, the explicit comparison of the theoretical and empirical properties of these probability distributions has not been the focus of a study to the best of our knowledge.

This paper aims to fill this gap in the literature. The comparison is important since choosing a probability distribution that accurately reflects the characteristics of RC data is crucial. Depending on the assumed distribution, different courses of action might result for practitioners. For instance, consider an investor who wants to invest in the predicted global minimum variance portfolio. If she is already invested in the underlying assets, the covariance matrix forecast almost surely implies different optimal weights than her current allocation. But should she reweight her portfolio, or is her current allocation "within reasonable distance" to the optimal weights? To answer this question, she must know the probability distribution of the forecasted covariance matrix, i.e. of the forecasted RC. Adopting a well-fitting distribution is also in itself an important aim. After all, since empirically, the realized variances (RVs), i.e. the diagonal elements of the RCs, have

high variance, are fat-tailed, and are right-skewed (Opschoor and Lucas 2022), the realized covariances (RCOVs) are more often positive than negative, and the RV of one asset tends to be large if the one of another asset is large (as is true in our data), it is logical to choose a probability distribution that can reflect these properties.

We outline the specific contributions of this paper. The first is, that we collate theoretical knowledge about the different hitherto used probability distributions into one place and common notation. This uncovers theoretical similarities and differences among the distributions, which is useful in explaining their differences in empirical fit and forecasting performance. Second, we derive a new distribution family, the *t-Riesz distribution family*, which is based on the t-Wishart distribution by Sutradhar and Ali (1989). We show that it has desirable theoretical and empirical properties, is closely related to the hitherto suggested distributions, and can be grounded in a realistic distributional assumption on the intraday return vectors from which an RC is constructed. Furthermore, we demonstrate that the t-Riesz distribution family offers a particularly good fit for the RC data. We contribute further by showing that all distributions, including the t-Riesz distribution family, can be regarded as belonging to a common overarching family. In this context, we rigorously show how exactly all distributions are related to each other. Many of these relations have not been previously derived. In the empirical part of this paper, we perform fit and forecasting comparisons of the different distributions in different datasets of time-series of RCs and explain how the theoretical differences translate into differences in fit and forecasting performance. Finally, other minor contributions of this paper are the discovery and derivation of another distribution (the Inverse F-Riesz) and showing that empirically it suffices to consider one of the two versions of Riesz distributions[1].

We now present the literature on probability distributions hitherto applied to time-series of RCs, citing not their original sources, but the ones that introduced them to financial econometrics. Probability distributions for RCs must generate symmetric positive definite random matrices. The first one proposed for RCs was the Non-Central Wishart in Gourieroux, Jasiak, and Sufana 2009. Their Wishart Autoregressive process (WAR) can be derived from a continuous-time multivari-

1. Each Riesz-type distribution has two versions.

ate Ornstein-Uhlenbeck price process. It can be seen as a multivariate extension of the Cox-Ingersoll-Ross model and is simple to estimate. The model essentially assumes that the non-centrality matrix varies over time, driven by lagged values of the RCs, while the scale matrix stays constant. Yu, Li, and Ng (2017) also assume the Non-Central Wishart distribution. Their model is a generalization of the WAR model by Gourieroux, Jasiak, and Sufana (2009) (and the CAW model by Golosnoy, Gribisch, and Liesenfeld (2012) presented below), where both the scale matrix and the non-centrality matrix depend on the lagged values of the RCs. However, this model is not applicable to dimensions of, say, more than three assets because the likelihood computation of the Non-Central Wishart distribution involves the numerical approximation of the matrix-variate hypergeometric function (see Koev and Edelman 2006), which is prohibitively slow. Golosnoy, Gribisch, and Liesenfeld (2012) propose the conditional autoregressive Wishart (CAW) model, which assumes that the conditional expected value matrix of the Wishart distribution follows BEKK dynamics (cf. Engle and Kroner 1995 for BEKK dynamics). Gorgi et al. (2019) also use the Wishart matrix but assume *generalized autoregressive score* (GAS) dynamics for the expected value matrix. The Wishart distribution has been well-studied (see e.g. Gupta and Nagar 2000) and is attractive because of its relative simplicity. It is a special case of the Non-Central Wishart distribution where the non-centrality matrix equals the zero matrix. Furthermore, it can be derived as the sum of outer products of independent and identically distributed (i.i.d.) normally distributed random vectors and arises naturally in many areas of statistics. However, the Wishart has the disadvantage of being a *thin-tailed* distribution, which is in contrast to the empirically *fat-tailed* RCs. Asai and So (2013) and Jin and Maheu (2016) use the Inverse Wishart distribution in various model frameworks. Jin and Maheu (2016) show that the Inverse Wishart distribution is better at modelling the conditional density of realized covariance matrices than the Wishart distribution. The Inverse Wishart can be considered fatter-tailed than the Wishart distribution.[2] The first large advancement in distributional fit was accomplished by Opschoor et al. (2018) who propose the matrix-F distribution[3] and show that it implies fatter tails for the RVs than

2. See Section 5.3.2
3. From now on we call the matrix-F simply the F distribution.

previously used distributions. They employ the F distribution in a GAS model and demonstrate that it significantly outperforms previously proposed (Inverse) Wishart distribution-based models. Zhou et al. (2019) confirm these results. Recently, Blasques et al. (2021) introduced the Riesz, Inverse Riesz, and F-Riesz distributions into financial econometrics. These Riesz-type distributions generalize, respectively, the Wishart, Inverse Wishart, and F distributions by featuring *degree of freedom* (d.o.f.) parameter vectors instead of scalars, thus adding flexibility. Blasques et al. (2021) propose an efficient algorithm to estimate the Riesz-type distributions. Furthermore, they show that there is again a large increase in fit and forecasting ability from the F to the F-Riesz distribution and attribute this to the fact that the F-Riesz distribution features heterogeneous tails for the realized variances. Finally, Gribisch and Hartkopf (2022) show that the Riesz distribution for the standard realized covariance matrix can be derived by assuming a normal distribution on the intraday return innovations and sorting the assets according to their liquidity, where the asset with the most zero intraday returns is interpreted to be least liquid.

We note here that an alternative to modelling the RCs themselves is to transform them via e.g. the Cholesky decomposition (Chiriac and Voev 2011), the matrix logarithm (Bauer and Vorkink 2011), the decomposition into realized correlation matrices and realized variances (Bauwens, Storti, and Violante 2012), or the matrix logarithm of the realized correlation matrix (Archakov and Hansen 2021), and then model those transformed time-series. Notably, the model of Archakov and Hansen (2021) is invariant to the ordering of the assets, naturally accommodates positive definiteness of the implied RCs, and the individual elements of the transformed time-series exhibit much-reduced dependence amongst each other.

The rest of this paper is structured as follows. The next section derives all probability distributions[4] in a unified framework based on their stochastic representations, analyzes the distribution's stochastic properties, shows how the distributions relate to each other, and derives the probability density functions. Section 3 shows how the newly derived t-Riesz distribution family can be derived from a reasonable low-level assumption on the intraday return vectors (as an alternative to deriving it from the random triangular matrices). Section 4 introduces time-

4. Excluding the Non-Central Wishart.

variation in the expected value parameter matrix of the distributions, in Section 5 we discuss estimation and perform the in-sample fit comparison and out-of-sample forecasting performance analysis of the different distributions, and 6 concludes.

2 Probability Distributions

Let $\mathbf{R}$ denote the $p \times p$ symmetric positive definite realized covariance matrix of p asset returns on a given day and assume that it follows a probability distribution $\mathcal{D}$ with support on symmetric positive semidefinite matrices. Later on, we will add a time index t indicating the day $t = 1, \ldots, T$ and write $\mathbf{R}_t$; for now, we opt for better readability.

All[5] hitherto in the literature considered probability distributions for $\mathbf{R}$ are characterized by a $p \times p$ positive definite *parameter matrix* $\mathbf{\Omega}$ and a distribution-specific set of *degree of freedom* (d.o.f.) parameters $\boldsymbol{\theta}_{\mathcal{D}}$, such that we write

$$\mathbf{R} \sim \mathcal{D}(\mathbf{\Omega}, \boldsymbol{\theta}_{\mathcal{D}}). \tag{1}$$

Their characterization in terms of $\mathbf{\Omega}$ and $\boldsymbol{\theta}_{\mathcal{D}}$ will become apparent by examining the distributions' stochastic representations, which we will do next. The stochastic representations are central to this paper since based on them we introduce the various distributions, analyze their stochastic properties, derive the distributions' expected values and probability density functions (p.d.f.s) as well as their relationships to one another.

2.1 Stochastic Representations

All hitherto in the literature considered probability distributions for $\mathbf{R}$ and the ones newly proposed in this paper can be generated from the $p \times p$ random trian-

5. Excluding the Non-Central Wishart. Since it doesn't fit into the theoretical framework of this paper, and because the empirical results indicate that other distributions fit much better to RC data, we exclude the Non-Central Wishart from the theoretical considerations in this paper.

gular 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

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

and/or

$$(\bar{\mathbf{B}})_{ij} \sim \begin{cases} \chi_{\nu_i-p+i} & \text{for } i = j, \\ \mathcal{N}(0,1) & \text{for } i > j, \end{cases} \tag{3}$$

where all random variables inside the matrices are independent of each other.[6] We refer to these random matrices as *Bartlett matrices* since Bartlett (1933) was the first to show that the Wishart random matrix can be generated from a special case of those triangular matrices. The corresponding decomposition of the Wishart random matrix is known in the literature as *Barlett decomposition*.

The parameters n_i and ν_i are the aforementioned d.o.f. parameters, which we collect in the $p \times 1$ vectors $\mathbf{n} = (n_1, \ldots, n_p)^\top$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)^\top$. The special cases of the Bartlett matrices, where for all i, $n_i = n$ and $\nu_i = \nu$, we denote as $\underline{\mathcal{B}}$ and $\bar{\mathcal{B}}$, respectively. That is $\underline{\mathcal{B}}$ and $\bar{\mathcal{B}}$ are governed by d.o.f. parameters n

6. The χ_n distribution is given in e.g. Walck (2007), Section 8.14.

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

Table 1: Stochastic representation kernels $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ of all distributions for RCs. The complete stochastic representations are given by $\mathbf{C}_{\Omega}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{C}_{\Omega}^{\top}$, where $\mathbf{C}_{\Omega}$ is the lower Cholesky factor of the $p \times p$ symmetric positive definite parameter matrix $\boldsymbol{\Omega} = \mathbf{C}_{\Omega}\mathbf{C}_{\Omega}^{\top}$. $\underline{b}$ and $\bar{b}$ are χ_n and χ_ν distributed random variables, thus can be interpreted as one-dimensional $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$, respectively. $\underline{\mathcal{B}}$ ($\bar{\mathcal{B}}$) is the special cases of $\underline{\mathbf{B}}$ ($\bar{\mathbf{B}}$) where for all i, $n_i = n$ ($\nu_i = \nu$).

and ν, and $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$ are governed by d.o.f. parameter vectors $\mathbf{n}$ and $\boldsymbol{\nu}$. Note that for $p = 1$, the Bartlett matrices reduce to the random variables χ_n and χ_ν, respectively. For the matrix distributions to exist, we must restrict $n_i > i - 1$ and $\nu_i > p - i$ since otherwise the χ distributions on the main diagonals would not exist.[7]

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 Table 1 for their full names). In Section 2.2 below we will have a closer look at the distributions one by one and analyze their stochastic properties. The ones in green are the novel distributions derived in this paper. Assuming that $\mathbf{R}$ follows one of the distributions $\mathcal{D}$, its stochastic representation can be written as

$$\mathbf{R} = \mathbf{C}_{\Omega}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{C}_{\Omega}^{\top}, \tag{4}$$

where $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ is a distribution-specific function of one or both of the Bartlett matrices or their special cases, and $\mathbf{C}_{\Omega}$ denotes the lower Cholesky factor of symmetric positive definite parameter matrix $\boldsymbol{\Omega}$. We call $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ *stochastic representation kernel*. The distribution parameters are thus given by the parameter matrix $\boldsymbol{\Omega}$ and the

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

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 2: Degree of freedom parameters of distributions for RCs.

d.o.f. parameter(s) in $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ (one or two of the set $(n, \nu, \mathbf{n}, \boldsymbol{\nu})$). We collect the distribution-specific d.o.f. parameters in the vector $\boldsymbol{\theta}_{\mathcal{D}}$. The exact composition of $\boldsymbol{\theta}_{\mathcal{D}}$ for the different distributions is given in Table 2.[8]

2.1.1 The Expected Value Matrix $\boldsymbol{\Sigma}$

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

$$\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top := \mathbb{E}\left[\mathbf{R}\right], \tag{5}$$

instead of in terms of $\boldsymbol{\Omega}$, where $\mathbf{C}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}$. This standardization allows for a simple two-step estimation strategy, where the $\mathcal{O}(p^2)$ expected value matrix $\boldsymbol{\Sigma}$ is estimated by its obvious method of moments estimator in the first step and the d.o.f. parameters in a second step. Furthermore, since $\boldsymbol{\Sigma}$ has the same interpretation across distributions (unlike $\boldsymbol{\Omega}$), standardization makes comparisons of the distributions easier. Additionally, the nesting relationships between the different distributions (see Figure 6 on p. 25) are only valid for the standardized, i.e. $\boldsymbol{\Sigma}$-parameterized, distributions.

We denote the expected value of the stochastic representation kernel by

$$\mathbf{M}_{\mathcal{D}} := \mathbb{E}\left[\boldsymbol{\mathcal{K}}_{\mathcal{D}}\right]. \tag{6}$$

8. It is easy to see from Table 1, which d.o.f. parameter(s) characterize each distribution.

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	$\mathrm{dg}(\mathring{\boldsymbol{\nu}})$
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	$\mathrm{dg}(\mathring{\boldsymbol{\nu}})n$
F	$\mathbf{I}\frac{n}{\nu-p-1}$	F-Riesz	$\mathrm{dg}(\mathring{\mathbf{n}})$
F	$\mathbf{I}\frac{n}{\nu-p-1}$	Inv.F-Riesz	$\mathrm{dg}(\mathring{\ddot{\mathbf{n}}})$

Table 3: Expected values of stochastic representation kernels, $\mathbf{M}_{\mathcal{D}} = \mathbb{E}[\boldsymbol{\mathcal{K}}_{\mathcal{D}}]$. For the definitions of $\mathring{\boldsymbol{\nu}}$, $\mathring{\mathbf{n}}$ and $\mathring{\ddot{\mathbf{n}}}$, see Theorem 2.1.

Then the stochastic representation of the standardized distributions is

$$\mathbf{R} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top}, \tag{7}$$

with

$$\mathbb{E}[\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top}] = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbb{E}\left[\boldsymbol{\mathcal{K}}_{\mathcal{D}}\right]\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}}\mathbf{C}^{\top} = \boldsymbol{\Sigma}. \tag{8}$$

This implies that

$$\mathbf{C}_{\Omega} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-\frac{1}{2}} \Leftrightarrow \mathbf{C} = \mathbf{C}_{\Omega}\mathbf{M}_{\mathcal{D}}^{\frac{1}{2}} \text{ and } \boldsymbol{\Omega} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}^{\top} \Leftrightarrow \boldsymbol{\Sigma} = \mathbf{C}_{\Omega}\mathbf{M}_{\mathcal{D}}\mathbf{C}_{\Omega}. \tag{9}$$

The expectations $\mathbf{M}_{\mathcal{D}}$ are straightforward to derive by applying Theorem 2.1 (p. 12). They are listed in Table 3. Notice that they are all 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 while the conditions for the existence of the distributions are fulfilled, then the expected value $\boldsymbol{\Sigma}$ does not exist. In this paper, we assume that the expected value always exists, and we can thus equivalently characterize the distribution in terms of $\boldsymbol{\Sigma}$ and write

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

Theorem 2.1. *Let* $\underline{\mathbf{B}}$ *and* $\bar{\mathbf{B}}$ *be defined as in equations (2) and (3). Then*

$$\begin{aligned}
\mathbb{E}[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top] &= \mathrm{dg}(\mathbf{n}), \textit{ (Díaz-García 2013)} \\
\mathbb{E}[\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\right)^{-1}] &= \mathrm{dg}(\mathring{\boldsymbol{\nu}}), \textit{ (Louati and Masmoudi 2015)} \\
\mathbb{E}[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}] &= \mathrm{dg}(\mathring{\mathbf{n}}), \textit{ (Blasques et al. 2021) and} \\
\mathbb{E}[\underline{\mathbf{B}}\left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\right)^{-1}\underline{\mathbf{B}}^\top] &= \mathrm{dg}(\mathring{\mathring{\mathbf{n}}}),
\end{aligned}$$

with the $p \times 1$ *vectors*[a]

$$\mathbf{n} = (n_1, n_2, \ldots, n_p)^\top,$$

$$\mathring{\boldsymbol{\nu}} = (\mathring{\nu}_1, \mathring{\nu}_2, \ldots, \mathring{\nu}_p)^\top,$$

$$\mathring{\nu}_i = \begin{cases} \frac{1}{\nu_i - p - 1}, & \textit{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(1 + \sum_{j=1}^{i-1} \mathring{\nu}_j\right) & \textit{for } i = 2, \ldots, p, \end{cases} \tag{11}$$

$$\mathring{\mathbf{n}} = (\mathring{n}_1, \mathring{n}_2, \ldots, \mathring{n}_p)^\top,$$

$$\mathring{n}_i = \begin{cases} \frac{n_1}{\nu_1 - p - 1}, & \textit{for } i = 1 \\ \frac{1}{\nu_i - p + i - 2}\left(n_i + \sum_{j=1}^{i-1} \mathring{n}_j\right) & \textit{for } i = 2, \ldots, p \end{cases} \quad \textit{and} \tag{12}$$

$$\mathring{\mathring{\mathbf{n}}} = (\mathring{\mathring{n}}_1, \mathring{\mathring{n}}_2, \ldots, \mathring{\mathring{n}}_p)^\top$$

$$\mathring{\mathring{n}}_i = \begin{cases} n_1\ \mathring{\nu}_1, & \textit{for } i = 1 \\ \sum_{j=1}^{i-1} \mathring{\nu}_j + (n_i - i + 1)\ \mathring{\nu}_i, & \textit{for } i = 2, \ldots, p. \end{cases} \tag{13}$$

Proof in Appendix.

a. If $\forall i$, $n_i = n$, $\nu_i = \nu$ then $\mathring{\nu}_i = (\nu - p - 1)^{-1}$ and $\mathring{n}_i = \mathring{\mathring{n}}_i = n(\nu - p - 1)^{-1}$. For $p = 1$ we obtain the expectations of a χ^2_n, an inverse χ^2_ν, and the χ^2_n/χ^2_ν ratio distribution.

2.2 The Individual Distributions

To understand the various distributions' stochastic properties, we now analyze and compare their stochastic representation kernels (Table 1). There are many properties amongst which we could compare the distributions. We focus on the marginal distribution of the RVs, i.e. the diagonal elements, since they are of special interest, and on the dependence amongst the individual elements of the

stochastic representation kernels.

We start with the Wishart distribution ($\mathcal{W}$), which we treat as our baseline case,

$$\underline{\mathcal{B}}\underline{\mathcal{B}}^\top.$$

The diagonal elements of the stochastic representation kernel of the Wishart follow independent χ^2_n-distributions, which is easy to see by noticing that

$$(\underline{\mathcal{B}}\underline{\mathcal{B}}^\top)_{ii} = \sum_{k=1}^{p}(\underline{\mathcal{B}})_{ik}(\underline{\mathcal{B}})_{ki} = \sum_{k=1}^{p}(\underline{\mathcal{B}})^2_{ik} = \sum_{k=1}^{i}(\underline{\mathcal{B}})^2_{ik}, \tag{14}$$

which is a sum of a χ^2_{n-i+1} and $i-1$ independent $(N(0,1))^2$ random variables, which yields a χ^2_n random variable. The Riesz distribution ($\mathcal{R}$),

$$\underline{\mathbf{B}}\underline{\mathbf{B}}^\top,$$

adds flexibility by allowing for different d.o.f. parameters of the χ-distributions on the main diagonal of the Bartlett matrix, which leads to the diagonal elements being distributed as $(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top)_{ii} \overset{iid}{\sim} \chi^2_{n_i}$.

The t-Wishart ($t\mathcal{W}$) has stochastic representation kernel

$$(\bar{b})^{-2}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top,$$

where the term $(\bar{b})^{-2}$ corresponds to an inverse χ^2_ν random variable. It is immediately obvious, that the scalar multiplication of every element in $\underline{\mathcal{B}}\underline{\mathcal{B}}^\top$ by $(\bar{b})^{-2}$ creates much more dependence among the individual elements than for the Wishart distribution. In particular, the diagonal elements of the stochastic representation kernel are now *dependent* random variables following a ratio distribution $\sim \chi^2_n/\chi^2_\nu$, where the χ^2_ν random variable is common across all diagonal elements and the χ^2_n random variable is specific to the index i. This implies that if one diagonal element has a large realization, the others are likely to have a large realization as well. Again, the Riesz version of the distribution, that is the t-Riesz ($t\mathcal{R}$) with

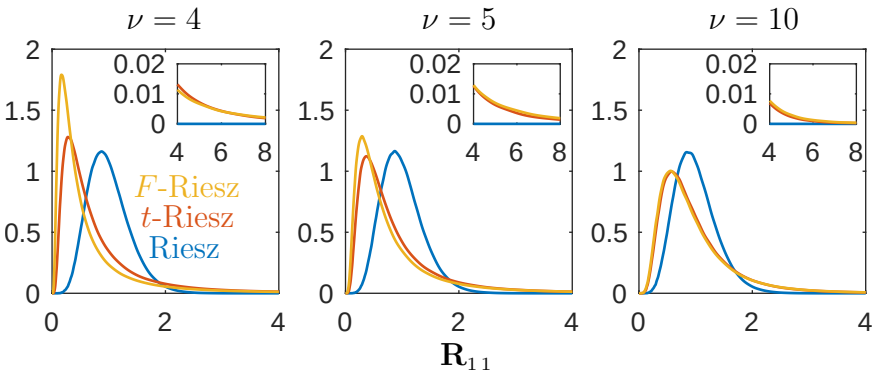

Figure 1: Marginal p.d.f.s of the first RV ($\mathbf{R}_{11}$) implied by the Riesz, t-Riesz, and F-Riesz distributions for a two-dimensional RC with the parameter setting $\mathbf{\Sigma} = \mathbf{I}_2$, $\mathbf{n} = (15, 30)^\top$, and with three different settings for the d.o.f. parameters, $\nu = 4, 5, 10$ (t-Riesz) and $\boldsymbol{\nu} = (4, 4)^\top, (5, 5)^\top, (10, 10)^\top$ (F-Riesz).

stochastic representation kernel

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

adds flexibility by allowing for the $\chi^2_{n_i}$ random variables to have different d.o.f. parameters.

The F distribution has stochastic representation kernel

$$\bar{\mathcal{B}}^{-\top}\underline{\mathcal{B}}\underline{\mathcal{B}}^\top\bar{\mathcal{B}}^{-1}.$$

Note that it is related to the t-Wishart in the sense that if the $p \times p$ Bartlett matrix $\bar{\mathcal{B}}$ was of dimension 1×1, i.e. a scalar, the stochastic representation kernel of the t-Wishart would arise. However, it is important to understand, that the F distribution does *not* nest the t-Wishart and can thus *not* be thought of as a more flexible version of it. Among the notable differences between the two distributions is that the t-Wishart features much higher dependence among the elements in the stochastic representation kernel. This is easy to see as for the t-Wishart every element in $\underline{\mathcal{B}}\underline{\mathcal{B}}^\top$ is scaled by $(\bar{b})^{-2}$, thus all elements are influenced by one random variable. For the F on the other hand $\underline{\mathcal{B}}\underline{\mathcal{B}}^\top$ is scaled by $\bar{\mathcal{B}}$ which itself consists of $p(p+1)/2$ independent random variables. The Riesz version of the F distribution, called the F-Riesz ($F\mathcal{R}$) adds flexibility by allowing the $\chi^2_{n_i}$

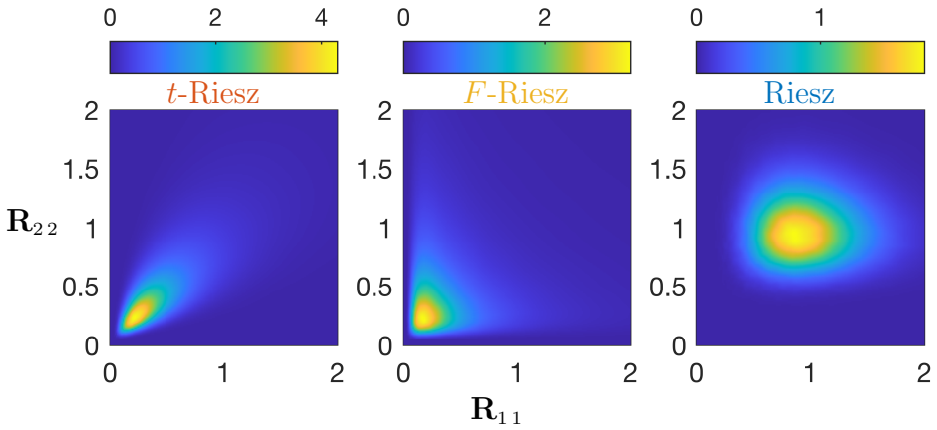

Figure 2: Heatmap of marginal joint p.d.f. of the first RV and second RV ($\mathbf{R}_{11}$, $\mathbf{R}_{22}$) implied by the Riesz, t-Riesz, and F-Riesz distributions for a two-dimensional RC with the parameter setting $\boldsymbol{\Sigma} = \mathbf{I}_2$, $\mathbf{n} = (15, 30)^\top$, $\nu = 5$ (t-Riesz), $\boldsymbol{\nu} = (5, 5)^\top$ (F-Riesz).

and $\chi^2_{\nu_i}$ random variables to have different d.o.f. parameters, thus has stochastic representation

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

In Figures 1 and 2 we illustrate the differences between the Riesz, t-Riesz, and F-Riesz. Figure 1 depicts the marginal p.d.f.s of the first RV ($\mathbf{R}_{11}$) implied by each distribution for the parameter setting $\boldsymbol{\Sigma} = \mathbf{I}_2$, $\mathbf{n} = (15, 30)^\top$, with three different settings for the d.o.f. parameters, $\nu = 5, 6, 10$ (t-Riesz) and, correspondingly $\boldsymbol{\nu} = (5, 5)^\top, (6, 6)^\top, (10, 10)^\top$ (F-Riesz). These parameter values are inspired by our estimation results presented in Section 5 below. We see that the F-Riesz and t-Riesz distributions feature similar marginal distributions for the RV, especially for larger ν. This is not surprising, as the two distributions converge to the same Riesz distribution as ν goes to infinity (see Section 2.5). The F-Riesz and t-Riesz distributions compared to the Riesz have fatter tails, which become smaller with increasing ν. Furthermore, the F-Riesz and t-Riesz have more probability mass on small RVs than the Riesz. In Figure 2 we take the setting with $\nu = 5$ (t-Riesz) and $\boldsymbol{\nu} = (5, 5)^\top$ (F-Riesz) from above and plot the marginal joint distribution of the two RVs ($\mathbf{R}_{11}$, $\mathbf{R}_{22}$). The plots look qualitatively the same for different settings of ν. We first note that the Riesz distribution is drastically different in

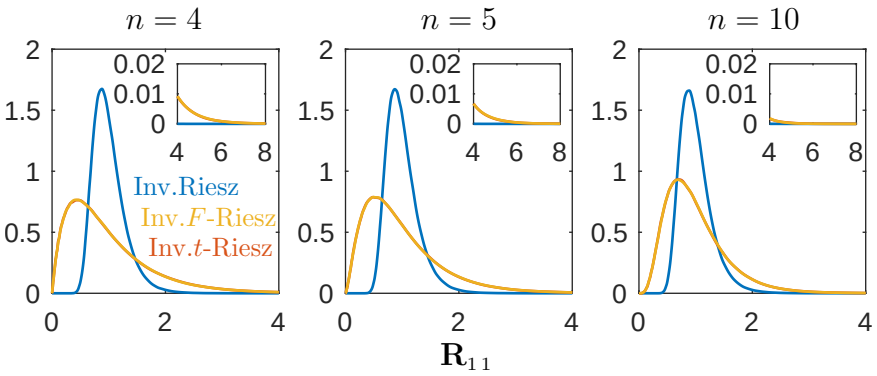

Figure 3: Marginal p.d.f.s of the first RV ($\mathbf{R}_{11}$) implied by the Inverse Riesz, Inverse t-Riesz, and Inverse F-Riesz distributions for a two-dimensional RC with the parameter setting $\boldsymbol{\Sigma} = \mathbf{I}_2$, $\boldsymbol{\nu} = (30, 15)^\top$, and with three different settings for the d.o.f. parameters, $n = 4, 5, 10$ (Inverse t-Riesz) and $\mathbf{n} = (4, 4)^\top, (5, 5)^\top, (10, 10)^\top$ (Inverse F-Riesz).

that most of its probability mass is centered around the value of one for both RVs but the probability mass is less peaked and is spread out more (elliptically around the coordinate (1,1)) than for the other two distributions. Comparing subplot one and two, the stark differences between the t-Riesz and F-Riesz distribution now become visible. The t-Riesz distributions' probability mass lies in an elliptical shape around the diagonal from bottom right to top left with a high peak in probability mass on values between 0 and 0.5 for both RVs. This shape implies that the two RVs are more likely to have similar values than drastically different ones for any size of RV realization, which is in line with our analysis of the stochastic representation kernels above. In contrast, the F-Riesz probability mass fans out in a triangular fashion from the bottom left corner with more probability mass along the coordinate axes than around the bottom-left to top-right diagonal. It also peaks on values between 0 and 0.5, but the peak is less pronounced than for the t-Riesz and for these small values the probability mass is much more spread out. We see for the F-Riesz, that if one RV is small the probability of the other being small is not nearly as high as for the t-Riesz and there lies substantial probability mass on large realizations for the other RV. Conversely, if one RV has a tail realization the other is more likely not to have one, a property Blasques et al. (2021) call *tail-heterogeneity*. Consequently we call the t-Riesz distribution *tail-homogeneous*. The correlation between the two RVs is 0 for the Riesz, 0.24 for the F-Riesz distribution and 0.87 for the t-Riesz. All these observations are in line with the

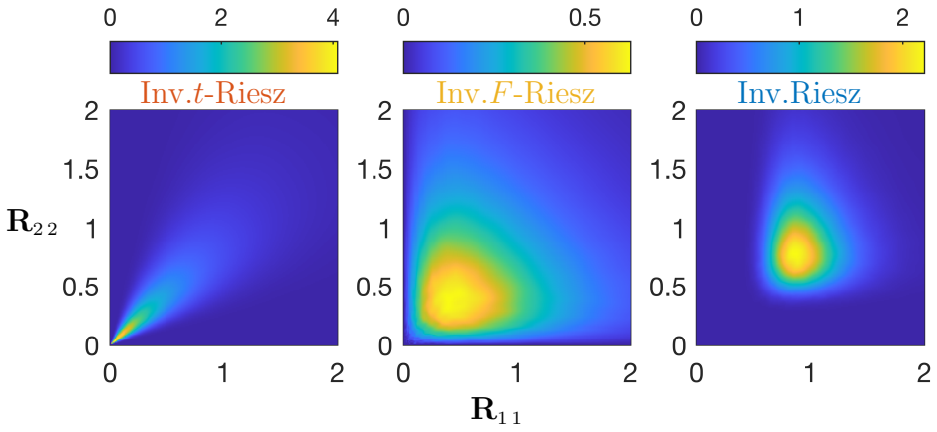

Figure 4: Heatmap of marginal joint p.d.f. of the first RV and second RV ($\mathbf{R}_{11}$, $\mathbf{R}_{22}$) implied by the Inverse Riesz, Inverse t-Riesz, and Inverse F-Riesz distributions for a two-dimensional RC with the parameter setting $\mathbf{\Sigma} = \mathbf{I}_2$, $\boldsymbol{\nu} = (30, 15)^\top$, $n = 5$ (Inverse t-Riesz), $\mathbf{n} = (5, 5)^\top$ (Inverse F-Riesz).

intuition we gained above from analyzing the stochastic representation kernels. We deduce that its plausible that the t-Riesz distribution family would perform better during market-wide crises where all assets experience high volatility, whereas the F-Riesz distribution could offer benefits when only a particular asset or subsections of the market experience distress.

To complete the analysis of all distribution, we still have to consider the inverse versions of the distributions mentioned so far. Unfortunately, it is difficult to gain an intuition of the stochastic properties of inverse distributions since, due to the inversion of the Bartlett matrices, the marginal distributions of the elements in $\boldsymbol{\mathcal{K}}_{\mathcal{D}}$ and their dependencies are not easily derived. However, we can visualize the same marginal distributions as we did above for non-inverted distributions. Figure 3 plots the marginal p.d.f.s of the first RV ($\mathbf{R}_{11}$) implied by the Inverse Riesz, Inverse t-Riesz, and Inverse F-Riesz distributions for the parameter setting $\mathbf{\Sigma} = \mathbf{I}_2$, $\boldsymbol{\nu} = (30, 15)^\top$, with three different settings for the d.o.f. parameters, $n = 5, 6, 10$ (Inverse t-Riesz) and, correspondingly $\mathbf{n} = (5, 5)^\top, (6, 6)^\top, (10, 10)^\top$ (Inverse F-Riesz). These parameter values are again inspired by our estimation results below. We see that the marginal distribution of the RV is almost identical for the Inverse F-Riesz and Inverse t-Riesz distributions. The Inverse F-Riesz and Inverse t-Riesz distributions have is more skewed and has fatter tails than the Inverse Riesz, which become smaller with increasing n. It is also noteworthy, that the Inverse Riesz,

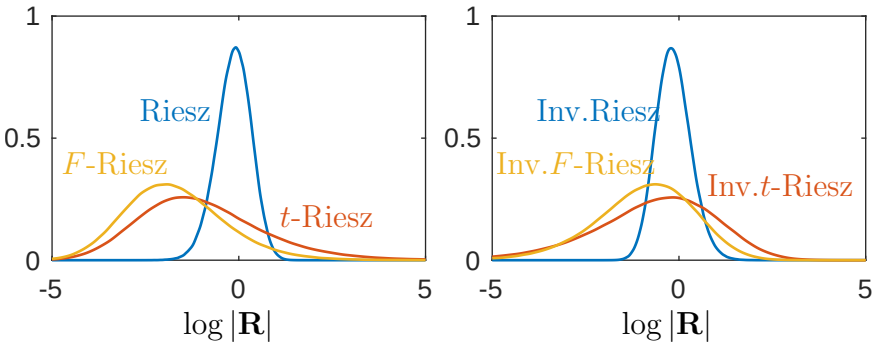

Figure 5: Marginal p.d.f. of the log-determinant of the RC for the (Inverse) Riesz, (Inverse) t-Riesz, and (Inverse) F-Riesz distributions for a two-dimensional RC with the parameter setting $\mathbf{\Sigma} = \mathbf{I}_2$, and for the non-inverted distributions, $\mathbf{n} = (15, 30)^\top$, $\nu = 5$ (t-Riesz), $\boldsymbol{\nu} = (5, 5)^\top$ (F-Riesz), and for the inverse distributions, $\boldsymbol{\nu} = (30, 15)^\top$, $n = 5$ (Inverse t-Riesz), $\mathbf{n} = (5, 5)^\top$ (Inverse F-Riesz).

compared to the Riesz in Figure 1, has larger tails. Even though the Inverse F-Riesz and Inverse t-Riesz have almost identical marginal distributions for the RV, their joint marginal distributions of the first and second RV ($\mathbf{R}_{1\,1}$ and $\mathbf{R}_{2\,2}$) differ substantially, as we can see in Figure 4. In particular, as for the t-Riesz, the Inverse t-Riesz's probability mass is very concentrated around the lower-left to upper-right diagonal with a very high peak on values below 0.3 for both RVs. This implies that $\mathbf{R}_{1\,1}$ and $\mathbf{R}_{2\,2}$ are highly correlated. Similarly to the non-inverted version, Inverse F-Riesz's probability mass is spread out in a triangular fashion with a much lower peak than the Inverse t-Riesz. The correlation between $\mathbf{R}_{1\,1}$ and $\mathbf{R}_{2\,2}$ is 0.70, 0.017, and 0.023 for the Inverse t-Riesz, Inverse F-Riesz, and Inverse Riesz, respectively. It is noteworthy, that the Inverse F-Riesz, compared to the F-Riesz in Figure 2, is more spread out and features much lower correlation between the RVs.

2.3 Fat-Tailedness

In the literature, fat-tailedness of RCs has been measured by or considered synonymous to fat-tailedness of their diagonal elements, i.e. the RVs (c.f. Opschoor et al. 2018 and Blasques et al. 2021). We extend this interpretation by taking the log-determinant of the RCs as the quantity to determine their fat-tailedness because the determinant of a matrix can be geometrically interpreted as its vol-

ume. Naturally, we call those matrix-variate distributions fat-tailed that imply fat-tailed marginal distributions for the log-determinant of their random matrices. Conveniently, for random matrices from any of our considered distribution for RCs, the log-determinant equals the sum of log-χ random variables. This can be seen from the stochastic representations (Table 1). For example, for the F-Riesz distribution, the stochastic representation of the log-determinant, omitting the subscript t, is equal to

$$\log|\mathbf{R}| = \log|\mathbf{C}_\Omega \bar{\mathbf{B}}^{-\top} \underline{\mathbf{B}} \underline{\mathbf{B}}^\top \bar{\mathbf{B}}^{-1} \mathbf{C}_\Omega^\top| \tag{15}$$

$$= 2\sum_{i=1}^{p} \log(\mathbf{C}_\Omega)_{ii} + 2\sum_{i=1}^{p} \log(\chi_{n_i-i+1}) - 2\sum_{i=1}^{p} \log(\chi_{\nu_i-p+i}), \tag{16}$$

and for the t-Riesz distribution it is

$$\log|\mathbf{R}| = \log|\mathbf{C}_\Omega (\bar{b})^{-2} \underline{\mathbf{B}} \underline{\mathbf{B}}^\top \mathbf{C}_\Omega^\top| \tag{17}$$

$$= 2\sum_{i=1}^{p} \log(\mathbf{C}_\Omega)_{ii} + 2\sum_{i=1}^{p} \log(\chi_{n_i-i+1}) - 2\,p \log(\chi_\nu). \tag{18}$$

We can then formalize the concept of fait-tailedness by considering a distribution $\mathcal{D}_1$ to be more fat-tailed than another, $\mathcal{D}_2$, if there exists some real number x, such that for all real $y > x$, the p.d.f.s are such that $p_{\log|\mathbf{R}|,\mathcal{D}_1}(y) > p_{\log|\mathbf{R}|,\mathcal{D}_2}(y)$. For the case of 1×1 random matrices, i.e. random variables, this definition corresponds to the classical definition of fat-tailedness for random variables.

In Figure 5 we plot the marginal p.d.f.s of the log-determinant implied by the Riesz-type distributions. We see, that the (Inverse) F-Riesz and (Inverse) t-Riesz distributions can indeed be considered fat-tailed, as they feature fatter right tails than the (Inverse) Riesz distribution. It is noteworthy that the t-Riesz distribution features a fatter right tail than the F-Riesz.

2.4 Asset Ordering

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 $\mathbf{R} \sim \mathcal{D}(\mathbf{\Omega}, \boldsymbol{\theta}_\mathcal{D})$, then any other ordering the assets, $\mathbf{PRP}^\top$, where $\mathbf{P}$ denotes an

arbitrary permutation matrix[9], has stochastic representation $\mathbf{P}\mathbf{C}_\Omega\boldsymbol{\mathcal{K}}_\mathcal{D}(\mathbf{C}_\Omega)^\top\mathbf{P}^\top$.[10] This new stochastic representation yields *for Riesz-type distributions* a different and generally unknown probability distribution, which we denote by $\mathcal{D}_\mathbf{P}(\boldsymbol{\Omega}, \boldsymbol{\theta}_\mathcal{D})$. Only for the special case where $\mathbf{R}$ is the exchange matrix we do know the resulting probability distributions (see next Section).

In practice, we are given a randomly ordered RC $\mathbf{PRP}^\top$ for which the model $\mathcal{D}(\boldsymbol{\Omega}, \boldsymbol{\theta}_\mathcal{D})$ 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 it as a parameter to optimize over. Blasques et al. (2021) have proposed an efficient algorithm fro maximizing the likelihood for many different orderings[11] and then choosing the ordering with the highest estimated likelihood value. In a simulation experiment, they find that their algorithms' 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.

2.4.1 Riesz Distribution Versions

For any Riesz-type distribution there are two versions; we call them version-I and version-II (c.f. Blasques et al. 2021). In this subsection we contribute to the literature by showing that the two versions are closely related since assuming $\mathbf{R}$ follows a version-I Riesz distribution is equivalent to assuming that $\mathbf{R}$ with the asset order reversed, denoted by $\overleftarrow{\mathbf{R}}$, follows the corresponding version-II distribution with "reversed" parameters. See Table 4 for the exact relationships.

To derive the equivalence we have to introduce some concepts. First note that the permutation matrix which achieves a reversal of the asset order in $\mathbf{R}$ is the is the exchange matrix, i.e. a matrix with ones on the diagonal from the upper

9. A permutation matrix is a square matrix that has exactly one entry of one in each row and each column and zeros elsewhere.

10. Recall the stochastic representation of $\mathbf{R}$ is $\mathbf{R} = \mathbf{C}_\Omega\boldsymbol{\mathcal{K}}_\mathcal{D}(\mathbf{C}_\Omega)^\top$.

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

Our versions		Alternative versions
$\mathbf{R} \sim \mathcal{R}^{I}(\mathbf{\Omega}, \mathbf{n})$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim \mathcal{R}^{II}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\mathbf{n}})$
$\mathbf{R} \sim i\mathcal{R}^{II}(\mathbf{\Omega}, \boldsymbol{\nu})$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim i\mathcal{R}^{I}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\boldsymbol{\nu}})$
$\mathbf{R} \sim t\mathcal{R}^{I}(\mathbf{\Omega}, \mathbf{n}, \nu)$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim t\mathcal{R}^{II}(\overleftarrow{\mathbf{\Omega}}, \nu, \overleftarrow{\mathbf{n}})$
$\mathbf{R} \sim it\mathcal{R}^{II}(\mathbf{\Omega}, n, \boldsymbol{\nu})$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim it\mathcal{R}^{I}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\boldsymbol{\nu}}, n)$
$\mathbf{R} \sim F\mathcal{R}^{I}(\mathbf{\Omega}, \mathbf{n}, \boldsymbol{\nu})$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim F\mathcal{R}^{II}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\boldsymbol{\nu}}, \overleftarrow{\mathbf{n}})$
$\mathbf{R} \sim iF\mathcal{R}^{II}(\mathbf{\Omega}, \mathbf{n}, \boldsymbol{\nu})$	$\Leftrightarrow$	$\overleftarrow{\mathbf{R}} \sim iF\mathcal{R}^{I}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\boldsymbol{\nu}}, \overleftarrow{\mathbf{n}})$

Table 4: Equivalence between the Riesz distribution versions. In this paper we choose version-I for the non-inverted distribution and version-II for the inverse ones. Outside of this subsection we do not refer to the different versions and consequently drop the version superscripts for better readability.

right-hand corner to the lower left-hand corner and zeros elsewhere,

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

We can visualize $\overleftarrow{\mathbf{R}} := \mathbf{P}_e \mathbf{R} \mathbf{P}_e^\top$ as the original matrix rotated by 180 degrees. Furthermore, note that $\mathbf{P}_e \bar{\mathbf{B}} \mathbf{P}_e$ is equal to $\underline{\mathbf{B}}$ but with degrees of freedom $\overleftarrow{\boldsymbol{\nu}} = (\nu_p, \nu_{p-1}, \ldots, \nu_1)$ instead of $\mathbf{n}$. We denote it by $\underline{\mathbf{B}}_{\overleftarrow{\boldsymbol{\nu}}}$. Similarly, $\mathbf{P}_e \underline{\mathbf{B}} \mathbf{P}_e$ is equal to $\bar{\mathbf{B}}$ but with $\overleftarrow{\mathbf{n}}$ instead of $\boldsymbol{\nu}$, denoted by $\bar{\mathbf{B}}_{\overleftarrow{\mathbf{n}}}$. Next, note that $\mathbf{P}_e = \mathbf{P}_e^\top$, and $\mathbf{P}_e \mathbf{P}_e = \mathbf{I}$, such that $\mathbf{P}_e = \mathbf{P}_e^{-1}$. Finally, note that we can write the reverse order $\mathbf{\Omega}$ as

$$\overleftarrow{\mathbf{\Omega}} = \mathbf{P}_e \mathbf{\Omega} \mathbf{P}_e = \mathbf{P}_e \mathbf{C}_\Omega \mathbf{P}_e \mathbf{P}_e \mathbf{C}_\Omega^\top \mathbf{P}_e = \mathbf{U}_{\overleftarrow{\mathbf{\Omega}}} \mathbf{U}_{\overleftarrow{\mathbf{\Omega}}}^\top,$$

where $\mathbf{P}_e \mathbf{C}_\Omega \mathbf{P}_e = \mathbf{U}_{\overleftarrow{\mathbf{\Omega}}}$ is the upper Cholesky factor of $\overleftarrow{\mathbf{\Omega}}$, since the decomposition of a symmetric positive definite matrix into an upper triangular matrix post-multiplied by its transpose, i.e. the Cholesky decomposition, is unique.

We now show the equivalence between the two versions using the example of the F-Riesz distribution. Assume that $\mathbf{R}$ follows our version-I F-Riesz distribution,

$\mathbf{R} \sim F\mathcal{R}(\mathbf{\Omega}, \mathbf{n}, \boldsymbol{\nu})$. Then

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

which is the stochastic representation of the version-II F-Riesz distribution with parameters $\overleftarrow{\mathbf{\Omega}}$, $\overleftarrow{\boldsymbol{\nu}}$, and $\overleftarrow{\mathbf{n}}$, $F\mathcal{R}^{II}(\overleftarrow{\mathbf{\Omega}}, \overleftarrow{\boldsymbol{\nu}}, \overleftarrow{\mathbf{n}})$ (Theorem 8 in Blasques et al. 2021). We can derive the distributions of $\overleftarrow{\mathbf{R}}$ for all other Riesz-type distributions in similar fashion. In general, the alternative versions not used in this paper (right column of Table 4) have stochastic representation kernels where $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$ (and their special cases) are interchanged.

Empirically it does not matter which distribution version we assume since, as we mentioned above, we take the asset order as a parameter to optimize over. That is, assuming version I (II) and estimating it via maximum likelihood should yield the "reversed" estimates for the asset order and the parameter values as compared to assuming version II (I).

2.4.2 Inverse Distribution Versions

As you can see in Table 4, in this paper we choose the version-II distributions for the inverted ones and the version-I distributions for the non-inverted ones. In this way our stochastic representations are all defined in terms of the lower Cholesky factor of $\mathbf{\Omega}$ rather than its upper Cholesky factor (as for the version-I distributions for the inverted ones and the version-II distributions for the non-inverted ones). To see this, note that assuming $\mathbf{R}$ follows a version-I (version-II) inverse distribution with parameter matrix $\mathbf{\Omega}^{-1}$ is equivalent to assuming that $\mathbf{R}^{-1}$ follows the corresponding version-I (version-II) non-inverted distribution with parameter matrix $\mathbf{\Omega}$. This is how the inverse distributions are defined (e.g. Blasques et al. 2021). As an example, let's take a version-I Riesz distribution, $\mathbf{R} \sim \mathcal{R}(\mathbf{\Omega}^{-1}, \mathbf{n})$, and use the stochastic representation to derive its inverse, the version-I Inverse

Riesz distribution, as defined above.

$$\begin{aligned}\mathbf{R} &= \mathbf{C}_{\boldsymbol{\Omega}^{-1}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathbf{C}_{\boldsymbol{\Omega}^{-1}}^{\top} \\ \Leftrightarrow \mathbf{R}^{-1} &= \mathbf{C}_{\boldsymbol{\Omega}^{-1}}^{-\top}\underline{\mathbf{B}}^{-\top}\underline{\mathbf{B}}^{-1}\mathbf{C}_{\boldsymbol{\Omega}^{-1}}^{-1}.\end{aligned}$$

The important thing to note here is, that the stochastic representation of the version-I Inverse Riesz distribution features $\mathbf{C}_{\boldsymbol{\Omega}^{-1}}^{-\top} = \mathbf{U}_{\boldsymbol{\Omega}}$, which is the *upper* Cholesky factor of $\boldsymbol{\Omega}$. In fact, we can derive in similar fashion, that the stochastic representations of all inverse version-I distributions feature the upper Cholesky factor of $\boldsymbol{\Omega}$, while all inverse version-II distributions use the lower Cholesky factor. Thus, in order to have *one* general stochastic representation formula based on the lower Cholesky factor that is valid for all distributions in our paper, i.e. equation (4) ($\mathbf{R} = \mathbf{C}_{\Omega}\boldsymbol{\mathcal{K}}_{\mathcal{D}}\mathbf{C}_{\Omega}^{\top}$), and since empirically it does not matter which distribution version we choose (see above), we use the version-II for the inverse and version-I for the non-inverted ones.

2.5 Distribution Relationships

Figure 6 shows how the distributions are related to each other. Every Wishart-type distribution is a special case of its Riesz-type counterpart and is obtained by setting the entries in each d.o.f. parameter vector all equal to each other. This is easily seen from the stochastic representations in Table 1, where the Wishart-type distributions have the same stochastic representation kernel as their Riesz-type counterparts but with $\underline{\mathcal{B}}$ and $\bar{\mathcal{B}}$ instead of $\underline{\mathbf{B}}$ and $\bar{\mathbf{B}}$. The proofs for the dashed arrows are also immediately evident from the stochastic representations since 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 one of the (Inverse) t-Riesz distribution. We derive the remaining relationships in the following Theorem 2.2.

Theorem 2.2.

$$\mathbf{C}\mathbf{M}_{F\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{F\mathcal{R}}\mathbf{M}_{F\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \xrightarrow[\forall i:\nu_i\to\infty]{d} \mathbf{C}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{R}}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top}$$

$$\mathbf{C}\mathbf{M}_{iF\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{iF\mathcal{R}}\mathbf{M}_{iF\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \xrightarrow[\forall i:n_i\to\infty]{d} \mathbf{C}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{i\mathcal{R}}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top}$$

$$\mathbf{C}\mathbf{M}_{t\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{t\mathcal{R}}\mathbf{M}_{t\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \xrightarrow[\nu\to\infty]{d} \mathbf{C}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{\mathcal{R}}\mathbf{M}_{\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top}$$

$$\mathbf{C}\mathbf{M}_{it\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{it\mathcal{R}}\mathbf{M}_{it\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top} \xrightarrow[n\to\infty]{d} \mathbf{C}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\boldsymbol{\mathcal{K}}_{i\mathcal{R}}\mathbf{M}_{i\mathcal{R}}^{-\frac{1}{2}}\mathbf{C}^{\top}.$$

Proof in Appendix.

As they are just special cases, Theorem 2.2 also holds for the corresponding Wishart-type distributions ($F \to \mathcal{W}$ or $i\mathcal{W}$, $t\mathcal{W} \to \mathcal{W}$, $it\mathcal{W} \to i\mathcal{W}$). Finally, note that an Inverse F distribution is again an F distribution with the degrees of freedom parameters switched and the expected value matrix inverted, as is easy to see from their stochastic representations. However, an Inverse F-Riesz distribution is not again an F-Riesz distribution (see Section 7.2 in the appendix). We derive the novel Inverse F-Riesz distribution in this paper.

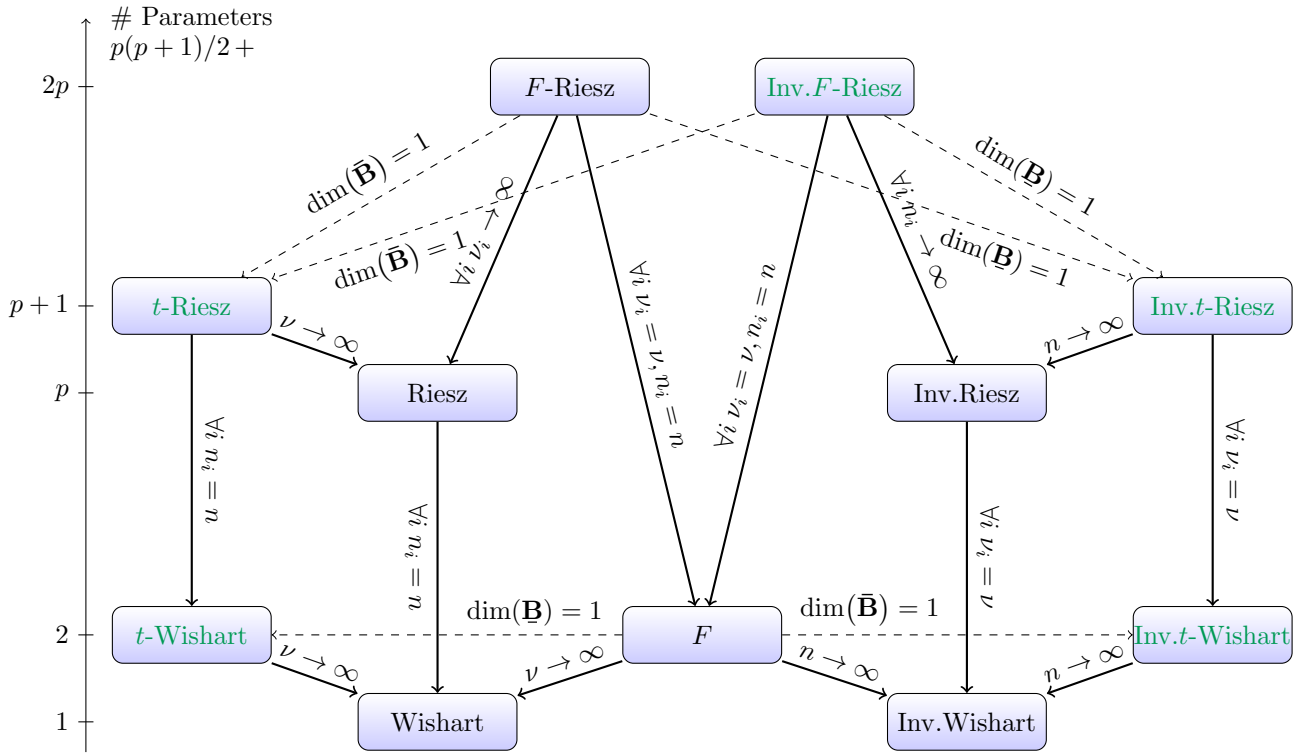

Figure 6: Relationships between the standardized probability distributions. Stochastic representations given by equation (7) in conjunction with Tables 1 and 3. Next to every arrow, we indicate how the distributions are related. Dashed arrows indicate a more general relationship; not a nesting. The vertical axis shows the number of parameters of the respective distributions. All distributions have $p(p+1)/2$ distinct parameters in the symmetric positive definite parameter matrix $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) plus the number of d.o.f. parameters.

2.6 Probability Density Functions

Let us define some special functions used in the probability density functions.

Definition 2.1 (Generalized Power Function)**.** *Let* $\mathbf{X}$ *be a real* $p \times p$ *matrix,* $\mathbf{n} = (n_1, \ldots, n_p)^\top$ *be a real vector of length* p *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*

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

The generalized power function is defined in e.g. Faraut and Korányi (1994). The determinant-with-subscript notation is taken from Blasques et al. (2021). It makes immediately visible the close relation of the generalized power function to the determinant since for $n_1 = n_2 = \ldots = n_p = n$, $|\mathbf{X}|_\mathbf{n} = |\mathbf{X}|^n$. The next lemma shows that in the case of positive definite $\mathbf{X}$, the generalized power function can be written as a function of the diagonal elements of the lower Cholesky decomposition of $\mathbf{X}$. Blasques et al. (2021) name this special case *power weighted determinant*.

Lemma 2.1 (Power Weighted Determinant)**.** *Let* $\mathbf{\Sigma}$ *be positive definite and* $\mathbf{\Sigma} = \mathbf{T}\mathbf{D}\mathbf{T}^\top$ *be the unique decomposition into a 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.

The next lemma lists algebraic equalities for the power weighted determinant.

Lemma 2.2 (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 $|\mathbf{R}|_{n\cdot\mathbf{1}} = |\mathbf{R}|^n$*

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

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

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. Then the vector-valued multivariate gamma function is defined as*

$$\Gamma_p(\mathbf{n}) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(n_i - \frac{i-1}{2}\right),$$

with $2n_i > i - 1$, $i = 1, \ldots p$.

Let n be a scalar. Then the scalar-valued multivariate gamma function is defined as

$$\Gamma_p(n) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(n - \frac{i-1}{2}\right),$$

with $2n > p - 1$.

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

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

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

Table 5: Probability density functions of distributions for RCs. Derivations from stochastic representations are in Section 7.2 (appendix). $\bar{\mathbf{n}} = \frac{1}{p}\sum_{i=1}^{p} n_i$.

Now that we have introduced all special functions used in the probability density functions (p.d.f.s), we list the p.d.f.s for all considered probability distributions in Table 5. 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). See Section 7.2 in the appendix for more details.

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

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

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

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

Proof in Appendix.

The p.d.f.s of the Wishart-type distributions follow by simply setting all entries

in the d.o.f. vectors equal to each other. The p.d.f.s of the standardized probability distributions, which we indicate by writing $p_{\mathcal{D}}(\boldsymbol{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}})$ instead of $p_{\mathcal{D}}(\boldsymbol{\Omega}, \boldsymbol{\theta}_{\mathcal{D}})$, are given in the appendix in Table 11. They are easily derived by replacing $\boldsymbol{\Omega}$ with $\mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}^{\top}$.[12]

3 The t-Riesz Distribution Family Based on Intraday Return Vectors

This section shows how our novel t-Riesz distribution family arises naturally as the distribution of the standard realized covariance matrix under a set of assumptions on the underlying intraday return vectors. To be precise, let us add subscripts for the days in our sample, $t = 1, \ldots, T$, to represent the time-series of RCs as $\{\mathbf{R}_t\}_{t=1}^{T}$. Each day is split into intraday sub-periods of equal length, $j = 1, \ldots, m$, to represent the j'th intraday return vector on day t as $\mathbf{r}_{t,j} = (r_{t,j,1}, \ldots, r_{t,j,p})^{\top}$. The day-$t$ standard realized covariance matrix is then defined as

$$\mathbf{R}_t = \sum_{j=1}^{m} \mathbf{r}_{t,j}\mathbf{r}_{t,j}^{\top}. \tag{19}$$

So $\mathbf{R}_t$ is almost surely positive definite if $m \geq p$.[13]

We denote the $mp \times 1$ vector of all intraday returns on day t as $\tilde{\mathbf{r}}_t = (\mathbf{r}_{t,1}^{\top}, \mathbf{r}_{t,2}^{\top}, \ldots, \mathbf{r}_{t,m}^{\top})^{\top}$, and the vector that collects all intraday returns as $\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_1^{\top}, \ldots, \tilde{\mathbf{r}}_T^{\top})^{\top}$. We start from the very general assumption that all intraday returns on all days jointly follow a multivariate elliptically contoured distribution[14] with zero mean vector and block diagonal scale matrix $\tilde{\boldsymbol{\Omega}} = (\mathbf{I}_{(Tm)} \otimes \boldsymbol{\Omega})$,

$$\tilde{\mathbf{r}} \sim \mathcal{E}(\mathbf{0}, \mathbf{I}_{(Tm)} \otimes \boldsymbol{\Omega}, \zeta), \tag{20}$$

where the exact functional form of $\zeta : [0, \infty) \to \mathbb{R}$ determines the specific elliptical distribution. The block diagonal scale matrix implies that there is no correlation

12. See Section 7.2 in the appendix for more details.
13. As mentioned in Section 5.2 below, we effectively make use of 385 intraday return vectors for each day by employing the 5-minute realized covariance matrix subsampled every minute of the trading day. Thus positive definiteness is given up to $p = 385$.
14. as defined in Definition 2.1 of Gupta, Varga, and Bodnar 2013

between the intraday returns of different intraday intervals, but there is correlation between those on the same intraday interval j. The joint distribution of all realized covariance matrices $\{\mathbf{R}_t\}_{t=1}^T$ implied by the above assumption on $\tilde{\mathbf{r}}$ is given by Theorem 3.1.

Theorem 3.1. *(Gupta, Varga, and Bodnar 2013) Let* $\tilde{\mathbf{r}} = (\tilde{\mathbf{r}}_1^\top, \ldots, \tilde{\mathbf{r}}_T^\top)^\top$ *follow a multivariate elliptically contoured distribution with zero mean vector,* $Tmp \times Tmp$ *scale matrix* $\tilde{\mathbf{\Omega}} = (\mathbf{I}_{(Tm)} \otimes \mathbf{\Omega})$ *and p.d.f.*

$$f(\tilde{\mathbf{r}}) = |\mathbf{\Omega}|^{-Tm/2} h(\tilde{\mathbf{r}}^\top \tilde{\mathbf{\Omega}}\, \tilde{\mathbf{r}}). \tag{21}$$

Then the joint p.d.f. of $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_T)$*, where* $\mathbf{R}_t = \sum_{j=1}^m \mathbf{r}_{t,j}\mathbf{r}_{t,j}^\top$*, obtains as*

$$\frac{\pi^{\frac{Tmp}{2}}}{\Gamma_p\left(\frac{m}{2}\right)^T} \prod_{t=1}^T |\mathbf{R}_t|^{\frac{m-p-1}{2}} f\left(tr\left(\sum_{t=1}^T \mathbf{R}_t \mathbf{\Omega}^{-1}\right)\right). \tag{22}$$

The marginal p.d.f. of $\mathbf{R}_t$ *is given by*

$$\frac{\pi^{\frac{mp}{2}}}{\Gamma_p\left(\frac{m}{2}\right)} |\mathbf{R}_t|^{\frac{m-p-1}{2}} f\left(tr\left(\mathbf{R}_t \mathbf{\Omega}^{-1}\right)\right). \tag{23}$$

[a] *Proof in Appendix.*

a. Note that these distributions for $\mathbf{R}_t$ are also defined if m is a real, rather than natural number (in the previous section we called this real parameter n).

For the special case, where $\tilde{\mathbf{r}}$ is follows a multivariate normal distribution, the marginal distribution of a specific $\mathbf{R}_t$ (equation 23) is a Wishart distribution $\mathbf{R}_t \sim \mathcal{W}(\mathbf{\Omega}, m)$. In fact, in the Wishart case, the joint distribution of all $\mathbf{R}_t$ given derived from equation (39) is equal to the product of their marginal distributions, which implies independence, $\mathbf{R}_t \overset{iid}{\sim} \mathcal{W}(\mathbf{\Omega}, m)$. However, the normality assumption for financial return vectors is strongly rejected by the data. Furthermore, the multivariate normality assumption with block-diagonal covariance matrix $\tilde{\mathbf{\Omega}}$ implicitly assumes that the individual intraday return vectors $\mathbf{r}_{t,j}$ are independent of each other also within each day, which is a very strong assumption. A more realistic multivariate elliptically contoured distribution is the multivariate t-distribution, as it accommodates the fat tails observed in financial return data. Although due to the block-diagonal structure of $\tilde{\mathbf{\Omega}}$ there is no correlation between the $\mathbf{r}_{t,j}$, the

multivariate t distribution does imply dependence between them. For this special case, i.e. $\tilde{\mathbf{r}} \sim mvt(\mathbf{0}, \tilde{\mathbf{\Omega}}, \nu)$, the marginal distributions correspond to what we call the t-Wishart distribution, denoted by $\mathbf{R}_t \sim t\mathcal{W}(\mathbf{\Omega}, m, \nu)$. For $t_1 \neq t_2$, $\mathbf{R}_{t_1}$ and $\mathbf{R}_{t_2}$ are dependent.[15] The superior performance of the t-Wishart distribution over the Wishart (as documented in Section 5) indicates that the assumption of fat-tails and dependence between the intraday return vectors (as implied by the multivariate t-distribution) is more realistic.[16]

Unfortunately the assumption of a joint distribution with the simple block-diagonal scale matrix for all intraday return vectors is unrealistic. Researchers have documented that the covariance matrix of asset returns varies over time (volatility clustering), which is at odds with the assumption that every intraday return vectors' marginal distribution has the same scale matrix $\mathbf{\Omega}$ as implied by the block-diagonal structure. Furthermore, we only observe returns during market hours. Thus, the equally spaced grid of intraday returns that *connects* across days is not a realistic assumption. For these reasons, we relax the assumption of a joint distribution for *all* intraday returns and assume instead that only those intraday returns on the *same* day follow jointly an elliptically contoured distribution, with a time-varying scale matrix, conditional on all previous intraday returns.

$$\tilde{\mathbf{r}}_t | \mathcal{F}_{t-1} \sim \mathcal{E}(\mathbf{0}, \mathbf{I}_{(m)} \otimes \mathbf{\Omega}_t, \zeta), \tag{24}$$

where $\mathcal{F}_{t-1} = \{\tilde{\mathbf{r}}_{t-1}, \tilde{\mathbf{r}}_{t-2}, \dots\}$. By Theorem 3.1[17], this assumption implies that $\mathbf{R}_t | \mathcal{F}_{t-1}$ follows a conditional distribution with time-varying scale matrices and p.d.f.

$$\frac{\pi^{\frac{mp}{2}}}{\Gamma_p\left(\frac{m}{2}\right)} |\mathbf{R}_t|^{\frac{m-p-1}{2}} f\left(tr\left(\mathbf{R}_t \mathbf{\Omega}_t^{-1}\right)\right). \tag{25}$$

So, for example, if we assume an intraday joint t-distribution (normal distribution)

15. Another version of the t-Wishart was first introduced by Sutradhar and Ali (1989).

16. We also tried the multivariate hyperbolic distribution (leading to the hyperbolic-Wishart), which is a generalization of the multivariate t distribution and the multivariate Laplace distribution (leading to the Laplace-Wishart). In both a low-dimensional and a high-dimensional estimation there were no substantial likelihood gains over the t-Wishart. The hyperbolic-Wishart numerical maximum likelihood parameter estimates converged to almost exactly its special case, the t-Wishart distribution.

17. Set $T = 1$, i.e. chooose one specific day only.

for $\tilde{\mathbf{r}}_t$, this implies a t-Wishart (Wishart) distribution with a time-varying scale matrix,

$$\mathbf{R}_t|\mathcal{F}_{t-1} \sim t\mathcal{W}(\mathbf{\Omega}_t, m, \nu). \tag{26}$$

To summarize, we assume a static model for the intraday return vectors of a given day t (elliptical distribution with static scale matrix $\mathbf{I}_{(m)} \otimes \mathbf{\Omega}_t$), but a dynamic model across days (the scale matrix varies from day to day).

Next, we are going to derive the t-Riesz distribution under the same assumption of a joint multivariate t-distribution of the intraday returns of a given day, but with the notion of asset liquidity included.[18] We measure liquidity for a given asset by the number of intraday intervals where at least one trade occurred, and thus a new price observation was recorded. Only those intraday returns in $\tilde{\mathbf{r}}_t$ for which there was a new price observation are assumed to jointly follow a multivariate t-distribution, while the others are replaced by zeros.[19]

In fact, zero returns are far too frequent in financial asset return data for the assumption of a continuous distribution for returns (like the t-distribution) to be realistic. This was shown by e.g. Sucarrat and Grønneberg (2020). Excluding returns that are zero due to missing price observations makes the assumption of a continuous distribution more realistic. The treatment of missing new price observations as deterministic zeros was proposed by Gribisch and Hartkopf (2022) and Hassairi, Ktari, and Zine (2022), who derive the Riesz distribution from the assumption of a joint normal distribution on all intraday returns of a given day t.

We can represent the missing price observations mathematically by using the stochastic representation of a multivariate t-distribution, $\tilde{\mathbf{r}}_t = \sqrt{y_t}\mathbf{C}_{\tilde{\mathbf{\Omega}}_t}\tilde{\mathbf{z}}_t$, where $y_t \sim \Gamma(\nu/2, 2/\nu)$ and with $\tilde{\mathbf{z}}_t = (\mathbf{z}_{t,1}^\top, \mathbf{z}_{t,2}^\top, \ldots, \mathbf{z}_{t,m}^\top)^\top$, $\mathbf{z}_{t,j} = (z_{t,j,1}, z_{t,j,2}, \ldots, z_{t,j,p})^\top$, and the innovation on day t on intraday-return interval j of asset i given by

$$z_{t,j,i} = \begin{cases} 0 \text{ if there is a missing observation,} \\ \overset{iid}{\sim} \mathcal{N}(0,1) \text{ else.} \end{cases}$$

18. Recall that the t-distribution is a generalization of the t-Wishart distribution with d.o.f. parameter vector $\mathbf{n}$ instead of scalar n, where the two distributions are equal if $\mathbf{n} = (n, \ldots, n)^\top$.

19. Since there was no new price observation, previous-tick interpolation would make these returns zero anyway.

Note that this is equivalent to saying that only those intraday returns follow jointly a multivariate t-distribution for which there exists a new price observation in the corresponding intraday interval.

Due to the block-diagonal structure of $\tilde{\mathbf{\Omega}}$, the $p \times m$ matrix of all m intraday return vectors stacked next to each other can be written as

$$\mathbf{X}_t = [\mathbf{r}_{t,1}, \mathbf{r}_{t,2}, \dots \mathbf{r}_{t,m}] = y_t \mathbf{C}_{\Omega_t} [\mathbf{z}_{t,1}, \mathbf{z}_{t,2}, \dots \mathbf{z}_{t,m}] = y_t \mathbf{C}_{\Omega_t} \mathbf{Z}_t \tag{27}$$

with $p \times m$ matrix $\mathbf{Z}_t$. We require that

1. at least one asset has a new price observation on all intraday intervals m,
2. the asset with the i'th most price observations has a price observation on a subset of the intraday intervals where the asset with the $(i+1)$'th most price observations, and
3. the assets are sorted from least to most liquid in the intraday return vectors $\mathbf{r}_{t,j}$.

That is, $\mathbf{Z}_t$ must have a structure like

$$\mathbf{Z}_t = \begin{bmatrix} z_{t,1,1} & \dots & z_{t,m_1,1} & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & & & \vdots & \vdots & & \vdots & \\ z_{t,1,p-1} & \dots & z_{t,m_1,p-1} & z_{t,m_1+1,p-1} & \dots & z_{t,m_{p-1},p-1} & 0 & \dots & 0 \\ z_{t,1,p} & \dots & z_{t,m_1,p} & z_{t,m_1+1,p} & \dots & z_{t,m_{p-1},p} & z_{t,m_{p-1}+1,p} & \dots & z_{t,m_p,p} \end{bmatrix}, \tag{28}$$

where the columns are "interchangeable". For example,

$$\begin{bmatrix} z_{t,1,1} & z_{t,2,1} & 0 \\ z_{t,1,2} & z_{t,2,2} & 0 \\ z_{t,1,3} & z_{t,2,3} & z_{t,3,3} \end{bmatrix} \text{ and } \begin{bmatrix} z_{t,1,1} & 0 & z_{t,3,1} \\ z_{t,1,2} & 0 & z_{t,3,2} \\ z_{t,1,3} & z_{t,2,3} & z_{t,3,3} \end{bmatrix} \tag{29}$$

are possible, but not

$$\begin{bmatrix} z_{t,1,1} & z_{t,2,1} & 0 \\ z_{t,1,2} & 0 & z_{t,3,2} \\ z_{t,1,3} & z_{t,2,3} & z_{t,3,3} \end{bmatrix}. \tag{30}$$

The vector $\mathbf{m} = (m_1, m_2, \ldots, m_p)^\top, m_1 \leq m_2 \leq \ldots \leq m_p$, contains for each asset i the number of intraday intervals with a new price observation m_i.

Theorem 3.2. *(The t-Riesz distribution.) For a given day t, let only those intraday returns for which there exists a new price observation in the corresponding intraday interval follow jointly a multivariate t-distribution with d.o.f. ν while replacing the others with zeros. More precisely, as explained above, assume that the vector of all intraday returns on day t is given by $\tilde{\mathbf{r}}_t = \sqrt{y}_t \mathbf{C}_{\tilde{\mathbf{\Omega}}_t} \tilde{\mathbf{z}}_t$, where $y_t \sim \Gamma(\nu/2, 2/\nu)$, $\mathbf{C}_{\tilde{\mathbf{\Omega}}_t}$ is the lower Cholesky decomposition of $\tilde{\mathbf{\Omega}}$, and $\tilde{\mathbf{z}}_t$ is a vector of standard normal distributed random variables and zeros with the restrictions on the placement of the zeros as explained in bulletpoints 1.-3. above. Then the RC on day t follows a t-Riesz distribution.*

$$\mathbf{R}_t \sim t\mathcal{R}(\mathbf{\Omega}_t, \mathbf{m}, \nu). \tag{31}$$

[a] *Proof in Appendix.*

a. The t-Riesz distribution is still defined if the d.o.f. parameter vector $\mathbf{m}$ is a vector of reals $\mathbf{n} = (n_1, \ldots, n_p)$, rather than of natural numbers m_i.

Note that for the t-Wishart distribution, which is obtained for the special case of of no missing new price observations ($m_1 = m_2 = m_p = m$), the assets can be ordered in any way. This is reflected in the fact that the t-Wishart is invariant to the ordering of the assets. Theorem 3.2 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 standard Riesz distribution can be generated by assuming a normal distribution on the intraday returns with heterogeneous liquidity. As for the t-Wishart compared to the Wishart, the superior performance of the t-Riesz compared to the Riesz distribution, shown in the empirical section below, mirrors that the assumption of a multivariate t compared

to the normal distribution on the intraday return vectors is more appropriate.

4 Time-Varying Mean

Now we specify the time-variation of $\mathbf{\Omega}_t$ to model the volatility dynamics. After having introduced the t-Riesz distribution family in the previous section, we return here to the full scope of all distributions considered in this paper. We assume

$$\mathbf{R}_t|\mathcal{F}_{t-1} \sim \mathcal{D}(\mathbf{\Omega}_t, \boldsymbol{\theta}_{\mathcal{D}}), \tag{32}$$

where $\mathcal{D}$ indicates the chosen distribution, $\mathcal{F}_{t-1} = \{\mathbf{R}_{t-1}, \mathbf{R}_{t-2}, \dots\}$ and the distribution-specific d.o.f. parameters ($\boldsymbol{\theta}_{\mathcal{D}}$) remain constant over time.

$\mathbf{\Omega}_t$ has a slightly different meaning for the various distributions, while $\mathbf{\Sigma}_t$ (the mean) has the same for all. So to achieve better comparability across distributions we assume that the time-variation in $\mathbf{\Omega}_t$ is driven by time-variation in $\mathbf{\Sigma}_t$ of the underlying distribution and focus on the equivalent standardized distribution representations

$$\mathbf{R}_t|\mathcal{F}_{t-1} \sim \mathcal{D}(\mathbf{\Sigma}_t, \boldsymbol{\theta}_{\mathcal{D}}). \tag{33}$$

A commonly used updating mechanism for $\mathbf{\Sigma}_t$, which we also assume in this paper, is the scalar-BEKK[20] recursion given by

$$\mathbf{\Sigma}_t = (1 - a - b)\mathbf{\Xi} + a\mathbf{R}_{t-1} + b\mathbf{\Sigma}_{t-1}, \tag{34}$$

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. For stationarity a necessary condition is that $a, b > 0 \wedge (a + b) < 1$ under which we have that the unconditional mean

$$\mathbb{E}\left[\mathbf{R}_t\right] = \mathbf{\Xi}. \tag{35}$$

20. 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.

In this paper, we assume stationarity. Note that if we restrict $a = b = 0$ in equation (34), we end up with a static distribution that is, $\boldsymbol{\Sigma} = \boldsymbol{\Xi}$, and

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

5 Empirical Analysis

5.1 Estimation

The number of parameters is dominated by the order $\mathcal{O}(p^2)$ matrix $\boldsymbol{\Xi}$, 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 $\boldsymbol{\Xi}$ (or $\boldsymbol{\Sigma}$ in the static model) with its obvious (see equation (35)) method-of-moments estimator

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

and estimate the remaining parameters ($\boldsymbol{\theta}_{\mathcal{D}}$, a, and b) conditional on $\widehat{\boldsymbol{\Xi}}$ via standard numerical maximum likelihood estimation. This multi-step estimation procedure, sometimes called *targeting*, reduces the size of the numerical optimization problem to the order of at most $\mathcal{O}(p)$, and for the Wishart-type distributions to at most four[21] parameters. The targeting two-step estimation procedure is common in the literature (see e.g. Noureldin, Shephard, and Sheppard 2012, Opschoor et al. 2018). Its consistency is intuitive and has been shown in the traditional multivariate GARCH framework by Francq, Horváth, and Zakoïan (2014). We expect consistency to carry over to the realized multivariate GARCH framework. In the empirical section of this paper, we always use this two-step estimation procedure, regardless of the cross-sectional dimension p.

A 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 (see Section 2.4).[22] This algorithm

21. These are n, and/or ν, a and b.
22. The seed for the random generation of permutations to try initially is kept the same for all

iteratively optimizes the likelihood for many different orderings of the assets (but not for all of the $p!$ possible ones, as this is computationally impossible). As such, it is clearly inconsistent but Blasques et al. (2021) have shown in a simulation experiment that its estimated likelihood value comes reasonably close to the one of the true asset ordering.

5.2 Data

Our original data are one-minute close prices from all trading days from 01 January 1998 to 05 February 2021 for every stock that was a constituent of the S&P 500 index during the sample period. We acquired the data from Quantquote[23], who combine, clean and process data directly obtained from various exchanges, where the biggest are NYSE, NASDAQ and AMEX[24].

We aim to generate the longest possible time-series of accurately estimated daily integrated covariance estimators. We exclude dates before 01 January 2002, because the NYSE fully implemented decimal pricing in 2001[25] and there are numerous other trading irregularities during 2001[26]. This leaves 4808 trading days. We then exclude stocks that have not been traded on one of the remaining days in the sample, which leaves 465 of 983 stocks. We only keep observations from official trading hours to be consistent across trading days. We then choose the 100 stocks with the most one-minute close price observations. Of those, the one with the least observations has, on average, 385.18 one-minute close price observations per trading day. Since the typical trading day has 390 minutes, on average less than five close prices are missing per day. Excluding illiquid stocks is common practice in creating time-series of RCs (see e.g. Lunde, Shephard, and Sheppard 2016). While this procedure biases the sample towards stocks that were very liquid over the entire sample period[27], it does ensure that the integrated covariance estimates are accurate for those stocks included.

Riesz-type distributions.

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

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

25. On 29 January 2001 to be precise.

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

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

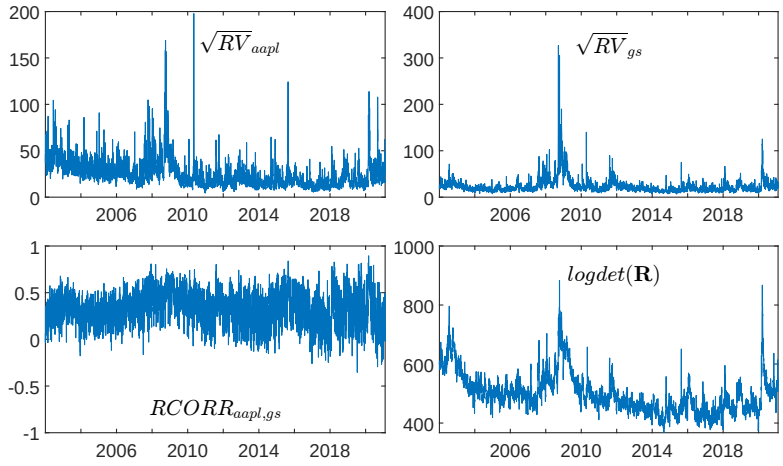

Figure 7: Top row: Annualized realized volatilities of Apple (aapl) and Goldman Sachs (gs), i.e. the square root of two of the elements on the main diagonal of the 100 asset $\mathbf{R}_t$ for the complete sample (01 January 2002 to 05 February 2021). Bottom row: Realized correlation between Apple and Goldman Sachs $RCORR_{aapl,gs} = \mathbf{R}_{aapl,gs}/(\sqrt{RV_{aapl}}\sqrt{RV_{gs}})$ and the natural logarithm of the determinant of the 100 assets $\mathbf{R}_t$ over time.

We follow Opschoor et al. (2018) and Blasques et al. (2021) and use five-minute returns to construct the 100-dimensional RCs. More specifically, we choose the subsampling realized covariance matrix which has been introduced by Zhang, Mykland, and Aït-Sahalia (2005) (see also Sheppard 2012). It is more efficient than the simple realized covariance matrix since it uses all our data (not just the data of one of the five-minute grids) by averaging the five distinct RCs obtained from these grids.[28] Furthermore, it produces positive definite matrices, even for high cross-sectional dimensions and low sampling frequencies.[29]

For a view of the data, see Figure 7, which shows the annualized realized volatility for Apple (appl) and Goldman Sachs (gs), as well as their realized correlation and the log-determinant of the 100-asset RCs. We see that the spikes in volatility are of similar magnitude for Apple and Goldman Sachs in the recent COVID-19-induced market turmoil, while the global financial crises of 2008/2009 caused volatility to spike much higher for Goldman Sachs than for Apple. The dot-com crisis (early 2000s), on the other hand, causes more volatility for Apple. We

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

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

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	135715	169570	337345	1055471	1077579	766533
Riesz	111237	140588	216604	691184	364470	251984
Inv.Wishart	101569	107661	223770	512282	328090	159749
Inv.Riesz	91149	93219	155817	361421	-69304	-87478
t-Wishart	49501	61117	77251	51997	-170922	-464564
t-Riesz	35264	46763	4442	-98244	-587842	-708956
Inv.t-Wishart	49650	55250	81863	24882	-381119	-773173
Inv.t-Riesz	41836	39405	16673	-113641	-750048	-887689
F	97642	104077	195081	444660	112776	-26494
F-Riesz	53016	58764	28253	61374	-612694	-663926
Inv.F-Riesz	59456	60140	45710	141882	-550025	-614263

Table 6: Bayes Information Criterion (BIC) values for the estimated static distributions on various datasets. The background shades are to be read column-wise, with the lowest BIC value shaded black and the highest shaded white, with linear gray-scaling in between. Largest values in red.

see that correlations are mainly positive and more stable around crisis 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-19, global financial crisis).

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

5.3 In-Sample

5.3.1 Static Distributions

As a first empirical exercise, we fit the different static distributions (i.e. $\mathbf{\Sigma}_t = \mathbf{\Xi}$) to the data. We use the two-step estimation method described above for all distributions and datasets. Table 14 in the appendix shows the estimated d.o.f. parameters. We focus now on Table 6, which shows the estimated Bayes Information Criterion (BIC) values. The distribution rankings according to BIC

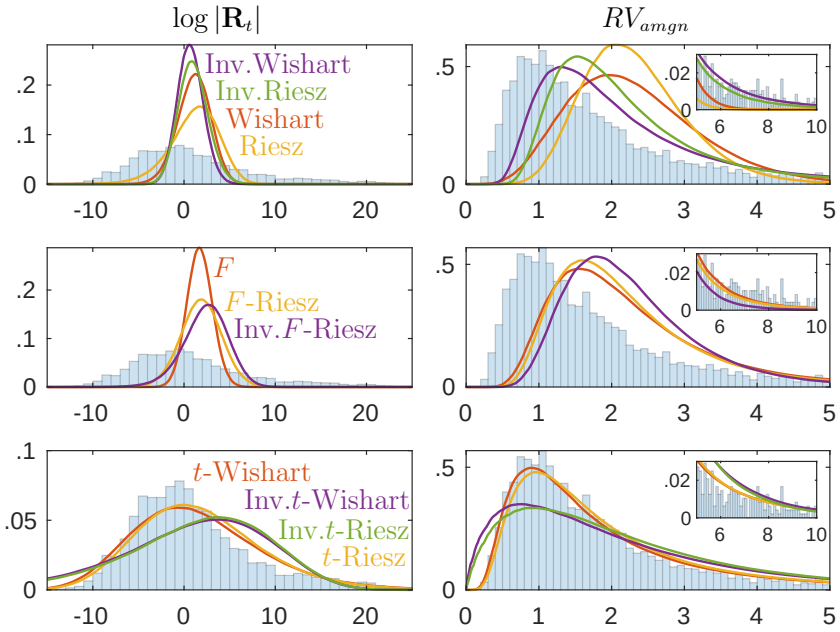

Figure 8: Histograms, normed to reflect p.d.f.s, of $\log|\mathbf{R}_t|$, and of the realized variance of Amgen, $RV_{amgn} = (\mathbf{R}_t)_{11}$ of the ten-dimensional dataset, as well as the respective marginal p.d.f.s implied by the fitted static matrix distributions. The p.d.f.s are kernel density estimates on 1.000.000 simulated realizations.

values are robust across the different datasets. We see a clustered pattern, where the (Inverse) Wishart, (Inverse) Riesz and F distribution could be considered as a group of similar fit and the (Inverse) t-Wishart, (Inverse) t-Riesz and (Inverse) F-Riesz as another group. The t-Riesz and Inverse t-Riesz distribution stand out from this second group since, for all datasets, one of these two obtains the best BIC value. The industry-specific datasets (Mining, Finance, and High-End Manufacturing) show a clear pattern favoring the t-Riesz distribution family. The F-Riesz and Inverse F-Riesz distributions also fit quite well overall, especially in the randomly drawn datasets. The Riesz and its special case, the Wishart, are the worst-fitting static distributions.

Next, we investigate how the different distributions for RCs match specific marginal characteristics of the data. In Figure 8, we plot the histograms of the log-determinant and the first diagonal element (RV of Amgen) of the ten-dimensional dataset. In the same figure we also plot the corresponding marginal p.d.f.s implied by the fitted probability distributions for RCs.[30] First, we observe that the

30. The plots for other datasets and companies look similar.

t-Wishart and t-Riesz distributions clearly mirror the empirical distribution of both the log-determinant and the RVs best. Second, only the t-Riesz distribution family (last row of the figure) exhibits a reasonable marginal distribution of the log-determinant; all other distributions have too much probability mass around their center and too little in the tails.

The F and F-Riesz distributions (second row, first column of the figure) do not match the (tail)-distribution of the log-determinant of the t-Wishart and t-Riesz because their ν and $\boldsymbol{\nu}$ d.o.f. parameter estimates are much higher (see Table 14 in the appendix). The d.o.f. parameters influence on the distribution of the log-determinant can be seen in the equations (15) and (17). In fact, the estimated ν parameter is fairly constant and close to five across all cross-sectional dimensions, whereas it increases for the F distribution from 11 for the five-dimensional dataset to 46 for the 25-dimensional dataset. A similar pattern is visible for the corresponding Riesz-type distributions. So we can see that in our definition of fat-tailedness, only the t-Riesz distribution family can be considered fat-tailed, as the other distributions do not have sufficient probability mass on large (> 10) log-determinants.

Considering the fat-tailedness of the RVs, the small subplots on the right column in Figure 8 show that all but the Wishart and Riesz distribution reasonably well match their empirical tail behavior. Again the t-Riesz distribution family implies the most probability mass on large RVs and the t-Riesz and t-Wishart match the empirical distribution in the tails best. For the entire distribution of the RVs the t-Riesz distribution family also clearly fits best. All other distributions assume too little probability mass on RVs smaller than one and too much on values between one and four. In particular, the fit of the t-Riesz and t-Wishart distribution marginal p.d.f.s for the empirical RVs are very good.

Next, in Figure 9, we plot the histogram of the realized covariances (RCOVS) between Cisco and Amgen ($(\mathbf{R}_t)_{21}$.[31] We see that the t-Riesz family also clearly fits best for the covariances, while the other distributions allocate too little probability mass RCOVs between 0 and 0.5 and too much probability mass on negative RCOVs and on RCOVs between 0.5 and 2.

31. The plots of the RV of Cisco in column one and the implied fitted distributions confirm our observations on RVs in the previous paragraph.

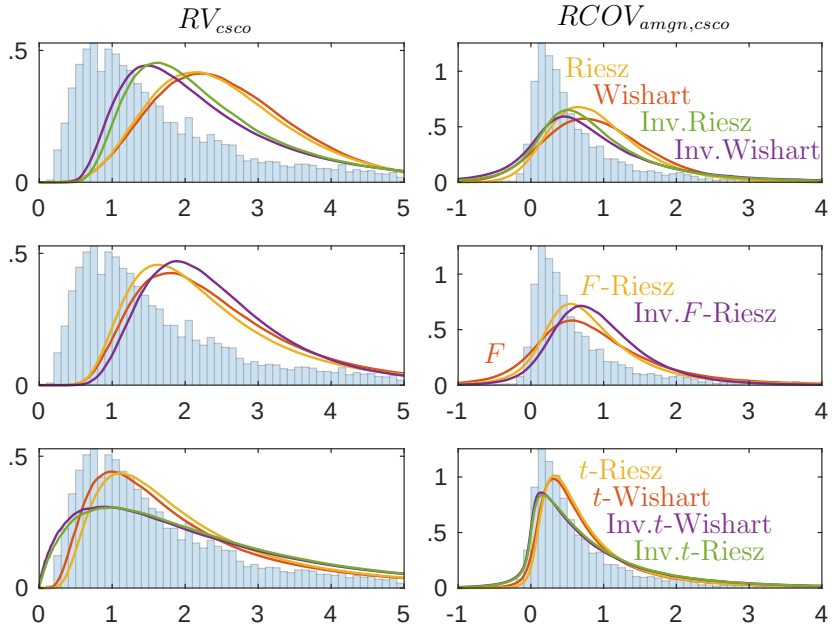

Figure 9: Histograms, normed to reflect p.d.f.s, of the realized variance of Cisco, $RV_{csco} = (\mathbf{R}_t)_{2\,2}$, and of the realized covariance between Cisco and Amgen, $RV_{amgen} = (\mathbf{R}_t)_{2\,1}$, in the ten-dimensional dataset, as well as the respective marginal p.d.f.s implied by the fitted static matrix distributions. The p.d.f.s are kernel density estimates on 1.000.000 simulated realizations.

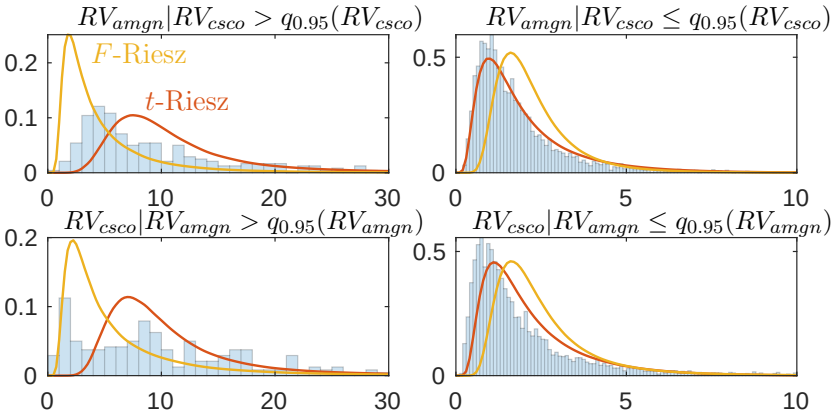

Figure 10: Histograms of real data, normed to reflect p.d.f.s, and the respective p.d.f.s of the realized variance of companies Amgen (first row, $RV_{amgn} = (\mathbf{R}_t)_{1\,1}$) and Cisco, (second row, $RV_{csco} = (\mathbf{R}_t)_{2\,2}$), conditional on the respective RV of the other company having a tail-realization (left column) or not (right column). The data comes from the random ten-dimensional dataset described in 5.2. The respective conditional p.d.f.s come from the fitted t-Riesz and F-Riesz distribution and are kernel density estimates on 1.000.000 simulated realizations.

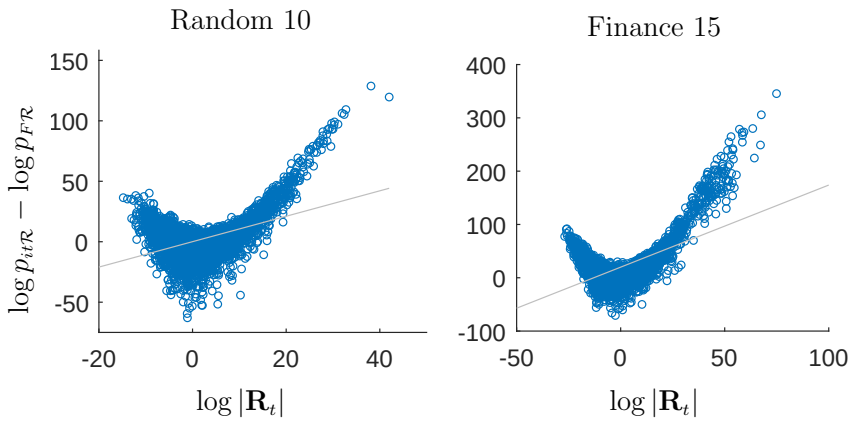

Figure 11: Difference in estimated log-likelihood contributions between Inverse t-Riesz and F-Riesz distributions, $\log p_{it\mathcal{R}}(\mathbf{R}_t|\widehat{\boldsymbol{\Sigma}}, \hat{\theta}) - \log p_{F\mathcal{R}}(\mathbf{R}_t|\widehat{\boldsymbol{\Sigma}}, \hat{\theta})$. Least-squares line is in gray.

Now we turn to tail-heterogeneity versus tail-homogeneity. We plot in Figure 10 histograms of the RV of Amgen (Cisco) given that the RV of Cisco (Amgen) has a tail-realization and given that it does not. We plot on top of each histogram the conditional p.d.f.s implied by the t-Riesz and F-Riesz distribution. A tail-realization is defined as an RV of a given asset that exceeds its empirical 95% quantile. Comparing the empirical distributions in the histograms, we see that an RV has a much higher probability of large realizations if another RV has a tail-realization (i.e. tail-homogeneity). The t-Riesz distribution can mirror this property of the data much better than the F-Riesz. While the F-Riesz has more probability mass on large RVs if another RV has a tail-realization, compared to if the other RV does not, this right-shift in probability mass is not as large as for the t-Riesz. Thus, as we explained intuitively by looking at the respective stochastic representation of the t-Riesz and F-Riesz in Section 3, it does make sense to call the t-Riesz tail-homogeneous and the F-Riesz tail-heterogeneous and our data favours tail-homogeneity.

Finally, we examine the *differences* in log-likelihood values between the tail-homogeneous Inverse t-Riesz and the tail-heterogeneous F-Riesz distribution more closely.[32] Figure 11 shows the difference in log-likelihood contributions between the two distributions depending on the log-determinant of the RCs for the random

32. We take the *Inverse* t-Riesz distribution in this comparison because, in the dynamic setting, it will turn out to be the distribution of the t-Riesz distribution family with the best log-likelihood values and will be in close competition with the F-Riesz.

	$\hat{a}$					
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 7: Estimated ARCH $\hat{a}$ parameters of the models in equations (33) and (34) for the different datasets and all distributions for RCs. All estimated $\hat{a}$ are highly significant, with the median (smallest) t-stat equalling 552 (123). The estimated persistence $(\hat{a} + \hat{b})$ is very similar across datasets and distributions and ranges from 0.976 (Wishart, dataset "Random 5") to 0.999 (Inverse t-Wishart, dataset "Finance 15"). All estimated $\hat{b}$ are highly significant, with the median (smallest) t-stat equalling 3060 (399).

ten-dimensional and the Finance 15-dimensional dataset. We see that the Inverse t-Riesz distribution gains its advantage in static fit mainly from the RCs with larger log-determinants. This is in line with our expectation that tail-heterogeneity is disadvantageous for crisis periods. The Inverse t-Riesz also fits better for very small RCs, which can be rationalized by the fact that in times of a very calm market, financial assets behave very homogeneously as well.

5.3.2 Time-Varying Mean

Now we fit the dynamic time-varying mean models for the different distributions to the datasets. The estimated d.o.f. parameters are given in Table 15 in the appendix. Table 7 contains the estimated ARCH parameters $(\hat{a})$. They are all highly significant. The estimated persistence parameters $(\hat{a} + \hat{b})$ are very close across distributions and datasets and range from 0.976 (Wishart, dataset "Random 5") to 0.999 (Inverse t-Wishart, dataset "Finance 15") with all GARCH parameters $(\hat{b})$ being highly significant. We see several clear patterns.

First, the estimated ARCH parameters become smaller with increasing cross-

Assets:	Random	Mining	Random	Finance	Random	Manuf.
#Assets:	5	6	10	15	25	25
Wishart	31334	22815	3269	-110723	-664333	-658311
Riesz	23901	12807	-25662	-176617	-843507	-844053
Inv.Wishart	15473	-3200	-64980	-263597	-1080371	-1092674
Inv.Riesz	11796	-8079	-79208	-295532	-1146456	-1164924
t-Wishart	10739	-11625	-66693	-297959	-935757	-933123
t-Riesz	5477	-16906	-87656	-355525	-1075614	-1079903
Inv.t-Wishart	2987	-22360	-109908	-390030	-1276488	-1308378
Inv.t-Riesz	-702	-26336	-123708	-425383	-1329647	-1365443
F	13960	-4742	-70770	-283792	-1140218	-1142989
F-Riesz	-1230	-24963	-124236	-379116	-1336077	-1345603
Inv.F-Riesz	456	-22958	-118306	-369412	-1322051	-1327107

Table 8: Bayes Information Criterion (BIC) values for the estimated dynamic distributions and different datasets. The background shades are to be read column-wise, with the lowest BIC value shaded black and the highest shaded white, with linear gray-scaling in between. Largest values in red.

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 $\widehat{\boldsymbol{\Xi}}$. They show that composite likelihood estimation can mitigate the bias. Unfortunately, composite likelihood estimation is not straightforward to apply on Riesz-type distributions due to their d.o.f. parameter vector(s). Another way to mitigate this bias might be to use a shrinkage estimator 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 above-mentioned methods to estimate the intercept matrix $\boldsymbol{\Xi}$.

The second pattern we observe is, that 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 $\boldsymbol{\Sigma}_t$, which indicates 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. In terms of fit and forecasting performance, a large mean-shifting reaction to previous RCs (as for the Wishart model) is actually beneficial in crisis periods, where RCs suddenly

spike in size and stay large for a short time. The good overall distributional fit of other distributions causes them to react more slowly to those volatility bursts.

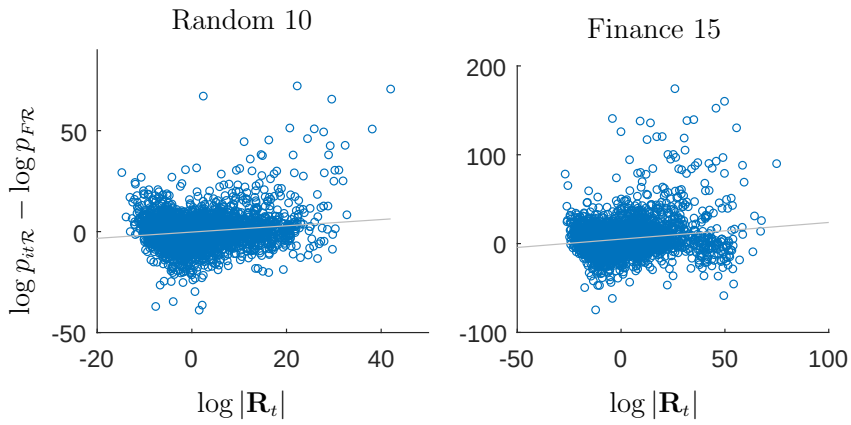

Figure 12: Difference in estimated log-likelihood contributions between Inverse t-Riesz and F-Riesz distributions with time-varying expected value matrix, $\log p_{it\mathcal{R}}(\mathbf{R}_t|\widehat{\mathbf{\Sigma}}_t, \hat{\theta}) - \log p_{F\mathcal{R}}(\mathbf{R}_t|\widehat{\mathbf{\Sigma}}_t, \hat{\theta})$. Least-squares line in gray.

Table 8 contains the BIC values for the estimated distributions with time-varying mean. As for the static distributions, the ranking across distributions is relatively stable over the cross-sectional dimension p. However, now in the dynamic setting, the ranking across distributions is less clustered. The Inverse t-Riesz and the F-Riesz distributions emerge as the clear winners exhibiting the smallest BIC 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 more advantageous for RCs of homogeneous assets.

The Riesz 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 since fitting the inverse distributions to $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_T\}$ is equivalent to fitting the corresponding 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 thinner tails; hence the good fit of inverse distributions. Obviously, by construction, every Wishart-type distribution has 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.

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

Finally, in Figure 12 we compare the log-likelihood contributions of the Inverse t-Riesz and F-Riesz distribution (similar to Figure 11 but for the time-varying mean specification). As in the static case, larger RCs are associated with higher log-likelihood contributions for the tail-homogeneous Inverse t-Riesz distributions, in line with our economic intuition that more volatile trading days exhibit more dependence among financial assets, favoring tail-homogeneity. Also in line with the static case, the smallest RCs are associated with higher log-likelihood contributions for the Inverse t-Riesz distribution, indicating that when markets are very calm, tail-homogeneity might be favored as well.

5.4 Out-of-Sample Forecasting Performance

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

For out-of-sample comparisons between different probability distributions, a natural loss function is the log-score, also known as the log posterior 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 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 look at the $t+1$ forecasting performance of the different distributions. To this end, Table 9 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.[33]

We see that the Inverse t-Riesz distribution emerges as the clear winner for log-score losses over a one-month forecasting period as it is the only member of the MCS for all datasets. Across datasets, the entire t-Riesz distribution family fits very well out-of-sample, and 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 16 and 17 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 (even here it has the lowest loss value), where also the F-Riesz distribution is in the MCS.

In Table 10 we also report the one-day-ahead log-score loss results for the entire sample. Here the Inverse t-Riesz distribution is the sole member of the MCS for the industry-specific Mining and Finance datasets, while for the other datasets, the F-Riesz distribution is also in the MCS. Still, the Inverse t-Riesz distribution has lower losses than the F-Riesz for all datasets except for the 25-dimensional one.

The above observations confirm our intuition that tail-homogeneity is a reasonable assumption. Clearly, the worst fitting distributions out-of-sample are the Riesz and its special case the Wishart distributions.

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

6 Conclusion

In conclusion, this paper provides a comprehensive comparison of probability distributions used to model realized covariance matrices (RCs) in financial applications. We reveal theoretical similarities and differences among the distributions, which are useful in explaining their disparity in empirical fit and forecasting performance. We derive the novel t-Riesz distribution family, which features tail-homogeneity as opposed to the tail-heterogeneity implied by the F-Riesz distribution. We show that 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 matrices are constructed. In the empirical part of the paper we perform fit and forecasting comparisons of the different distributions in different datasets and explain how the theoretical differences translate into differences in fit and forecasting performance. It emerges that when assuming a static distribution for the RCs, the t-Riesz distribution fits best in terms of BIC values. It matches the fat tails, the marginal distributions of the realized variances and covariances, and the tail-homogeneity that financial data exhibits. In the dynamic setting where the mean of the distributions is assumed to be time-varying, the distribution rankings are less pronounced. Here, the Inverse t-Riesz and F-Riesz distribution fit best in-sample and out-of-sample with an advantage in favor of

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

the former. We show that, especially in times of high market volatility and for assets of the same industry sector, tail-homogeneity is a more fitting assumption to the RC time-series data. Overall, the paper provides important insights for practitioners and researchers who want to model RCs of financial asset returns.

References

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

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. "The Distribution of Realized Stock Return Volatility." *Journal of Financial Economics* 61 (1): 43–76.

Archakov, Ilya, and Peter Reinhard Hansen. 2021. "A New Parametrization of Correlation Matrices." *Econometrica* 89 (4): 1699–1715.

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

Bartlett, Maurice S. 1933. "On the Theory of Statistical Regression." *Proc. R. Soc. Edinburgh* 53:260–283.

Bauer, Gregory H., and Keith Vorkink. 2011. "Forecasting Multivariate Realized Stock Market Volatility." *Journal of Econometrics* 160 (1): 93–101.

Bauwens, Luc, Giuseppe Storti, and Francesco Violante. 2012. "Dynamic Conditional Correlation Models for Realized Covariance Matrices." Working Paper, CORE.

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

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.

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.

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.

Engle, Robert F. 2002. "Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models." *Journal of Business and Economic Statistics* 20 (3): 339–350.

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 and Economic Statistics* 37 (2): 363–375.

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

Francq, Christian, Lajos Horváth, and Jean-Michel Zakoïan. 2014. "Variance Targeting Estimation of Multivariate GARCH Models." *Journal of Financial Econometrics* 14 (2): 353–382.

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.*

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

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

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

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

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* 106 (4).

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.

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

Koev, Plamen, and Alan Edelman. 2006. "The Efficient Evaluation of the Hypergeometric Function of a Matrix Argument." *Mathematics of Computation* 75 (254): 833–846.

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.

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

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

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

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.

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

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

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

Sucarrat, Genaro, and Steffen Grønneberg. 2020. "Risk Estimation with a Time-Varying Probability of Zero Returns." *Journal of Financial Econometrics* 20 (2): 278–309.

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.

Walck, Christian. 2007. *Hand-book on Statistical Distributions for Experimentalists.* University of Stockholm.

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

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

7 Appendix

7.1 Proofs

7.1.1 Proof of Theorem 2.1

Proof.

Proof of $\mathbb{E}[\underline{\mathbf{B}}\underline{\mathbf{B}}^\top]$: 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 Bartlett matrices. We have

$$(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top)_{ij} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}(\underline{\mathbf{B}}^\top)_{kj} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}. \tag{36}$$

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

$$\mathbb{E}[(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top)_{ij}] = \sum_{k=1}^{p} \mathbb{E}[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}] = \sum_{k=1}^{p} \mathbb{E}[\underline{\mathbf{B}}_{ik}]\mathbb{E}[\underline{\mathbf{B}}_{jk}] = 0, \tag{37}$$

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

$$(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top)_{ii} = \sum_{k=1}^{p} \underline{\mathbf{B}}_{ik}^2 = \sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^2, \tag{38}$$

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

$$\sum_{k=1}^{i} \underline{\mathbf{B}}_{ik}^2 \sim \chi^2_{n_i} \tag{39}$$

with expectation n_i. Thus

$$\mathbb{E}[(\underline{\mathbf{B}}\underline{\mathbf{B}}^\top)_{ii}] = n_i. \tag{40}$$

Proof of $\mathbb{E}[(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top)^{-1}]$: See Louati and Masmoudi (2015).
Proof of $\mathbb{E}[\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}]$: See Theorem 10 in Blasques et al. (2021).
Proof of $\mathbb{E}[\underline{\mathbf{B}}(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top)^{-1}\underline{\mathbf{B}}^\top]$: Due to independence, we have

$$\mathbb{E}[\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^\top] = \mathbb{E}[\underline{\mathbf{B}}\mathrm{dg}(\mathring{\boldsymbol{\nu}})\underline{\mathbf{B}}^\top],$$

where $\mathring{\boldsymbol{\nu}}$ is given in (11). Denote

$$\mathbf{T} = \underline{\mathbf{B}}(\mathring{\boldsymbol{\nu}})^{1/2},$$

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

$$\mathbf{R}_{ij} = \sum_{k=1}^{p} \mathbf{T}_{ik}(\mathbf{T}^\top)_{kj} = \sum_{k=1}^{p} \mathbf{T}_{ik}\mathbf{T}_{jk} = \sum_{k=1}^{p} \mathring{\nu}_k\, \underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk},$$

which for $i \neq j$ we have

$$\mathbb{E}[\mathbf{R}_{ij}] = \sum_{k=1}^{p} \mathring{\nu}_k \, \mathbb{E}[\underline{\mathbf{B}}_{ik}\underline{\mathbf{B}}_{jk}] = \sum_{k=1}^{p} \mathring{\nu}_k \, \mathbb{E}[\underline{\mathbf{B}}_{ik}]\mathbb{E}[\underline{\mathbf{B}}_{jk}] = 0,$$

because of the 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

$$\mathbb{E}[\mathbf{R}_{ii}] = \sum_{k=1}^{p} \mathring{\nu}_k \, \mathbb{E}[\underline{\mathbf{B}}_{ik}^2] = \sum_{k=1}^{i} \mathring{\nu}_k \, \mathbb{E}[\underline{\mathbf{B}}_{ik}^2],$$

with

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

Thus the elements of $\mathbb{E}_{_{iFR}}[\mathbf{R}] = \mathrm{dg}(\mathring{\mathring{\mathbf{n}}})$ are given by

$$\begin{aligned} \mathbb{E}[\mathbf{R}_{1\,1}] &= (n_1 - 1 + 1)\, \mathring{\nu}_1, \\ \mathbb{E}[\mathbf{R}_{2\,2}] &= \mathring{\nu}_1 + (n_2 - 2 + 1)\, \mathring{\nu}_2, \\ \mathbb{E}[\mathbf{R}_{3\,3}] &= \mathring{\nu}_1 + \mathring{\nu}_2 + (n_3 - 3 + 1)\, \mathring{\nu}_3, \\ &\vdots \end{aligned}$$

or

$$\mathring{\nu}_i = \sum_{j=1}^{i-1} \mathring{\nu}_j + (n_i - i + 1)\, \mathring{\nu}_i$$

or more precisely

$$\mathring{\nu}_i = \begin{cases} n_1 \, \mathring{\nu}_1, & \text{for } i = 1 \\ \sum_{j=1}^{i-1} \mathring{\nu}_j + (n_i + i - 1)\, \mathring{\nu}_i, & \text{for } i > 1, \end{cases} \tag{41}$$

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

7.1.2 Proof of Lemma 2.1

Proof. The equivalence between the two different representations is proofed in Maaß (1971), pp. 69-70. This proof is closely based on it. Using the LDL and Cholesky decomposition of $\mathbf{\Sigma}$,

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

Then

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

where $\mathbf{X}_{[i]}$ denotes the square submatrix created by taking the first i rows and columns of $\mathbf{X}$. So

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

and thus

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

Finally

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

□

7.1.3 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 $\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underset{\sim}{\mathbf{B}}$,

$i < j$ are given by

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

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 $n_i \to \infty$,

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

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

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

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

$$\operatorname*{plim}_{\mathbf{n}\to\infty}(\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}) = \mathbf{I}, \tag{42}$$

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,

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

and consequently

$$\operatorname*{plim}_{\boldsymbol{\nu}\to\infty}((\bar{\mathbf{B}}\mathrm{dg}(\boldsymbol{\nu})^{-\frac{1}{2}})^{-1}) = \mathbf{I}. \tag{43}$$

Now, we have

$$\mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}} = \underbrace{\mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}}\mathrm{dg}(\boldsymbol{\nu})^{-\frac{1}{2}}}_{\xrightarrow[\boldsymbol{\nu}\to\infty]{} \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}}\underbrace{\mathrm{dg}(\boldsymbol{\nu})^{\frac{1}{2}}\,\bar{\mathbf{B}}^{-\top}}_{\xrightarrow[\boldsymbol{\nu}\to\infty]{p} \mathbf{I}}\,\underline{\mathbf{B}} \xrightarrow[\boldsymbol{\nu}\to\infty]{d} \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}$$

and

$$\mathrm{dg}(\mathring{\bar{\mathbf{n}}})^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top} = \underbrace{\mathrm{dg}(\mathring{\bar{\mathbf{n}}})^{-\frac{1}{2}}\mathrm{dg}(\mathbf{n})^{\frac{1}{2}}}_{\xrightarrow[\mathbf{n}\to\infty]{} \mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}}\underbrace{\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}}_{\xrightarrow[\mathbf{n}\to\infty]{p} \mathbf{I}}\,\bar{\mathbf{B}}^{-\top} \xrightarrow[\mathbf{n}\to\infty]{d} \mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}$$

Finally,

$$\mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}} \xrightarrow[\boldsymbol{\nu}\to\infty]{d} \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \quad \text{and}$$
$$\mathrm{dg}(\mathring{\bar{\mathbf{n}}})^{-\frac{1}{2}}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathring{\bar{\mathbf{n}}})^{-\frac{1}{2}} \xrightarrow[\mathbf{n}\to\infty]{d} \mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}},$$

which are the stochastic representations of the Riesz and Inverse Riesz, respectively.

The proofs for

$$(\nu-2)(\chi^2_\nu)^{-1}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \xrightarrow[\nu\to\infty]{d} \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}}\underline{\mathbf{B}}\underline{\mathbf{B}}^{\top}\mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \quad \text{and}$$
$$\frac{\chi^2_n}{n}\mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}} \xrightarrow[n\to\infty]{d} \mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathrm{dg}(\mathring{\boldsymbol{\nu}})^{\frac{1}{2}}$$

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

□

7.1.4 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 p.d.f.s of the Riesz-type distributions.

Lemma 7.1. *(Faraut and Korányi 1994) For* $\mathbf{n}$ *with* $n_i > i-1$ *we have,*

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|^{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = 2^{p\bar{\mathbf{n}}/2}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)|\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}} \tag{44}$$

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

$$\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}^{-1}|^{\frac{\mathbf{n}+p+1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) d\mathbf{A} = \frac{1}{2^{p\bar{\mathbf{n}}/2}}\Gamma_p\left(-\frac{\overleftarrow{\mathbf{n}}}{2}\right)|\mathbf{B}|_{\frac{\mathbf{n}}{2}}\,. \tag{45}$$

Proof. The proofs can be found in Faraut and Korányi (1994) chapter VII.[34]. 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 $d = 1$.[35] 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.[36] In particular, from their Theorem VII.1.1.

$$\Gamma_\Omega(\mathbf{n}) = 2^{p(p-1)/4}\Gamma_p(\mathbf{n}), \tag{46}$$

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

$$\begin{aligned}\int_{\mathbf{A}>\mathbf{0}} |\mathbf{A}|^{\frac{\mathbf{n}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{B}\mathbf{A}\right) 2^{p(p-1)/4}d\mathbf{A} &= 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)|2\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}} \\ &= 2^{p(p-1)/4}\Gamma_p\left(\frac{\mathbf{n}}{2}\right)2^{p\bar{\mathbf{n}}/2}|\mathbf{B}^{-1}|_{\frac{\mathbf{n}}{2}}.\end{aligned}$$

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

34. 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).

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

36. I thank Jacques Faraut for pointing this out to me.

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

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

□

t-Riesz distribution: The stochastic representation is $\mathbf{R} = \mathbf{C}_\Omega(\bar{b})^{-2}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\mathbf{C}_\Omega^\top$, 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 p.d.f. of w and $\mathbf{A}$ is given by

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

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

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

where

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

Now integrating out w we get the p.d.f. of $\mathbf{R}$ as

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

where we used equation (5.9.1) of the NIST Digital Library of Mathematical Func-

tions.

Inverse t-Riesz distribution: We have $\mathbf{R} = \mathbf{C}_\Omega(\underset{\sim}{b})^2 \left(\bar{\mathbf{B}}\bar{\mathbf{B}}^\top\right)^{-1} \mathbf{C}_\Omega^\top$, 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 p.d.f. of w and $\mathbf{A}$ is given by

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

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

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

where

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

Now integrating out w we get the p.d.f. of $\mathbf{R}$ as

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

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 d.o.f. 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 $\mathbf{\Omega}^{-1}$.[37] Thus the stochastic representation of the Inverse F-Riesz distribution of type II is given by

$$\mathbf{R} = \mathbf{C}_{\Omega}\underline{\mathbf{B}}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\underline{\mathbf{B}}^{\top}\mathbf{C}_{\Omega}^{\top}, \tag{47}$$

which translate to $\mathbf{R} \sim i\mathcal{R}^{II}(\mathbf{Y}, \boldsymbol{\nu})$, $\mathbf{Y} \sim \mathcal{R}^{I}(\mathbf{\Omega}, \mathbf{n})$.[38] For the p.d.f. we can consequently use

$$\begin{aligned}
p_{iF\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, \mathbf{n}, \boldsymbol{\nu}) &= \int_{\mathbf{Y}>\mathbf{0}} p_{i\mathcal{R}^{II}}(\mathbf{R}|\mathbf{Y}, \boldsymbol{\nu}) p_{\mathcal{R}^{I}}(\mathbf{Y}|\mathbf{\Omega}, \mathbf{n}) d\mathbf{Y} \\
&= \int_{\mathbf{Y}>\mathbf{0}} \left(|\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{Y}|_{\frac{\nu}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}\mathbf{R}^{-1}\right) \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2) 2^{p\bar{\nu}/2}} \right. \\
&\qquad \left. \times |\mathbf{Y}|_{\frac{\mathbf{n}-p-1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{\Omega}^{-1}\mathbf{Y}\right) \frac{1}{\Gamma_p(\mathbf{n}/2)\, 2^{p\bar{\mathbf{n}}/2}} \right) d\mathbf{Y} \\
&= \frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)\, 2^{p(\bar{\boldsymbol{\nu}}+\bar{\mathbf{n}})/2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \\
&\quad \times \int_{\mathbf{Y}>\mathbf{0}} |\mathbf{Y}|_{\frac{\mathbf{n}+\boldsymbol{\nu}-p-1}{2}} \operatorname{etr}\left(-\frac{1}{2}\mathbf{Y}(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1})\right) d\mathbf{Y} \\
&= \frac{2^{p(\bar{\boldsymbol{\nu}}+\bar{\mathbf{n}})/2}\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)\, 2^{p(\bar{\boldsymbol{\nu}}+\bar{\mathbf{n}})/2}} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} \left| \left(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1}\right)^{-1} \right|_{\frac{\nu+\mathbf{n}}{2}} \\
&= \frac{\Gamma_p((\boldsymbol{\nu}+\mathbf{n})/2)}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)\Gamma_p(\mathbf{n}/2)} |\mathbf{R}|_{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}|_{-\frac{\mathbf{n}}{2}} |(\mathbf{\Omega}^{-1}+\mathbf{R}^{-1})^{-1}|_{\frac{\nu+\mathbf{n}}{2}},
\end{aligned} \tag{48}$$

where we used Theorem 7.1. Now rewrite using Lemma 2.2. □

7.1.5 Proof of Theorem 3.1

Proof. This Theorem is closely based on Gupta, Varga, and Bodnar (2013). If $\tilde{\mathbf{r}}$ follows an elliptically contoured distribution, $\tilde{\mathbf{r}} \sim E_{Tmp}(\mathbf{0}, \mathbf{I}_{Tm} \otimes \mathbf{\Omega}_p, \psi)$, then according to their Theorem 2.1 $\mathbf{X}^{\top} \sim E_{Tm,p}(\mathbf{0}, \mathbf{I}_{Tm} \otimes \mathbf{\Omega}_p, \psi)$ and then according to their Theorem 2.3 $\mathbf{X} \sim E_{p,Tm}(\mathbf{0}, \mathbf{\Omega}_p \otimes \mathbf{I}_{Tm}, \psi)$. Then the first part of our Theorem follows from their Theorem 5.5 and by noticing that

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

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

The fact that the marginal distribution of all daily RCs has the p.d.f. as shown in the Theorem, follows from Theorem 5.2 of Gupta, Varga, and Bodnar (2013) with their parameter $m = 1$.[39]

□

7.1.6 Proof of Theorem 3.2

Proof. The RC can be written as

$$\mathbf{R}_t = \sum_{j=1}^{m} \mathbf{r}_{t,j}\mathbf{r}_{t,j}^\top = \mathbf{X}_t\mathbf{X}_t^\top = y_t\mathbf{C}_{\Omega_t}\mathbf{Z}_t\mathbf{Z}_t^\top\mathbf{C}_{\Omega_t}^\top.$$

Hassairi, Ktari, and Zine (2022) show that if the assets in $\mathbf{Z}_t$ are sorted according to their liquidity with the least liquid asset in the first row, then $\mathbf{Z}_t\mathbf{Z}_t^\top$ follows a Riesz distribution with parameter matrix $\mathbf{\Omega}_t = \mathbf{I}$ and d.o.f. parameter vector $\mathbf{m}$, which implies that $\mathbf{C}_{\Omega_t}\mathbf{Z}_t\mathbf{Z}_t^\top\mathbf{C}_{\Omega_t}^\top \sim \mathcal{R}(\mathbf{\Omega}_t, \mathbf{m})$ follows a Riesz distribution with parameter matrix $\mathbf{\Omega}_t$ and d.o.f. parameter vector $\mathbf{n}$. Then according to Theorem 2.3 $\mathbf{R} \sim t\mathcal{R}(\mathbf{\Omega}_t, \mathbf{m}, \nu)$. □

7.2 Probability Density Functions

Our first aim is to derive the p.d.f.s in Table 5 from the stochastic representations of the respective distribution given by equation (4) in conjunction with Table 1.

For the derivation of the Riesz, Inverse Riesz and F-Riesz p.d.f.s we refer to Blasques et al. (2021), where to translate their notation to ours we use $\bar{\Gamma}_U(\mathbf{n}) = \Gamma_p(\overleftarrow{\mathbf{n}})$ (Lemma 7.2), ${}_U|\mathbf{X}|_{\mathbf{n}} = |\mathbf{X}^{-1}|_{-\mathbf{n}}$ (Lemma 2.2 (iv)) and $\mathbf{\Sigma}_{Blasques} = \mathbf{\Omega}$.

Lemma 7.2. *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*

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

39. Note that in part one of our Theorem, i.e. equation (), (and in Theorem 5.5 of Gupta, Varga, and Bodnar (2013)) the individual $\mathbf{R}_t$ are interchangeable.

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

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

□

In particular, the p.d.f. of the Riesz distribution $(\mathbf{C}_\Omega \underline{\mathbf{B}}\underline{\mathbf{B}}^\top \mathbf{C}_\Omega^\top)$ has been derived in Theorems 4 (i) of Blasques et al. (2021).

Our stochastic representation of the Inverse Riesz distribution $(\mathbf{C}_\Omega \bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathbf{C}_\Omega^\top)$ is the same as the one of the Inverse Riesz type II in Blasques et al. (2021) $(\mathbf{U}^{-\top}\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathbf{U}^{-1})$ (see their Theorem 4 (ii) and Definition 6 (ii)), because $\mathbf{U}$ is the upper Cholesky factor of $\mathbf{\Omega}^{-1}$ and consequently $\mathbf{U}^{-\top} = \mathbf{C}_\Omega$ is the lower Cholesky factor of $\mathbf{\Omega}$. The corresponding p.d.f. is given in their Theorem 7 (ii).

For the F-Riesz distribution $(\mathbf{C}_\Omega \bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\mathbf{C}_\Omega^\top)$ use their Theorem 8 (i) and notice that their $\mathbf{Y} = \mathbf{C}_\Omega\bar{\mathbf{B}}^{-\top}\bar{\mathbf{B}}^{-1}\mathbf{C}_\Omega^\top$ and thus, according to their Theorem 4 (i) their $\mathbf{X}|\mathbf{Y} = \mathbf{C}_\Omega\bar{\mathbf{B}}^{-\top}\underline{\mathbf{B}}\underline{\mathbf{B}}^\top\bar{\mathbf{B}}^{-1}\mathbf{C}_\Omega^\top$. So the stochastic representations are identical, and the corresponding p.d.f. is given in their Theorem 8 (i) and can be rewritten using Lemma 2.2 (iv) as

$$\begin{aligned}p_{F\mathcal{R}} &= \frac{\Gamma_p((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{\Omega}|^{\frac{\nu}{2}}|\mathbf{R}|^{\frac{\mathbf{n}-p-1}{2}}|\mathbf{\Omega}+\mathbf{R}|^{-\frac{\mathbf{n}+\nu}{2}} \\ &= \frac{\Gamma_p((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}|\mathbf{\Omega}|^{-\frac{\mathbf{n}}{2}}|\mathbf{R}|^{\frac{\mathbf{n}-p-1}{2}}|\mathbf{I}+\mathbf{C}_\Omega^{-1}\mathbf{R}\mathbf{C}_\Omega^{-\top}|^{-\frac{\mathbf{n}+\nu}{2}}.\end{aligned} \tag{49}$$

The p.d.f.s of the t-Riesz, Inverse t-Riesz, and Inverse F-Riesz distributions are derived in Theorem 2.3. The one of the Inverse F-Riesz can be rewritten using Lemma 2.2 (iv) as

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

Now that we have derived all Riesz-type p.d.f.s, it is easy to get the Wishart-type p.d.f.s since they are just special cases where all elements in the d.o.f. parameter vectors are equal to each other and using that for $\mathbf{n} = (n, n, \ldots, n)$, $\Gamma_p(\mathbf{n}) = \Gamma_p(n)$ and $|\mathbf{X}|_{\mathbf{n}}| = |\mathbf{X}|^n$ (see Definitions 2.1 and 2.2).

Notice that the stochastic representations of the F-Riesz and Inverse F-Riesz are, if $\forall i$, $n_i = n$ and $\nu_i = \nu$, (i.e. in case of an F distribution) $\mathbf{C}_\Omega \bar{\mathcal{B}}^{-\top} \underline{\mathcal{B}} \underline{\mathcal{B}}^\top \bar{\mathcal{B}}^{-1} \mathbf{C}_\Omega^\top$ and $\mathbf{C}_\Omega \underline{\mathcal{B}} \bar{\mathcal{B}}^{-\top} \bar{\mathcal{B}}^{-1} \underline{\mathcal{B}}^\top \mathbf{C}_\Omega^\top$, respectively, and their p.d.f.s are identical,

$$
\begin{aligned}
&p_{F\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, (n, \ldots, n), (\nu, \ldots, \nu)) \\
&\quad = \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Omega}|^{-\frac{n}{2}} |\mathbf{R}|^{\frac{n-p-1}{2}} |\mathbf{I} + \mathbf{C}_\Omega^{-1} \mathbf{R} \mathbf{C}_\Omega^{-\top}|^{-\frac{n+\nu}{2}} \\
&\quad = \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Omega}|^{\frac{\nu}{2}} |\mathbf{R}|^{\frac{n-p-1}{2}} |\mathbf{\Omega} + \mathbf{R}|^{-\frac{n+\nu}{2}} \\
&\quad = \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Omega}|^{\frac{\nu}{2}} |\mathbf{R}|^{-\frac{\nu+p+1}{2}} |\mathbf{\Omega}\mathbf{R}^{-1} + \mathbf{I}|^{-\frac{n+\nu}{2}} \\
&\quad = \frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)} |\mathbf{\Omega}|^{\frac{\nu}{2}} |\mathbf{R}|^{-\frac{\nu+p+1}{2}} |(\mathbf{I} + \mathbf{C}_\Omega^\top \mathbf{R}^{-1} \mathbf{C}_\Omega)^{-1}|^{\frac{\nu+n}{2}} \\
&\quad = p_{iF\mathcal{R}}(\mathbf{R}|\mathbf{\Omega}, (n, \ldots, n), (\nu, \ldots, \nu)) \\
&\quad = p_F(\mathbf{R}|\mathbf{\Omega}, n, \nu).
\end{aligned}
$$

This proves that both its alternative stochastic representations given in Table 1 yield the F distribution. Note that $\mathbf{R} \sim F\mathcal{R}(\mathbf{\Omega}^{-1}, \mathbf{n}, \boldsymbol{\nu}) \nRightarrow \mathbf{R}^{-1} \sim F\mathcal{R}(\mathbf{\Omega}, \boldsymbol{\nu}, \mathbf{n})$ for either type.[40] This is in contrast to the F distribution. Also, note that the standardized F-Riesz distribution cannot be obtained by mixing a standardized Riesz with a standardized Inverse Riesz but only by mixing the non-standardized versions and then standardizing the resulting distribution, as done above. This is also in contrast to the F distribution and can be seen since

$$
\begin{aligned}
&\mathbf{C}\mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}} \bar{\mathbf{B}}^{-\top} \underline{\mathbf{B}} \underline{\mathbf{B}}^\top \bar{\mathbf{B}}^{-1} \mathrm{dg}(\mathring{\mathbf{n}})^{-\frac{1}{2}} \mathbf{C}^\top \\
&\qquad\qquad \neq \mathbf{C}\mathrm{dg}(\mathring{\boldsymbol{\nu}})^{-\frac{1}{2}} \bar{\mathbf{B}}^{-\top} \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \underline{\mathbf{B}} \underline{\mathbf{B}}^\top \mathrm{dg}(\mathbf{n})^{-\frac{1}{2}} \bar{\mathbf{B}}^{-1} \mathrm{dg}(\mathring{\boldsymbol{\nu}})^{-\frac{1}{2}} \mathbf{C}^\top.
\end{aligned}
$$

Next, in Table 11 we list the p.d.f.s of the standardized distributions $p_{\mathcal{D}}(\mathbf{R}|\mathbf{\Sigma}, \boldsymbol{\theta}_{\mathcal{D}})$. They can be derived by replacing in the non-standardized p.d.f.s $\mathbf{\Omega} = \mathbf{C}\mathbf{M}_{\mathcal{D}}^{-1}\mathbf{C}$,

40. See the derivation of the Inverse F-Riesz type II. The derivation of the Inverse F-Riesz type I is very similar.

where $\mathbf{C}$ is the lower Cholesky factor of $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^\top$ and using Lemma 2.2 (iv).

Distribution	Probability Density Function $p_{\mathcal{D}}(\mathbf{R}\|\mathbf{\Sigma},\boldsymbol{\theta}_{\mathcal{D}})$
Wishart	$\frac{n^{np/2}}{2^{np/2}}\frac{1}{\Gamma_p(n/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|^{\frac{n}{2}}\,\operatorname{etr}(-\frac{1}{2}n\mathbf{Z})$
Riesz	$\frac{\prod_{i=1}^p n_i^{n_i/2}}{2^{p\bar{\mathbf{n}}/2}}\frac{1}{\Gamma_p(\mathbf{n}/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{\frac{\mathbf{n}}{2}}\,\operatorname{etr}(-\frac{1}{2}\operatorname{dg}(\mathbf{n})\mathbf{Z})$
Inv.Wishart	$\frac{(\nu-p-1)^{\nu p/2}}{2^{\nu p/2}}\frac{1}{\Gamma_p(\nu/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|^{-\frac{\nu}{2}}\,\operatorname{etr}(-\frac{1}{2}(\nu-p-1)\mathbf{Z}^{-1})$
Inv.Riesz	$\frac{\prod_{i=1}^p \mathring{\nu}_i^{-\nu_i/2}}{2^{p\bar{\boldsymbol{\nu}}/2}}\frac{1}{\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{-\frac{\boldsymbol{\nu}}{2}}\,\operatorname{etr}(-\frac{1}{2}\operatorname{dg}(\mathring{\boldsymbol{\nu}})^{-1}\mathbf{Z}^{-1})$
t-Wishart	$(\frac{n}{\nu-2})^{pn/2}\frac{\Gamma((\nu+pn)/2)}{\Gamma_p(n/2)\Gamma(\nu/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|^{\frac{n}{2}}\,(1+\frac{n}{\nu-2}\operatorname{tr}(\mathbf{Z}))^{-\frac{\nu+pn}{2}}$
t-Riesz	$\frac{\prod_{i=1}^p n_i^{n_i/2}}{(\nu-2)^{p\bar{\mathbf{n}}/2}}\frac{\Gamma((\nu+p\bar{\mathbf{n}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma(\nu/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{\frac{\mathbf{n}}{2}}\,(1+\frac{1}{\nu-2}\operatorname{tr}(\operatorname{dg}(\mathbf{n})\mathbf{Z}))^{-\frac{\nu+p\bar{\mathbf{n}}}{2}}$
Inv.t-Wishart	$(\frac{\nu-p-1}{n})^{\frac{\nu p}{2}}\frac{\Gamma((n+p\nu)/2)}{\Gamma(n/2)\Gamma_p(\nu/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|^{-\frac{\nu}{2}}\,(1+\frac{\nu-p-1}{n}\operatorname{tr}(\mathbf{Z}^{-1}))^{-\frac{n+p\nu}{2}}$
Inv.t-Riesz	$\frac{\prod_{i=1}^p \mathring{\nu}_i^{-\nu_i/2}}{n^{p\bar{\boldsymbol{\nu}}/2}}\frac{\Gamma((n+p\bar{\boldsymbol{\nu}})/2)}{\Gamma(n/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{-\frac{\boldsymbol{\nu}}{2}}\,(1+\frac{1}{n}\operatorname{tr}(\operatorname{dg}(\mathring{\boldsymbol{\nu}})^{-1}\mathbf{Z}^{-1}))^{-\frac{n+p\bar{\boldsymbol{\nu}}}{2}}$
F	$(\frac{n}{\nu-p-1})^{\frac{np}{2}}\frac{\Gamma_p((n+\nu)/2)}{\Gamma_p(n/2)\Gamma_p(\nu/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|^{\frac{n}{2}}\,\|\mathbf{I}+\frac{n}{\nu-p-1}\mathbf{Z}\|^{-\frac{\nu+n}{2}}$
F-Riesz	$\prod_{i=1}^p \mathring{n}_i^{\frac{n_i}{2}}\,\frac{\Gamma_p((\overleftarrow{\mathbf{n}}+\overleftarrow{\boldsymbol{\nu}})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{\frac{\mathbf{n}}{2}}\,\|\mathbf{I}+\operatorname{dg}(\mathring{\mathbf{n}})^{\frac{1}{2}}\mathbf{Z}\operatorname{dg}(\mathring{\mathbf{n}})^{\frac{1}{2}}\|_{-\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$
Inv.F-Riesz	$\prod_{i=1}^p \mathring{\mathring{n}}_i^{-\frac{\nu_i}{2}}\,\frac{\Gamma_p((\mathbf{n}+\boldsymbol{\nu})/2)}{\Gamma_p(\mathbf{n}/2)\Gamma_p(\overleftarrow{\boldsymbol{\nu}}/2)}\,\|\mathbf{R}\|^{-\frac{p+1}{2}}\,\|\mathbf{Z}\|_{-\frac{\boldsymbol{\nu}}{2}}\,\|(\mathbf{I}+\operatorname{dg}(\mathring{\mathring{\mathbf{n}}})^{-\frac{1}{2}}\mathbf{Z}^{-1}\operatorname{dg}(\mathring{\mathring{\mathbf{n}}})^{-\frac{1}{2}})^{-1}\|_{\frac{\mathbf{n}+\boldsymbol{\nu}}{2}}$

Table 11: Standardized probability density functions. We define $\mathbf{Z}=\mathbf{C}^{-1}\mathbf{R}\mathbf{C}^{-\top}$, where $\mathbf{C}$ is the lower Cholesky factor of $\mathbf{\Sigma}$. For the definition of $\mathring{\boldsymbol{\nu}}$, $\mathring{\mathbf{n}}$ and $\mathring{\mathring{\mathbf{n}}}$ Theorem 2.1. To derive these representations from the ones in Table 5 use Lemma 2.2 (iv).

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 12: Log-likelihood values for the estimated static distributions on various datasets. The background shades are to be read column-wise, with the lowest log-likelihood value shaded black and the highest shaded white, with linear gray-scaling in between. Largest values in red.

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 13: 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 shaded white, with linear gray-scaling in between. Largest values in red.

7.3 Other Empirical Results

Assets:		Rnd	Mngn	Rnd	Fin	Rnd	Manf
#Assets:		5	6	10	15	25	25
Wishart	n	7.0	8.3	12.8	15.0	27.7	29.4
Riesz	$\mathbf{n}_{min}$	1.4	1.9	1.2	0.8	1.1	1.5
	$\bar{\mathbf{n}}$	6.9	7.9	12.8	12.8	25.0	25.9
	$\mathbf{n}_{max}$	12.4	16.0	24.4	27.8	49.5	52.5
t-Wishart	n	17.1	17.5	22.7	28.4	38.2	42.3
t-Riesz	$\mathbf{n}_{min}$	3.9	4.4	2.2	2.2	2.1	4.4
	$\bar{\mathbf{n}}$	17.5	16.8	23.1	27.5	35.5	37.4
	$\mathbf{n}_{max}$	30.8	31.0	43.3	51.3	63.9	68.2
Inv.t-Wishart	n	3.1	4.1	3.6	1.7	3.8	3.5
Inv.t-Riesz	n	3.3	4.1	3.8	1.8	3.9	3.7
F	n	30.1	45.8	40.6	53.1	78.7	85.0
F-Riesz	$\mathbf{n}_{min}$	3.1	4.6	2.7	1.7	5.2	2.3
	$\bar{\mathbf{n}}$	43.3	41.9	57.5	63.6	74.5	75.0
	$\mathbf{n}_{max}$	110.4	108.8	168.0	191.2	180.5	190.4
Inv.F-Riesz	$\mathbf{n}_{min}$	1.7	2.8	1.5	1.0	3.2	2.6
	$\bar{\mathbf{n}}$	1297.5	1098.2	1658.0	981.2	697.4	700.0
	$\mathbf{n}_{max}$	6429.7	6344.4	8213.8	5285.6	7480.9	7440.0
Inv.Wishart	ν	9.2	11.2	16.2	19.3	34.6	36.3
Inv.Riesz	$\boldsymbol{\nu}_{min}$	4.0	4.2	4.2	4.1	4.2	4.3
	$\bar{\boldsymbol{\nu}}$	8.5	11.0	15.7	19.3	31.7	32.4
	$\boldsymbol{\nu}_{max}$	11.9	16.8	24.1	31.0	48.2	46.6
t-Wishart	ν	4.1	4.8	4.9	3.0	5.2	5.3
t-Riesz	$\boldsymbol{\nu}_{min}$	4.3	4.5	5.4	3.1	5.7	5.5
Inv.t-Wishart	ν	17.3	18.4	22.6	29.4	41.2	47.7
Inv.t-Riesz	$\boldsymbol{\nu}_{min}$	8.0	5.3	3.9	6.2	4.4	6.7
	$\bar{\boldsymbol{\nu}}$	18.7	19.6	24.3	32.8	42.3	45.5
	$\boldsymbol{\nu}_{max}$	27.9	30.9	36.4	50.5	70.3	67.9
F	ν	10.9	12.9	20.7	21.5	43.1	45.9
F-Riesz	$\boldsymbol{\nu}_{min}$	4.0	4.2	5.0	3.7	3.6	3.8
	$\bar{\boldsymbol{\nu}}$	7.9	9.8	14.1	15.1	31.8	31.8
	$\boldsymbol{\nu}_{max}$	10.7	13.7	18.6	25.2	51.4	53.8
Inv.F-Riesz	$\boldsymbol{\nu}_{min}$	4.1	3.6	4.6	4.2	3.6	3.9
	$\bar{\boldsymbol{\nu}}$	22.4	25.2	34.6	33.2	47.5	50.4
	$\boldsymbol{\nu}_{max}$	49.4	62.3	83.9	77.4	103.7	111.3

Table 14: Estimated degree of freedom parameters of static distributions for the different datasets.

Assets:		Rnd	Mngn	Rnd	Fin	Rnd	Manf
#Assets:		5	6	10	15	25	25
Wishart	n	17.6	20.9	26.3	33.2	45.2	45.1
Riesz	$\mathbf{n}_{min}$	6.3	7.6	6.0	5.4	5.0	5.0
	$\bar{\mathbf{n}}$	16.6	19.6	24.0	29.7	39.5	39.5
	$\mathbf{n}_{max}$	26.4	33.2	41.0	51.1	66.9	66.9
t-Wishart	n	24.8	30.4	33.4	42.8	50.7	50.7
t-Riesz	$\mathbf{n}_{min}$	8.3	10.3	7.9	7.1	6.3	6.2
	$\bar{\mathbf{n}}$	23.1	27.9	30.1	38.5	44.3	44.3
	$\mathbf{n}_{max}$	35.2	41.1	49.0	64.5	71.7	73.2
Inv.t-Wishart	n	24.2	24.3	23.9	16.5	24.2	21.8
Inv.t-Riesz	n	23.1	23.9	22.1	14.4	22.5	20.9
F	n	81.6	109.4	122.1	134.2	171.3	179.4
F-Riesz	$\mathbf{n}_{min}$	17.2	18.5	14.5	12.8	16.2	12.9
	$\bar{\mathbf{n}}$	71.1	84.7	93.9	100.5	117.3	124.4
	$\mathbf{n}_{max}$	142.4	169.3	203.2	210.3	225.8	264.6
Inv.F-Riesz	$\mathbf{n}_{min}$	10.2	12.4	8.5	7.8	6.8	6.7
	$\bar{\mathbf{n}}$	3078.3	2237.5	1005.8	1604.8	1965.0	1187.7
	$\mathbf{n}_{max}$	15248.2	13161.6	9356.7	12236.0	21896.1	10330.9
Inv.Wishart	ν	20.9	25.4	31.9	39.9	54.1	54.4
Inv.Riesz	$\boldsymbol{\nu}_{min}$	9.9	11.1	10.1	8.7	9.2	9.7
	$\bar{\boldsymbol{\nu}}$	19.6	23.8	29.0	35.8	48.1	48.9
	$\boldsymbol{\nu}_{max}$	25.2	32.6	38.7	46.3	62.7	64.7
t-Wishart	ν	19.8	20.3	20.6	15.3	23.7	20.4
t-Riesz	ν	19.8	18.8	20.6	13.7	23.7	19.9
Inv.t-Wishart	ν	26.8	33.0	37.9	48.3	58.7	59.6
Inv.t-Riesz	$\boldsymbol{\nu}_{min}$	11.0	13.5	11.2	8.7	10.5	11.7
	$\bar{\boldsymbol{\nu}}$	25.5	31.3	35.5	45.7	54.1	55.2
	$\boldsymbol{\nu}_{max}$	34.1	42.7	46.9	61.6	70.6	74.0
F	$\boldsymbol{\nu}_{min}$	27.3	32.3	41.1	53.4	70.8	69.7
F-Riesz	$\boldsymbol{\nu}_{min}$	9.3	10.4	9.5	10.6	10.7	10.5
	$\bar{\boldsymbol{\nu}}$	19.2	22.6	27.4	35.7	50.0	49.7
	$\boldsymbol{\nu}_{max}$	26.0	28.6	34.7	53.3	73.2	72.7
Inv.F-Riesz	$\boldsymbol{\nu}_{min}$	9.0	10.1	9.5	10.1	9.5	9.5
	$\bar{\boldsymbol{\nu}}$	33.5	39.3	45.1	51.3	63.4	63.5
	$\boldsymbol{\nu}_{max}$	56.8	82.1	79.0	83.7	102.3	102.1

Table 15: Estimated degree of freedom parameters of dynamic mean shifting distributions for the different datasets.

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 16: Average of log-score loss over a one-month forecasting period (22 trading days), $-\sum_{j=1}^{22} p_{\mathcal{D}}(\mathbf{R}_{t+j}|\widehat{\boldsymbol{\Sigma}}_{j+1}, \hat{\boldsymbol{\theta}}_{\mathcal{D},j+1})$, for the forecasting window from 1 January 2007 to 31 December 2011, where each model is re-estimated every ten 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 17: Average of log-score loss over a one month forecasting period (22 trading days), $-\sum_{j=1}^{22} p_{\mathcal{D}}(\mathbf{R}_{t+j}|\widehat{\boldsymbol{\Sigma}}_{j+1}, \hat{\boldsymbol{\theta}}_{\mathcal{D},j+1})$, for the forecasting window from 1 January 2012 to 31 December 2019, where each model is re-estimated every ten trading days. 90% model confidence sets in red.