Tail Heterogeneity for Dynamic Covariance-Matrix-Valued Random Variables Using the F-Riesz Distribution*

Francisco Blasques[a,b] Andre Lucas[a,b] Anne Opschoor[a,b] Luca Rossini[c]

[a]Vrije Universiteit Amsterdam, The Netherlands

[b]Tinbergen Institute, The Netherlands

[c]University of Milan, Italy

August 20, 2022

Abstract

We introduce the dynamic conditional F-Riesz distribution to model the dynamics of (realized) covariance matrices under tail-heterogeneity and fat-tails. Unlike earlier matrix-valued distributions from the econometric literature, the F-Riesz distribution allows for different tail behavior across all variables in the system using vector-valued degrees of freedom parameters. We study the consistency properties of the maximum likelihood estimator for a two-step (targeting) approach. Empirically, we find that tail-heterogeneity and tail-fatness are both empirically relevant: when applying the new F-Riesz distribution to realized covariance matrices of 30 U.S. stock returns, we find large likelihood increases both in-sample and out-of-sample compared to all competing distributions, including the Wishart, inverse Wishart, Riesz, inverse Riesz, and the matrix-F distribution.

Key words: Matrix Distributions, Tail Heterogeneity, (inverse) Riesz Distribution, Fat-Tails, Realized Covariance Matrices.

*Blasques thanks the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.195.099.

1 Introduction

Covariance matrix modeling and estimation play an important role in many areas of economics and statistics, such as financial risk assessment and decision making under uncertainty Markowitz (1991); Engle et al. (2019). Today's data-rich environment has led to a shift in ambition from estimating static covariance matrices to estimating covariance matrices on a frequent basis over many short spans of data, also known as realized covariance matrix estimation. Examples include Andersen et al. (2003); Barndorff-Nielsen and Shephard (2004); Chiriac and Voev (2011); Lunde et al. (2016); Callot et al. (2017); Bollerslev et al. (2018, 2020) and the references cited therein.

An important challenge is to design parsimonious yet flexible time-series models for such series of realized covariances that can be used for forecasting purposes and for decision models. A complication is that the time-series observations are matrix-valued (rather than vector-valued), have positive (semi)-definite outcomes only, and may be subject to fat-tailed behavior and outliers. Most of the models currently available cannot cope with all of these challenges simultaneously or are highly restrictive. Recent work on tensor-valued time-series such as Wang et al. (2019) and Chen et al. (2021) can deal with matrix valued time series, but not with restrictions on positive definiteness of the observations or with the fat-tailed nature of these data in many applications. Other approaches that can deal with positive definite random matrices build on distributional theory, but are typically highly restrictive. For instance, the often used Wishart or inverse Wishart distributions for matrix-valued time series only features a mean and a single degrees of freedom parameter for tail behavior (Golosnoy et al., 2012; Jin and Maheu, 2013, 2016), while the matrix-F distribution only features two tail parameters (Konno, 1991; Opschoor et al., 2018). While such distributions might be suitable for low-dimensional matrices, in moderate to high dimensions the implied constraints on tail behavior in the cross-section are typically too restrictive empirically.

An often used approach in the literature to flexibilize multivariate distributions is to split the marginal distributions from the dependence structure by using copulas (Patton, 2009; Oh and Patton, 2017, 2018; Opschoor et al., 2020). That approach, however, cannot easily be transferred to the current setting. Most copula methods related to vector-valued observations and cannot deal with the positive definiteness of covariance-matrix-

valued observations. Second, typical copula structures available in the literature for higher dimensional data are again tightly parameterized with very little heterogeneity in the tail-dependence structure, such as the Gaussian, (skewed) Student's t, and Archimedean copula.

In this paper, we introduce the dynamic F-Riesz distribution to model sequences of realized covariance matrices. We do so by building on (and correcting) the beta type II Riesz distribution of Díaz-García (2016). We then introduce dynamics for the key scale parameter of this distribution and derive the consistency properties of the maximum likelihood estimator for the parameters of this model. A key property of the F-Riesz distribution is that it allows for different tail-heterogeneity in each of its coordinates. It does so by replacing the two scalar degrees of freedom parameters of the matrix-F distribution of Konno (1991) by two *vectors* of degrees of freedom parameters. In our empirical study we show that across a wide range of dimensions the F-Riesz distribution leads to large likelihood increases with respect to other well-known matrix distributions such as the Wishart and inverse Wishart. It illustrates that allowing for tail-heterogeneity is empirically relevant.

We obtain the F-Riesz distribution by mixing a Riesz distribution (Hassairi and Lajmi, 2001; Díaz-García, 2013) and Inverse Riesz distribution (Tounsi and Zine, 2012; Louati and Masmoudi, 2015), which are generalizations of the Wishart and Inverse Wishart distributions. Their domain is the space of positive (semi)-definite matrices. The Riesz distribution has mainly been used in the physics literature (Andersson and Klein, 2010). In economic statistics, Gribisch and Hartkopf (2022) also recently apply the Riesz distribution to financial data. They introduce a state-space version of the dynamic Riesz distribution and estimate the model using Bayesian techniques. We differ from their approach in two important ways. First, we use the more flexible F-Riesz-distribution rather than the standard Riesz. This allows for additional distributional flexibility, which appears empirically relevant in our application. Second , we use an observation-driven rather than a parameter-driven approach to model the dynamics of realized covariance matrices. As a result we can obtain the likelihood in closed form and can stick to standard maximum likelihood techniques for the estimation of model parameters.

The main difference between the Riesz and the Wishart distribution is that the Riesz distribution is characterized by as many degrees of freedom as the dimension of the data, whereas the Wishart only has one single degrees of freedom parameter. This allows for much

more heterogeneity in tail behavior. The Riesz distribution, however, still has thin-tailed behavior for its coordinates. This might be problematic for many empirical data sets in finance and economics, which typically exhibit fat-tailed behavior. Be mixing a Riesz and inverse Riesz distribution, the F-Riesz distribution as constructed in this paper allows for heterogeneous fat-tailed behavior in all directions. The F-Riesz distribution is characterized by a matrix-valued mean, and two *vectors* of degrees of freedom parameters, thus allowing for considerable extra flexibility in tail behavior. If each of these vectors is scalar (i.e., has the same elements), then the F-Riesz reduced to the matrix-F distribution (see Konno, 1991; Opschoor et al., 2018).

We apply the F-Riesz distribution to a sample of daily realized covariance matrices of dimension 5 to 30 using U.S. stock data over 2001–2014. The results indicate that both the tail heterogeneity and fat-tailedness of the F-Riesz-distribution are emprically relevant compared to tail heterogeneity only (Riesz) or tail fatness only (matrix-F). We strongly reject both the Riesz and the matrix-F distributions for the dynamics of realized covariance matrices, despite each of these already being substantially better than the Wishart and inverse Wishart distributions. We conclude that F-Riesz distribution can thus provide a relevant tool for the statistical modeling of covariance-matrix-valued time-series, both in a classical framework as in this paper, or in a Bayesian framework as in Gribisch and Hartkopf (2022).

The rest of this paper is set-up as follows. In Section 2 we introduce the model. Section 3 considers consistency properties of the maximum likelihood estimator using a two-step targeting approach. Section 4 presents the new model's performance using simulated data. Section 5 presents the empirical results. Section 6 concludes. An appendix gathers all the technical results. As a general notational guide, scalars have normal type face, vectors are bolded, and matrices are bolded and capitalized. We number theorems, propositions, definitions and assumptions consecutively.

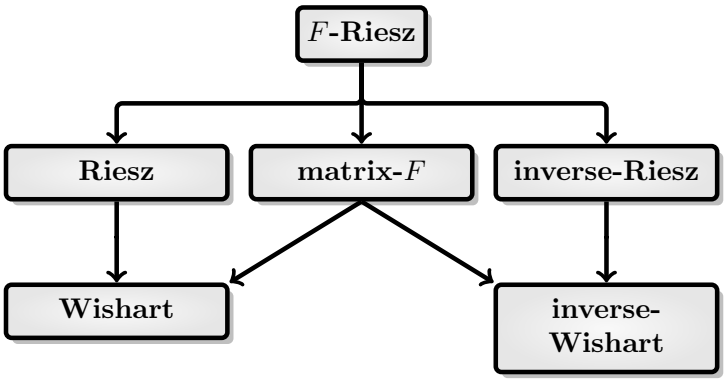

Figure 1: Family of matrix distributions
This figure shows a family tree of the F-Riesz distributions. Connected lines means that distributions are related by generalization.

2 The dynamic F-Riesz model

2.1 The F-Riesz distribution

A useful family tree of the different distributions considered in this paper is provided in Figure 1. Whereas the Wishart (and possibly the matrix-F) distribution are more familiar, the Riesz distribution is typically less well known. Therefore, we first briefly introduce the Riesz before deriving the F-Riesz distribution. A more extensive introduction to the different distributions, as well as more technical derivations, can be found in the online appendix.

The Riesz distribution is characterized by two parameters: a positive definite scaling matrix $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$ with lower triangular Cholesky decomposition $\boldsymbol{L}$, and a vector of degrees of freedom parameters $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^\top$, with $\nu_i > i-1$ for $i = 1, \dots, k$. Arguably the easiest way to introduce the Riesz distribution is via its so-called Bartlett decomposition, which is well known as a simulation device for the standard Wishart distribution; see Anderson (1962). Consider a random matrix $\boldsymbol{G} \in \mathbb{R}^{k \times k}$, defined as

$$\boldsymbol{G} = \begin{pmatrix} \sqrt{\chi^2_{\nu_1}} & 0 & \cdots & 0 \\ \mathcal{N}(0,1) & \ddots & 0 & \vdots \\ \vdots & \mathcal{N}(0,1) & \ddots & 0 \\ \mathcal{N}(0,1) & \cdots & \mathcal{N}(0,1) & \sqrt{\chi^2_{\nu_k - k + 1}} \end{pmatrix}, \tag{1}$$

where all elements of $\boldsymbol{G}$ are independent random variables. Then $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{G}^\top\boldsymbol{L}^\top$ has a so-called Riesz distribution of type I, where type I relates to the fact that we have taken a lower triangular Cholesky decomposition in the Bartlett decomposition. We denote this as $\boldsymbol{Y} \sim \mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$. Type-II versions of the distribution based on the upper triangular decomposition also exist and we refer to the appendix for details.

Given the use of the Cholesky decomposition, it is clear that the ordering of the variables matters. This is well-known and accepted in the Riesz literature. For instance for $k = 2$ and $\boldsymbol{\Sigma} = \mathbf{I}_2$, switching both rows and columns of $\boldsymbol{Y} \sim \mathcal{R}^I(\mathbf{I}_2, \boldsymbol{\nu})$ for $\boldsymbol{\nu} = (\nu_1, \nu_2)^\top$ does not yield a Riesz distribution with $(\nu_2, \nu_1)^\top$ degrees of freedom. The order of the variables can be recovered from the data under the assumption of correct specification by maximizing the likelihood also over the order of the variables in the system, rather than over $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}$ only. In high dimensional systems, this additional optimization is of combinatorial complexity. Empirically, we found that changes in the orders of the variables in the system typically result only in second order improvements of the likelihood function. The dominant increase in the likelihood is obtained by generalizing the Wishart into a Riesz, or the matrix-F into an F-Riesz. Nevertheless, we also propose a heuristic approach later in the paper to optimize efficiently over the ordering of the variables in order to maximize the likelihood.

If $\boldsymbol{\nu} = (\nu, \ldots, \nu)$ for some $\nu > k - 1$, the Riesz distribution collapses to the well-known Wishart distribution with ν degrees of freedom. In that case, the degrees of freedom parameters ν_i on the diagonal of the Bartlett decomposition $\boldsymbol{G}$ all have the same value ν. The Riesz distribution thus generalizes the Wishart by allowing heterogeneous tail behavior in the cross section. The tail-fatness, however, is left unaffected and is inherited from the exponential tail behavior of the chi-squared distributions on the diagonal of the Bartlett decomposition $\boldsymbol{G}$ of the Wishart and Riesz.

To introduce fat tails as well as tail-heterogeneity into the distribution, it is instructive to consider the case of the matrix-F distribution of Konno (1991). If $\boldsymbol{X}$ conditional on $\boldsymbol{Y}$ has a Wishart distribution, $\boldsymbol{X} \mid \boldsymbol{Y} \sim \mathcal{W}(\boldsymbol{Y}^{-1}, \mu)$, and $\boldsymbol{Y}$ has a Wishart distribution, $\boldsymbol{Y} \sim \mathcal{W}(\mathbf{I}_k, \nu)$, then the unconditional distribution of $\boldsymbol{X}$ is the matrix-F, or $\boldsymbol{X} \sim \mathcal{F}(\mathbf{I}_k, \mu, \nu)$. Replacing the Wishart distribution by Riesz-type distributions, it is intuitive that a similar result might be obtainable that allows for fat tails under tail-heterogeneity. This is precisely the result of our first theorem, which we prove in the appendix.

To formulate the theorem, we first need to define the concepts of the generalized (lower and upper) gamma function and the (lower and upper) power weighted determinant.

Definition 1 (**generalized multivariate gamma functions**)**.** The *lower* generalized multivariate gamma function for a vector-valued argument $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top \in \mathbb{R}^{k\times 1}$ is defined as

$$\overline{\Gamma}(\boldsymbol{\nu}) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\nu_i + \frac{1-i}{2}\right), \tag{2}$$

with $2\nu_i > i - 1$ for $i = 1, \ldots, k$.
The *upper* generalized multivariate gamma function is defined similarly as

$$\overline{\Gamma}_U(\boldsymbol{\nu}) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\nu_i + \frac{i-k}{2}\right) = \overline{\Gamma}(\boldsymbol{\nu} + \tilde{\boldsymbol{\gamma}}), \tag{3}$$

for $2\nu_i > k - i$ for $i = 1, \ldots, k$, and

$$\tilde{\boldsymbol{\gamma}} = \tfrac{1}{2}(k+1) - \Big(k, k-1, \ldots, 1\Big)^\top = \Big(-\tfrac{1}{2}(k-1), \ldots, \tfrac{1}{2}(k-1)\Big)^\top. \tag{4}$$

The upper and lower generalized multivariate gamma functions enter the integrating constant of the Riesz and F-Riesz distributions. Note that if $\boldsymbol{\nu} = (\nu, \ldots, \nu)^\top$ in (2) such that all its elements are the same, then $\overline{\Gamma}(\boldsymbol{\nu}) = \Gamma_k(\nu)$, with $\Gamma_k(\nu)$ the standard multivariate gamma function.

Next, we introduce the concept of Lower Power Weighted Determinants (LPWD) and Upper Power Weighted Determinants (UPWD). These power weighted determinants for the F-Riesz density take a similar role as the standard determinants found in the density expressions of the matrix-F density.

Definition 2 (**power weighted determinants**)**.** Consider the vector $\boldsymbol{\nu} \in \mathbb{R}^{k\times 1}$ and a positive definite matrix $\boldsymbol{Y}$. Let $\boldsymbol{L}$ and $\boldsymbol{U}$ be the (unique) *lower* and *upper* triangular Cholesky decompositions of $\boldsymbol{Y}$, i.e., $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{U}\boldsymbol{U}^\top$, with $\boldsymbol{L}$ and $\boldsymbol{U}$ a lower and upper triangular matrix, respectively, each with positive diagonal elements. Then the Lower Power Weighted Determinant (LPWD) and Upper Power Weighted Determinant (UPWD) of $\boldsymbol{Y}$,

denoted as $|\boldsymbol{Y}|_{\boldsymbol{\nu}}$ and ${}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}}$, respectively, are given by

$$|\boldsymbol{Y}|_{\boldsymbol{\nu}} = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_i}, \qquad {}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}} = \prod_{i=1}^{k} \boldsymbol{U}_{i,i}^{2\nu_i}. \tag{5}$$

In the physics literature, the power weighted determinants are commonly introduced via so-called weight functions; see for instance Gross and Richards (1987). In this paper, we instead use the notation of power weighted determinants as it is closer to the econometric literature and stresses the analogy between the Wishart and Riesz, and between the matrix-F and F-Riesz density expressions. It also allows us to easily see that the matrix-F (respectively the Wishart) is a special case of the F-Riesz(respectively the Wishart). To facilitate the latter, Lemma 20 in the appendix provides manipulation rules for power weighted determinants that will prove convenient. In particular, we note that the power weighted determinant is *not* a regular determinant. Simple properties like $|\boldsymbol{A}\cdot\boldsymbol{B}| = |\boldsymbol{A}|\cdot|\boldsymbol{B}|$ for matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{k\times k}$ need (and in general will) no longer hold for power weighted determinants.

We can now formulate our first main result.

Theorem 3 (F**-Riesz distribution**).

(i) Assume $\boldsymbol{X}|\boldsymbol{Y} \sim \mathcal{R}^{I}(\boldsymbol{Y}, \boldsymbol{\mu})$ *has a conditional Riesz type-I distribution, and* $\boldsymbol{Y} \sim i\mathcal{R}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ *has an inverse Riesz type-II distribution, then* $\boldsymbol{X}$ *is* $\mathcal{FR}^{I}$ *distributed with density function*

$$p_{\mathcal{FR}^{I}}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \frac{\overline{\Gamma}_U\left(\frac{\boldsymbol{\mu}+\boldsymbol{\nu}}{2}\right) \cdot |\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}}}{\overline{\Gamma}_U\left(\frac{\boldsymbol{\nu}}{2}\right)\overline{\Gamma}\left(\frac{\boldsymbol{\mu}}{2}\right)} |\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)} |\boldsymbol{\Sigma}+\boldsymbol{X}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}.$$

(ii) Assume $\boldsymbol{X}|\boldsymbol{Y} \sim \mathcal{R}^{II}(\boldsymbol{Y}, \boldsymbol{\mu})$ *and* $\boldsymbol{Y} \sim i\mathcal{R}^{I}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$*, then* $\boldsymbol{X}$ *is* $\mathcal{FR}^{II}$ *distributed with density function*

$$p_{\mathcal{FR}^{II}}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \frac{\overline{\Gamma}\left(\frac{\boldsymbol{\mu}+\boldsymbol{\nu}}{2}\right) \cdot {}_U|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}}}{\overline{\Gamma}\left(\frac{\boldsymbol{\nu}}{2}\right)\overline{\Gamma}_U\left(\frac{\boldsymbol{\mu}}{2}\right)} {}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)} {}_U|\boldsymbol{\Sigma}+\boldsymbol{X}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}.$$

(iii) Let $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{\top} = \boldsymbol{U}\boldsymbol{U}^{\top}$ *for lower and upper triangular matrix* $\boldsymbol{L}$ *and* $\boldsymbol{U}$*, respectively. If* $\boldsymbol{X} \sim \mathcal{FR}^{I}(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ *then* $\boldsymbol{L}^{-1}\boldsymbol{X}(\boldsymbol{L}^{\top})^{-1} \sim \mathcal{FR}^{I}(\mathbf{I}_k, \boldsymbol{\mu}, \boldsymbol{\nu})$*. Similarly, if* $\boldsymbol{X} \sim$

$\mathcal{FR}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ *then* $\boldsymbol{U}^{-1}\boldsymbol{X}(\boldsymbol{U}^{\top})^{-1} \sim \mathcal{FR}^{II}(\mathbf{I}_k, \boldsymbol{\mu}, \boldsymbol{\nu})$.

Theorem 3 corrects a result from the literature on the generalized beta II distribution of Díaz-García (2016). This earlier result is only valid for $\boldsymbol{\Sigma} = \mathbf{I}_k$, and incorrect otherwise. In our application it is crucial that the density is correct for dealing with time-varying scale matrices, which is why we use our corrected expression in Theorem 3. Interestingly, except for the use of the non-standard generalized gamma functions and power weighted determinants, the shape of the F-Riesz density mirrors and generalizes that of the standard matrix-F distrubution one-on-one. The following corollary establishes this link and shows that the matrix-F distribution of Konno (1991) as used by Opschoor et al. (2018) is just a special case of the F-Riesz distribution.

Corollary 4. *Under the conditions of Theorem 3 part (i) or (ii), if we assume* $\boldsymbol{\mu} = \mu \cdot \boldsymbol{\iota}_k$ *and* $\boldsymbol{\nu} = \nu \cdot \boldsymbol{\iota}_k$*, then* $\boldsymbol{X}$ *has a matrix-F distribution* $\mathcal{F}(\mu, \nu)$*, and*

$$p_{\mathcal{FR}^I}(\boldsymbol{X}; \boldsymbol{\Sigma}, \mu \cdot \boldsymbol{\iota}_k, \nu \cdot \boldsymbol{\iota}_k) = p_{\mathcal{F}}(\boldsymbol{X}; \boldsymbol{\Sigma}, \mu, \nu) = \frac{\Gamma_k\left(\frac{\mu+\nu}{2}\right) \cdot |\boldsymbol{\Sigma}|^{0.5\nu}}{\Gamma_k\left(\frac{\nu}{2}\right)\Gamma_k\left(\frac{\mu}{2}\right)} \cdot \frac{|\boldsymbol{X}|^{0.5(\mu-k-1)}}{|\boldsymbol{\Sigma}+\boldsymbol{X}|^{0.5(\mu+\nu)}}, \tag{6}$$

where $\Gamma_k(\cdot)$ *is the multivariate gamma function.*

The corollary makes clear that it is possible to test whether the F-Riesz collapses to the matrix-F distribution by testing whether all elements in $\boldsymbol{\mu}$ are the same, as well as all elements in $\boldsymbol{\nu}$. This is a simple likelihood ratio test, and we revisit this in the simulation section later on.

Similar to the scalar-F or matrix-F distributions, the F-Riesz does not always have an expectation. In the following theorem, we derive the appropriate expression for the F-Riesz's expectation for type I and II if it exists. These moments turn out to be useful for re-parameterizing the dynamic specification of the model and for designing a two-step targeting approach to estimation in the next sections.

Theorem 5 (**Expectation of the F-Riesz distribution**).

(i) Let $\boldsymbol{Y} \sim \mathcal{FR}^{I}(\mathbf{I}, \boldsymbol{\mu}, \boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{A}_k$*, where* $\boldsymbol{A}_k$ *is a diagonal matrix with ith*

diagonal element a_i *equal to*

$$a_i = \begin{cases} \frac{\mu_1}{\nu_1 - k - 1}, & \text{for } i = 1, \\ \frac{1}{\nu_i - k + i - 2} \left(\mu_i + \sum_{i=1}^{i-1} a_i \right), & \text{for } i = 2, \ldots, k, \end{cases}$$

provided $\mu_i > 0$ *and* $\nu_i > k + 2 - i$ *for* $i = 1, \ldots, k$.

(ii) Let $\boldsymbol{Y} \sim \mathcal{FR}^{II}(\mathbf{I}, \boldsymbol{\mu}, \boldsymbol{\nu})$, *then* $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{A}_k$, *where* $\boldsymbol{A}_k$ *is a diagonal matrix with* i*th diagonal element* a_i *equal to*

$$a_i = \begin{cases} \frac{1}{\nu_i - i - 1} \left(\mu_i + \sum_{i=i+1}^{k} a_i \right), & \text{for } i = 1, \ldots, k-1, \\ \frac{\mu_k}{\nu_k - k - 1}, & \text{for } i = k, \end{cases}$$

provided $\mu_i > 0$ *and* $\nu_i > i + 1$ *for* $i = 1, \ldots, k$.

Combining this result with Theorem 3.(iii), the expectation of a general $\mathcal{FR}^{I}(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ random variable equals $\boldsymbol{L}\boldsymbol{A}_k\boldsymbol{L}^\top$ for $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, with $\boldsymbol{L}$ lower triangular. Similarly, the expectation of a $\mathcal{FR}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ random variable equals $\boldsymbol{U}\boldsymbol{A}_k\boldsymbol{U}^\top$ for $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{U}^\top$, with $\boldsymbol{U}$ upper triangular.

2.2 Conditional Autoregressive F-Riesz models

We apply the F-Riesz distribution later in this paper to observations $\boldsymbol{X}_t$ of realized covariance matrices. Of course, one can assume that these are i.i.d. observations from the F-Riesz or related distributions. This, however, is empirically implausible for realized covariance matrices, which typically exhibit quite some persistence over time. To illustrate the usefulness of the new distributions, we therefore consider a model specification that allows for more persistence in $\boldsymbol{X}_t$. As an example, we consider the F-Riesz type I distribution. A similar set-up can be made for the other distributions discussed in this paper. Let $\boldsymbol{\Sigma}_t = \boldsymbol{L}_t \boldsymbol{L}_t^\top$ for a lower triangular matrix, then we consider the model

$$\boldsymbol{X}_t \mid \mathcal{F}_{t-1} \sim \mathcal{FR}^{I}(\boldsymbol{\Sigma}_t, \boldsymbol{\mu}, \boldsymbol{\nu}), \tag{7}$$

$$\boldsymbol{V}_t = \mathbb{E}[\boldsymbol{X}_t \mid \mathcal{F}_{t-1}] = \boldsymbol{L}_t \boldsymbol{M}(\boldsymbol{\mu}, \boldsymbol{\nu}) \boldsymbol{L}_t^\top, \tag{8}$$

$$\boldsymbol{V}_{t+1} = (1 - A - B)\boldsymbol{\Omega} + A \boldsymbol{X}_t + B \boldsymbol{V}_t, \tag{9}$$

where $\mathcal{F}_{t-1} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{t-1}\}$ contains the lagged observations, and $\boldsymbol{M}(\boldsymbol{\mu}, \boldsymbol{\nu})$ is a diagonal matrix-valued function containing the expectation of the standard ($\boldsymbol{\Sigma}_t = \mathbf{I}_k$) F-Riesz type I distribution. We take $\boldsymbol{\Omega}$ as a symmetric positive definite parameter matrix, and A and B as scalar parameters. The model will be labeled as the Conditional Autoregressive F-Riesz model (CAFr). Equation 9 creates persistence in the conditional expectation of $\boldsymbol{X}_t$, which is a typical empirical feature of realized covariance matrices. This persistent conditional mean enters the observation equation (7), causing persistence in $\boldsymbol{X}_t$ itself. If we consider the Wishart rather than the F-Riesz distribution in (7), the model resembles the Conditional Autoregressive Wishart (CAW) model of Golosnoy et al. (2012) In that case, the model is also one of the two core equations of the Multivariate HEAVY model of Noureldin et al. (2012). Of course, the model can easily be extended to have non-scalar A and B, more lags of $\boldsymbol{V}_t$ and/or $\boldsymbol{X}_t$, and other dynamics including long-memory or HAR type dynamics (Corsi, 2009). For the sake of this paper, the current specification suffices and already shows that we obtain large increases in likelihood by allowing for tail-heterogeneity using the F-Riesz distribution. The size of additional gains obtained by changing to more persistent dynamics can be found in for instance Corsi (2009) and is not the core of our contribution in this paper.

Model (7)–(9) is observation driven and thus allows for easy parameter estimation via maximum likelihood using a standard prediction error decomposition. Standard errors are obtained using the standard sandwich covariance matrix estimator.

To reduce the dimensionality of the optimization, we use a targeting approach to estimate the matrix $\boldsymbol{\Omega}$. First note that there is a simple mapping between $\boldsymbol{\Sigma}_t$ and $\boldsymbol{V}_t$ for given $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. In particular for the F-Riesz type I distribution, $\boldsymbol{L}_t = \boldsymbol{L}_{\boldsymbol{V}_t} \boldsymbol{M}(\boldsymbol{\mu}, \boldsymbol{\nu})^{-1/2}$, where $\boldsymbol{L}_{\boldsymbol{V}_t}$ is a lower triangular matrix such that $\boldsymbol{V}_t = \boldsymbol{L}_{\boldsymbol{V}_t} \boldsymbol{L}_{\boldsymbol{V}_t}^\top$. Similar expressions hold for the other distributions. The targeting approach works as follows. Assuming stationarity and the existence of unconditional first moments, we can take unconditional expectations of the left and right-hand sides of (9) to obtain $\overline{\boldsymbol{V}} = \mathbb{E}[\boldsymbol{V}_t] = \boldsymbol{\Omega}$. We use this to estimate $\boldsymbol{\Omega}$ by the sample average, $\hat{\boldsymbol{\Omega}} = n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_t$. The likelihood then holds only in the remaining parameters A, B, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$.

2.3 Ordering of variables

A final issue for likelihood maximization is determining the order of the variables in the system. As mentioned earlier, the order of the variable matters for the specification of the Riesz type distributions. Enumeration of all possible orders and picking the highest likelihood value is typically unfeasible in high dimensions. To approximate the optimal order, we propose a heuristic approach, the core of which is described in the following algorithm.

Algorithm 6 (**Approximating the optimal ordering of variables in the system**). *Let $o = (o_1, \ldots, o_k)$ be a permutation of the first k integers, indicating the order of the variables in the system that make up the covariance matrix observations $\boldsymbol{X}_t$. Also, let $\boldsymbol{\theta}$ denote the static parameters that characterize the model and that need to be estimated by maximum likelihood.*

Step 0 *Set $j = 0$.*

Step 1 *Select a random order $o^{(j)} = (o_1^{(j)}, \ldots, o_k^{(j)})$.*

Step 2 *Given the ordering $o^{(j)}$, estimate $\boldsymbol{\theta}$ by maximum likelihood (possibly combined with a targeting approach) to obtain $\hat{\boldsymbol{\theta}}^{(j)}$.*

Step 3 *Loop over asset i, $i = 1, \ldots, k$:*

Step 3a *Find $i^\star$ such that $i = o_{i^\star}^{(j)}$, i.e., find the position of asset i in the current ordering $o^{(j)}$.*

Step 3b *Put asset i in each of the possible positions $1, \ldots, k$, while keeping the order of the other variables as in $o^{(j)}$, i.e., consider the permutations $(o_{i^\star}^{(j)}, o_1^{(j)}, \ldots, o_{i^\star-1}^{(j)}, o_{i^\star+1}^{(j)}, \ldots, o_k^{(j)})$, $(o_1^{(j)}, o_{i^\star}^{(j)}, o_2^{(j)}, \ldots, o_{i^\star-1}^{(j)}, o_{i^\star+1}^{(j)}, \ldots, o_k^{(j)})$, up to $(o_1^{(j)}, \ldots, o_{i^\star-1}^{(j)}, o_{i^\star+1}^{(j)}, \ldots, o_k^{(j)}, o_{i^\star}^{(j)})$. Retain the ordering that yields the highest log-likelihood value for* given $\hat{\boldsymbol{\theta}}^{(j)}$ *and store it as $o^{(j+1)}$.*

Step 3c *Increase j to $j + 1$ and re-estimate $\hat{\boldsymbol{\theta}}^{(j)}$.*

Step 3d *Continue the loop by proceeding to the next asset $i + 1$.*

Step 4 *Repeat steps 1–3 for p different random initial orderings, retaining the final order that yields the highest log-likelihood. Call this final order $o^{(opt)}$ with corresponding parameter estimate $\hat{\boldsymbol{\theta}}^{(opt)}$.*

Some more details and possible variations on this heuristic algorithm are provided in the online appendix. The algorithm ensures that the maximized likelihood never decreases for the different orderings of the variables in the system considered during the search. Moreover, the algorithm is efficient in that it limits the number of times we re-estimate the parameter $\boldsymbol{\theta}$. The latter is costly due to the (possibly high-dimensional) non-linear optimization problem. In particular, we only re-estimate $\boldsymbol{\theta}$ $p(k+1)$ times. This is substantially smaller for large k than the full $k!$ enumerated possible orderings and therefore provides an enormous computational gain. Though no guarantee is given that we arrive at the true optimum using this algorithm, the simulation evidence in the next section shows that it results in some considerable likelihood increases and typically gets close to the correct order of the variables measured in terms of rank correlations.

3 Estimation and inference

In this section, we establish the consistency of several estimators for the parameters of the F-Riesz distribution. For the case of a sample $\{\boldsymbol{X}_t\}_{t=1,\dots,T}$ composed of random i.i.d. draws from a F-Riesz distribution, we study the consistency of the estimator $\hat{\boldsymbol{V}}_T$ of the matrix $\boldsymbol{V} = \mathbb{E}[\boldsymbol{X}_t]$, the estimator $\hat{\boldsymbol{\Sigma}}_T$ of the scale matrix $\boldsymbol{\Sigma}$, and the estimator $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ of the vector $(\boldsymbol{\mu}, \boldsymbol{\nu})$, as a two-step maximum likelihood estimator relying on a plug-in formulation of the log likelihood function. Additionally, for the case of a sample $\{\boldsymbol{X}_t\}_{t=1,\dots,T}$ generated by the conditional autoregressive F-Riesz model in (7)-(9), we obtain the consistency of the MLE for all the static parameters.

3.1 MLE consistency for an i.i.d. sample

We first consider the consistent estimation of the unknown parameters of interest for an i.i.d. random sample from a F-Riesz distribution using the maximum likelihood estimator (MLE). Assumption 7 below states the distributional nature of the data generating process.

Assumption 8 imposes the finiteness of the true $\boldsymbol{\Sigma}_0$ as well as restrictions on the parameter space for the vector $(\boldsymbol{\mu}, \boldsymbol{\nu})$.

Assumption 7. *The sequence* $\{\boldsymbol{X}_t\}_{t=1,...,T}$ *is i.i.d. with* $X_t \sim \mathcal{FR}^I(\boldsymbol{\Sigma}_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ *for every* $t = 1, ..., T$.

Assumption 8. *The positive definite matrix* $\boldsymbol{\Sigma}_0$ *satisfies* $\|\boldsymbol{\Sigma}_0\| < \infty$ *and* $(\boldsymbol{\mu}, \boldsymbol{\nu})$ *lie on a compact set satisfying* $\boldsymbol{\nu}_i > i + 1 \ \forall \ i$, *and* $\boldsymbol{\mu}_j > K + 1 \ \forall \ j$ *and containing* $(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$.

Proposition 9 now establishes the strong consistency of the sample average as an estimator of $\boldsymbol{V}_0 = \mathbb{E}[\boldsymbol{X}_t]$.

Proposition 9. *Let assumptions 7 and 8 hold. Then* $\hat{\boldsymbol{V}}_T = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_t \overset{a.s.}{\to} \boldsymbol{V}_0 = \mathbb{E}[\boldsymbol{X}_t]$ *as the sample diverges,* $T \to \infty$.

To prove the consistency of the two-step estimation procedure, we reparameterize the model using the fact that $\boldsymbol{\Sigma} = \boldsymbol{L}_{\boldsymbol{V}} \boldsymbol{A}_k^{-1} \boldsymbol{L}_{\boldsymbol{V}}^\top$, where $\boldsymbol{L}_{\boldsymbol{V}}$ is the (unique) Cholesky decomposition of $\boldsymbol{V}$. We define $\tilde{p}_{\mathcal{FR}^I}$ as the reparameterized density function which takes $\boldsymbol{V}$ rather than $\boldsymbol{\Sigma}$ as an argument, i.e., $\tilde{p}_{\mathcal{FR}^I}(\,\cdot\,; \boldsymbol{V}, \boldsymbol{\mu}, \boldsymbol{\nu}) \equiv p_{\mathcal{FR}^I}(\,\cdot\,; \boldsymbol{L}_{\boldsymbol{V}} \boldsymbol{A}_k^{-1} \boldsymbol{L}_{\boldsymbol{V}}^\top, \boldsymbol{\mu}, \boldsymbol{\nu})$.

Theorem 10 below establishes the strong consistency of the two-step MLE as $T \to \infty$. Note that in this two-step estimator $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ depends on the first-step estimator for $\boldsymbol{V}_0$. Specifically, we define the MLE $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ as the maximizer of the plug-in estimates $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_T, \boldsymbol{\mu}, \boldsymbol{\nu})$,

$$(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) = \arg \max_{(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathcal{U} \times \mathcal{V}} \frac{1}{T} \sum_{t=1}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_T, \boldsymbol{\mu}, \boldsymbol{\nu}), \tag{10}$$

rather than of the true log-likelihood contributions $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_0, \boldsymbol{\mu}, \boldsymbol{\nu})$. In Theorem 10, we build on standard M-estimation theory to obtain the strong consistency of our MLE to $(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$.

Theorem 10. *Let assumptions 7 and 8 hold. Then the MLE* $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ *satisfies* $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) \overset{a.s.}{\to} (\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ *as* $T \to \infty$.

Corollary 11 builds on the strong consistency of $\hat{\boldsymbol{V}}_T$ from Proposition 9 and the weak consistency of $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ from Theorem 10 to obtain the consistency of the estimator $\hat{\boldsymbol{\Sigma}}_T$ towards Σ_0.

Corollary 11. *Let assumptions 7 and 8 hold. Then* $\hat{\boldsymbol{\Sigma}}_T \xrightarrow{p} \Sigma_0$ *as* $T \to \infty$.

3.2 MLE consistency for the conditional autoregressive F-Riesz model

We now turn to the consistency of the MLE for the unknown static parameter of the conditional autoregressive F-Riesz model. Assumption 12 below defines the conditional autoregressive F-Riesz model in (7)-(9) as the data generating process. Assumption 13 imposes the finiteness of the true intercept parameter $\boldsymbol{\Omega}_0 = \mathbb{E}[\boldsymbol{V}_t]$ and the compactness of the parameter space of the remaining parameters. Additionally, Assumption 14 imposes conditions on the parameters A_0 and B_0 which ensure positivity and contracting or stable stochastic behavior for the data generated by our dynamic model in (7)–(9).

Assumption 12. *The sequence* $\{\boldsymbol{X}_t\}_{t=1,...,T}$ *is generated by (7)-(9) under some* $(\boldsymbol{\Omega}_0, A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ *for every* $t = 1, ..., T$.

Assumption 13. $\boldsymbol{\Omega}_0$ *is positive definite and* $(A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ *lies in a compact parameter space* Θ *ensuring that* $\hat{\boldsymbol{\Sigma}}_t$ *is positive definite almost surely, uniformly over* Θ, *where* $\boldsymbol{\nu}_i > i + 1 \ \forall \ i$, *and* $\boldsymbol{\mu}_j > K + 1 \ \forall \ j$.

Assumption 14. $A_0 \geq 0$, $B_0 \geq 0$ *and* $|A_0 + B_0| < 1$.

Under the conditions laid down in assumptions 12-14, we can obtain the strong consistency of the sample average $\hat{\boldsymbol{\Omega}} = n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_t$ with respect to $\boldsymbol{\Omega}_0$ by application of the ergodic theorem.

Proposition 15. *Let assumptions 12-14 hold. Then* $\hat{\boldsymbol{\Omega}} = n^{-1} \sum_{t=1}^{n} \boldsymbol{X}_t \xrightarrow{a.s.} \boldsymbol{\Omega}_0$ *as* $T \to \infty$.

We now turn to the invertibility of the filtering equation (9). Following the literature (e.g. Straumann and Mikosch (2006) and Wintenberger (2013)), invertibility ensures that the filter 'forgets' the incorrect initialization; i.e. the filtered sequence $\{\hat{\boldsymbol{V}}_t\}_{t\in\mathbb{N}}$ initialized at some point $\hat{\boldsymbol{V}}_1$ converges path-wise and exponentially fast to a unique stationary and ergodic limit sequence $\{\boldsymbol{V}_t\}_{t\in\mathbb{N}}$. This means that $c^t \|\hat{\boldsymbol{V}}_t - \boldsymbol{V}_t\| \xrightarrow{a.s.} 0$ as $t \to \infty$ for some $c > 1$, regardless of the fact that the initialization $\hat{\boldsymbol{V}}_1$ is almost surely incorrect (as the true $\boldsymbol{V}_1$ is an unobserved continuous random variable).

In the current setting, filter invertibility can be obtained by ensuring that the following conditions hold:

(i) stationarity of the data $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$;

(ii) a *logarithmic bounded moment* for $\boldsymbol{X}_t \ \forall\ t$;

(iii) a *contraction condition* for the filtering equation.

The stationarity of the data in (i), and the logarithmic moment in (ii), follow from assumptions 12-14. The contraction condition for the filtering equation, however, requires additional restrictions on the parameter space. Assumption 16 ensures that the filtered $\hat{\boldsymbol{V}}_t$ is positive definite and the contracting behavior of the stochastic sequence $\{\hat{\boldsymbol{V}}_t\}_{t\in\mathbb{N}}$.

Assumption 16. *The parameter space $\mathcal{A}\times\mathcal{B}$ is such that $A \geq 0$, $B \geq 0$ and $\sup_{B\in\mathcal{B}} |B| < 1$.*

Proposition 17 now establishes the invertibility of the filter $\hat{\boldsymbol{V}}_t$ for $\boldsymbol{V}_t$, and opens the door to the consistency of the MLE.

Proposition 17. *Let assumptions 12-16 hold. Then the filter $\{\hat{\boldsymbol{V}}_t\}_{t\in\mathbb{N}}$ is invertible.*

We are now ready to formulate our consistency result of the MLE $(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ for the vector $(A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$. Again, this MLE takes the form of a targeted two-step estimator as it depends on the first-step estimator for $\boldsymbol{\Omega}_0$. Additionally, as is common for filtering models, the log-likelihood depends directly on the properties of the filter $\hat{\boldsymbol{V}}_t$, which is itself a function of the estimated $\hat{\boldsymbol{\Omega}}_T$, the parameters A and B, and the initialization $\hat{V}_1$ as noted above. Specifically, we define the MLE as the maximizer of the plug-in loglikelihood $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})$,

$$(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) = \arg\max_{(A,B,\boldsymbol{\mu},\boldsymbol{\nu})} \sum_{t=2}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}). \tag{11}$$

Just as for the i.i.d. case, we apply standard theory of M-estimation to obtain the consistency of the MLE.

Theorem 18. *Let assumptions 12-16 hold. Then the targeted MLE $(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ defined in (11) satisfies*

$$(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) \xrightarrow{a.s.} (A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0) \quad \text{as } T \to \infty.$$

3.3 Test-sample inference

Given the large availability of data, we propose the use of an independent test sample to conduct inference on model parameters and specification. Specifically, for any given pair of model specifications or parameter restrictions, we calculate their loglikelihood on the test sample and obtain a test statistic by measuring the difference in the loglikelihoods of the two models. This test, often known as the *logarithmic scoring rule* (see e.g. Mitchell and Hall, 2005, Amisano and Giacomini, 2007, Bao et al. 2004, 2007, and Gneiting and Raftery, 2007), is a special case of the Diebold-Marianno (DM) test statistic introduced in Diebold and Marianno (1995, 1998). The DM test statistic compares the test-sample fit of two models and was further formalized by West (1996), McCracken (2000), Clark and McCracken (2001), Corradi, Swanson, and Olivetti (2001), and Chao, Corradi, and Swanson (2001), and further generalized in Giacomini and White (2006).

We now assume that additional data is available beyond time T, which we shall use as a test sample. In particular, we assume that $T' > T$ observations are available in total, and let $t = 1, ..., T$ be the *train sample* used for parameter estimation, while $H = T' - T$ observations ranging from $t = T + 1$ to $t = T'$ are used as a *test sample* to evaluate model performance. There are several sets of assumptions which can validate the use of a Diebold-Mariano test statistic. In general, the stationarity, invertibility and bounded moments discussed in the previous section set the stage for using the DM test statistic as a basis of comparison of performance of the autoregressive F-Riesz model against the performance of other models, which must also satisfy suitable regularity conditions. For completeness, we repeat well known conditions in Lemma 19 for the DM test statistic to be asymptotically Gaussian. This lemma is stated without proof.

Below we let $d_t = \ell_t^1 - \ell_t^2$ denote the difference between the loglikelihoods of two competing models at time t. For simplicity we adopt the shorthand notation $\ell_t = \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\Omega, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})$, for $t = 2, ..., T$. The null hypothesis assumes that the model fit the data equally well. This is stated in the remark below in terms of the difference in loglikelihoods d_t being a martingale difference sequence. Together with two bounded moments the DM test statistic can then be shown to satisfy Billingsley's central limit theorem. Weaker dependence conditions can however be considered since central limit

theorems are also available e.g. for mixing sequences and near epoch dependence sequences, at the expense of additional moment requirements.

Lemma 19. *Consider two models with log-likelihoods satisfying* $\mathbb{E}|\ell_t^i|^2 < \infty$*, for* $i = 1, 2$*. Under the null hypothesis* $\mathbb{H}_0 : \mathbb{E}(d_t|\mathcal{F}_{t-1}) = 0$ *we have,*

$$\sqrt{H}\frac{\bar{d}_H}{\hat{\sigma}_{\bar{d}}} \xrightarrow{d} N(0,1) \quad as\ H \to \infty,$$

where $\bar{d}_H$ *denotes the average loglikelihood difference based on* H *observations,* $\bar{d} = \frac{1}{H}\sum_{t=T+1}^{T'} d_t$*, and* $\hat{\sigma}_{\bar{d}}$ *of the standard deviation of* $\bar{d}$*.*[1]

Lemma 19 states that the DM test statistic based on the difference in loglikelihood contributions is asymptotically normal under known suitable regularity conditions. This remark considers two loglikelihoods, $\sum_{t=1}^{T} \ell_t^1$ and $\sum_{t=1}^{T} \ell_t^2$, obtained from two different specifications of the model. One important application of this test is in the regularization of the ordering of the variables in $\boldsymbol{X}_t$. Indeed, one can use the statistic above to test for significant differences in likelihood across any pair of orderings. More generally, one can find the optimal ordering which maximizes the test-sample loglikelihood among the finite set of $k!$ orderings available, and conduct inference on the significance of the achieved likelihood gains.

4 Simulation experiment

This section shows results of a Monte Carlo study which investigates the statistical properties of the maximum likelihood estimator of the parameters corresponding with the newly developed distributions. We perform five simulation experiments.

The first experiment aims to assess the small sample properties of the MLE for all degrees-of-freedom (DoF) parameters plus the parameters from the covariance matrix in a static model setting ($A = B = 0$). We simulate covariance matrices $\boldsymbol{X}_t$ of dimension 2 from the Riesz, inverse Riesz and F-Riesz distributions respectively. We set $\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{U}\boldsymbol{U}^\top$ with $\boldsymbol{L}$ ($\boldsymbol{U}$) the lower (upper) Cholesky matrix with elements $\boldsymbol{L}_{11} = 2.752$, $\boldsymbol{L}_{21} = 2.125$,

[1] It is well known that the sequence $\{d_t\}_{t=T+1}^{T'}$ is often autocorrelated, so that an autocorrelation-robust estimator $\hat{\sigma}_{\bar{d}}$ is usually employed; see e.g. Newey and West (1988).

$\boldsymbol{L}_{22} = 3.006$ and $\boldsymbol{U}_{11} = 2.247$, $\boldsymbol{U}_{21} = 1.588$, $\boldsymbol{U}_{22} = 3.682$. We set $\boldsymbol{\nu} = (10, 20)$ for the Riesz, as well as for the inverse Riesz distribution. For the F-Riesz distribution, we set $\boldsymbol{\mu} = (10, 15)$ and $\boldsymbol{\nu} = (15, 10)$.

The second experiment focuses on the estimation of the DoF parameters in a 5-variate case, where now the elements of $\boldsymbol{V}$ are estimated using a targeting approach as explained in Section 2.2. We set $\boldsymbol{\nu} = (10, 15, 20, 14, 12)$ for the (inverse) Riesz distribution, while $\boldsymbol{\mu} = (10, 15, 20, 14, 12)$ and $\boldsymbol{\nu} = (10, 15, 20, 12, 14)$ for the F-Riesz distribution.

Both simulation experiments are based on samples of 1000 observations from the three distributions. We then use Maximum Likelihood to estimate the parameters of interest. In addition, we estimate their standard errors by computing the inverse of the (negative) Hessian at the optimum. We replicate each experiment 1000 times.

Table 1 presents the results of the first two simulation experiments. Panels A and B correspond to the type I distributions, while Panels C and D show results for the type II distributions. In all panels, we find that all parameters are estimated near their true values. Comparing the Monte-Carlo standard error of the estimates (std column in Table 1) with the mean of the estimated standard error over all replications (mean(s.e.) column), we find that our computed standard errors fairly reflect estimation uncertainty. Only the true variability of the $\boldsymbol{\nu}$ parameters for the inverse Riesz and F-Riesz appears slightly higher than estimated by the usual standard errors, but the difference is minor.

The third and fourth simulation experiments are designed to address the different possible orderings of the variables in the system that make up the covariance matrix $\boldsymbol{X}_t$. To investigate this sensitivity in experiment three, we study the full enumeration approach for all available orderings in a low-dimensional setting. We simulate 1000 matrices $\boldsymbol{X}_t$ from a 5-variate $\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ distribution with $\boldsymbol{\nu} = (10, 20, 15, 18, 12)^\top$ and an arbitrarily chosen matrix $\boldsymbol{\Sigma}$. In each simulation run, we consider all 120 possible orderings of the variables in the system and estimate $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}$ using the targeting approach. We retain the ordering that has the highest maximized log-likelihood. We obtain such a final, optimal ordering $o^{(opt)}$ for each simulation run. To check whether the optimally estimated order coincides with the true DGP ordering, we compute the rank correlation between $o^{(opt)}$ and $o^{(dgp)}$ in each simulation run, and average across simulation runs. We also compute the difference between the maximum value of the optimized log-likelihood and the log-likelihood of the

Table 1: Parameter estimations of (inverse) Riesz and the F-Riesz distributions

This table shows Monte Carlo averages and standard deviations (in parentheses) of parameter estimates of simulated covariance matrices from the Riesz, inverse Riesz and F-Riesz distributions of dimension two and five. Panel A and B show results of the type I distributions, while panels C and D list results of the type II distributions. Panels A and C correspond to the bivariate case, where both the (Cholesky elements $\boldsymbol{L}_{11}$ ($\boldsymbol{U}_{11}$), $\boldsymbol{L}_{21}$ ($\boldsymbol{U}_{21}$) and $\boldsymbol{L}_{22}$ ($\boldsymbol{U}_{22}$) of) $\boldsymbol{V}_t$ as well as the degrees of freedom (DoF) parameters $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ are estimated. Panels B and D shows results of the five-variate case, where the elements of $\boldsymbol{V}$ are estimated by targeting in a first step, and the DoF parameters are estimated in a second step by maximum likelihood. The table reports the true values, the mean and standard deviation of the estimated coefficients, as well as the mean of the computed standard error. Results are based on 1000 Monte Carlo replications.

	Panel A: dimension 2					Panel B: dimension 5 (targeting)				
Distribution	Coef.	True	mean	std	mean(s.e.)	Coef.	True	mean	std	mean(s.e.)
Riesz I	L_{11}	2.752	2.752	0.019	0.019	ν_1	10	10.02	0.43	0.43
	L_{21}	2.125	2.126	0.025	0.026	ν_2	20	20.03	0.61	0.60
	L_{22}	3.006	3.006	0.015	0.015	ν_3	15	15.02	0.35	0.35
						ν_4	18	18.01	0.37	0.36
	ν_1	10	10.03	0.43	0.43	ν_5	12	12.01	0.18	0.19
	ν_2	20	20.01	0.60	0.60					
Inv Riesz I	L_{11}	2.752	2.752	0.019	0.020	ν_1	10	10.05	0.42	0.35
	L_{21}	2.125	2.126	0.027	0.027	ν_2	20	20.01	0.61	0.54
	L_{22}	3.006	3.006	0.018	0.018	ν_3	15	15.03	0.33	0.28
						ν_4	18	18.02	0.33	0.29
	ν_1	10	10.03	0.41	0.44	ν_5	12	12.00	0.18	0.11
	ν_2	20	20.04	0.61	0.61					
F-Riesz I	L_{11}	2.752	2.752	0.028	0.028	μ_1	10	10.02	0.57	0.55
	L_{21}	2.125	2.126	0.038	0.037	μ_2	15	15.03	0.64	0.62
	L_{22}	3.006	3.006	0.036	0.036	μ_3	20	20.05	0.73	0.72
						μ_4	14	14.01	0.40	0.40
	μ_1	10	10.08	0.87	0.86	μ_5	12	12.00	0.28	0.27
	μ_2	15	15.16	1.17	1.12					
	ν_1	15	15.23	1.85	1.87	ν_1	10	10.06	0.54	0.47
	ν_2	10	10.06	0.72	0.69	ν_2	15	15.06	0.73	0.66
						ν_3	20	20.10	0.96	0.93
						ν_4	12	12.05	0.43	0.39
						ν_5	14	14.12	0.75	0.69

DGP for each simulation run. Finally, we apply Algorithm 6 on 20 randomly chosen orders o^j and report the final order o^{opt} with the associated maximized log-likelihood.

Results of the third experiment are shown in Panel A of Table 2. The average rank correlation between the ordering with the highest log-likelihood and the true ordering is almost 1. Moreover, in more than 99% of the cases we are able to find the true ordering using the full set of 120 enumerated different orderings. Interestingly, applying the Heuristic gives exactly the same results, indicating that it works well in this simulation exercise. Panel

(continued from previous page)

	Panel C: dimension 2					Panel D: dimension 5 (targeting)				
Distribution	Coef.	True	mean	std	mean(s.e.)	Coef.	True	mean	std	mean(s.e.)
Riesz II	U_{11}	2.247	2.248	0.016	0.016	ν_1	10	10.01	0.15	0.14
	U_{21}	1.588	1.589	0.023	0.024	ν_2	20	20.01	0.39	0.40
	U_{22}	3.682	3.681	0.019	0.018	ν_3	15	15.02	0.36	0.35
						ν_4	18	18.03	0.53	0.54
	ν_1	10	10.02	0.28	0.29	ν_5	12	12.02	0.53	0.52
	ν_2	20	20.07	0.87	0.88					
Inv Riesz II	U_{11}	2.247	2.247	0.020	0.020	ν_1	10	10.01	0.15	0.07
	U_{21}	1.588	1.589	0.028	0.029	ν_2	20	20.00	0.42	0.34
	U_{22}	3.682	3.682	0.027	0.028	ν_3	15	15.02	0.35	0.28
						ν_4	18	17.98	0.58	0.47
	ν_1	10	10.02	0.29	0.29	ν_5	12	12.03	0.53	0.44
	ν_2	20	20.05	0.89	0.89					
F-Riesz II	U_{11}	2.247	2.247	0.025	0.025	μ_1	10	9.99	0.21	0.20
	U_{21}	1.588	1.588	0.039	0.039	μ_2	15	14.97	0.44	0.44
	U_{22}	3.682	3.682	0.043	0.041	μ_3	20	20.00	0.74	0.72
						μ_4	14	13.98	0.61	0.57
	μ_1	10	10.06	0.71	0.69	μ_5	12	12.00	0.67	0.67
	μ_2	15	15.14	1.39	1.40					
	ν_1	15	15.29	1.87	1.82	ν_1	10	10.06	0.43	0.29
	ν_2	10	10.08	0.70	0.69	ν_2	15	15.11	0.70	0.63
						ν_3	20	20.17	1.04	0.94
						ν_4	12	12.05	0.53	0.46
						ν_5	14	14.14	0.89	0.78

A.2 shows that the average difference in log-likelihood between the highest and the DGP log-likelihood equals 2.6 points on average across all simulations, which is very low.

In the fourth experiment, we test the performance of Algorithm 6 in improving the likelihood and in finding the true variable ordering. In each replication in this experiment, we simulate 1000 matrices $\boldsymbol{X}_t$ from a 15-variate $\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ distribution with ordering $o_{DGP} = (1, \ldots, 15)$ for a given $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}$. Note that for 15 variables there are more than 1.3 trillion possible orderings, such that trying all of them becomes impractical to impossible. We apply our algorithm for $p = 50$ different randomly chosen initial orderings. After obtaining the optimal order $o^{(opt)}$ from the Heuristic, we apply the Heuristic again using this order as the starting ordering. The final outcome is labeled as Heuristic(2). We report the average (across replications) of the rank correlation between $o^{(opt)}$ and $o^{(dgp)}$. In addition, we list the percentage of cases where (part of) the optimal ordering exactly matches (part of) the ordering of the DGP. Moreover, to check the sensitivity of the algorithm to the

Table 2: Simulation results on the ordering of variables

This table shows Monte Carlo results of two simulation experiments. Panels A.1 and A.2 present results for estimating parameters of a 5-dimensional $\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ distribution across 120 possible orderings of $\boldsymbol{X}$. In addition, it lists results of applying Heuristic I with 20 randomly chosen orders o^j. Panel A.1 shows the average rank correlation between the true ordering and 1) the ordering with the highest maximized log-likelihood, or 2) the final order o^{opt} in case of the Heuristic. Also the percentage of cases that this ordering matches the ordering of the DGP exactly is shown. Panel A.2 shows summary statistics of the range of the highest minus the log-likelihood of the DGP across all 120 possible orderings, summarized over simulation replications. Moreover, we also report the difference between the associated maximized log-likelihood of the final order of the Heuristic and the DGP. Panel B lists results of applying the Heuristic to a 15-variate $\mathcal{R}^I$ distribution with ordering $1, \ldots, 15$, using $p = 50$ randomly chosen initial orderings, labeled as Heu(1). Having obtained the final order and associated maximized log-likelihood, we repeat the Heuristic with the final order as the starting order. We label this as Heu(2). Panel B.1 shows the average rank correlation between the true ordering and the optimal ordering estimated using our algorithm once and twice. In addition, it presents the rank correlation between $o^{(opt)}$ for a pair of two random initial orderings, averaged across all pairs, and across all simulations. Finally, it lists the percentage of cases that part of the optimal ordering (1-5), (1-10) and the full ordering (1-15) matches exactly (part of the) ordering of the DGP. Panel B.2 reports summary statistics of the difference between the optimized log likelihood after applying Algorithm 6 and the log likelihood from the correct DGP. We run 1000 Monte Carlo replications for Panel A, while 500 Monte Carlo replications are used for panel B.

Panel A.1: Rank correlations

	FE	Heu
average rank corr $(o^{(opt)}, o^{(dgp)})$	0.999	0.999
perc correct rank $(o^{(opt)} = o^{(dgp)})$	0.994	0.994

Panel A.2: Summary statistics on $\log \mathcal{L}_{max} - \log \mathcal{L}_{DGP}$

	mean	sd	min	max
Full Enumeration	2.6	1.6	0.1	10.7
Heuristic	2.6	1.6	0.1	10.7

Panel B.1: Testing Heuristic I on $k = 15$: rank correlations

	Heu(1)	Heu(2)
average rank corr $(o^{(opt)}, o^{(dgp)})$	0.982	0.989
average rank corr of $o^{(opt)}$ across simulation pairs	0.807	0.951
perc correct rank $(o^{(opt)} = o^{(dgp)})$(1-5)	0.944	0.966
perc correct rank $(o^{(opt)} = o^{(dgp)})$(1-10)	0.676	0.784
perc correct rank $(o^{(opt)} = o^{(dgp)})$(1-15)	0.348	0.510

Panel B.2: Summary statistics on $\log \mathcal{L}_{opt} - \log \mathcal{L}_{DGP}$

	mean	sd	min	max
Heuristic(1)	7.10	3.85	-9.53	15.63
Heuristic(2)	8.37	3.05	1.05	16.38

p random initial orderings, we also compute the rank correlation of $o^{(opt)}$ for all pairs of two simulation replications, and then average over all pairs. The closer this number is to one, the smaller is the dependence on the precise initial random orderings. Also this fourth experiment is replicated 1000 times.

Panel B of Table 2 shows the results. Again, we find a very high rank correlation of 0.982 between the true order of the variables and the optimal estimated order. This implies that the algorithm works adequately. Second, the average (across all pairs) rank correlation for any combination of $o^{(opt)}$ based on two different initial random orderings is 0.802. This is high, and indicates that there is limited dependence of the final optimal ordering $o^{(opt)}$ on the $p = 50$ random initial orderings. Applying the Heuristic again increases this number to 0.951, hence the influence of the random initial ordering becomes very small. Furthermore, panel B.1 also indicates that repeating the Heuristic considerably improves the optimal ordering in the middle and in the end of the DGP ordering vector, as the percentages of correct rank increases from 0.676 (0.348) to 0.784 (0.510) respectively. As long as p is not chosen too small, the algorithm typically produces substantial likelihood increases. This pattern is corroborated by the difference between the maximized likelihood for all 50 different initial orderings and the maximized likelihood based on the DGP ordering ($\log \mathcal{L}_{dgp}$. The range of likelihoods is only 25 points wide (differences ranging form -9.63 to 15.63), and decreases if we repeat the Heuristic.

This is also illustrated by Figure 2, which shows a histogram of log likelihoods of the $\mathcal{R}^I$ distribution. The blue bars show $\log \mathcal{L}$ of the Riesz distribution of 50 randomly chosen initial orderings. These exhibit substantial dispersion, and are clearly inferior to the maximized likelihood value for the DGP ordering (red bar). However, after applying Algorithm 6 to the 50 initial random orderings twice, we obtain twice 50 optimal orderings $o^{(opt)}$. The latter are plotted with the green and yellow bars and are clearly much more concentrated, and close to the maximized log-likelihood value corresponding to the DGP ordering. The algorithm thus succeeds very well in eliminating much of the randomness of the initial orderings, and getting close to the optimal ordering.

The last simulation experiment investigates the statistical gain of the F-Riesz distribution over the matrix F distribution. Guided by the empirical application, we focus on a 5-variate F-Riesz I distribution with degrees of freedom vectors $\boldsymbol{\mu} = (18.7, 35.8, 58.2, 89.4, 143.9)^\top$ and $\boldsymbol{\nu} = (22.8, 24.3, 28.6, 22.3, 18.2)^\top$. We define $\bar{\mu} = 69.2$ and $\bar{\nu} = 23.3$ as the average values of the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ respectively and $\boldsymbol{\mu}_{range} = \boldsymbol{\mu} - \bar{\mu}\boldsymbol{\iota}_k$ and $\boldsymbol{\nu}_{range} = \boldsymbol{\nu} - \bar{\nu}\boldsymbol{\iota}_k$. The simulation experiment now consists of the following steps. First, we simulate 1000 matrices $\boldsymbol{X}_t$ from a $\mathcal{FR}^I(\boldsymbol{\Sigma}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$ with $\tilde{\boldsymbol{\mu}} = \bar{\mu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\mu}_{range}$ and

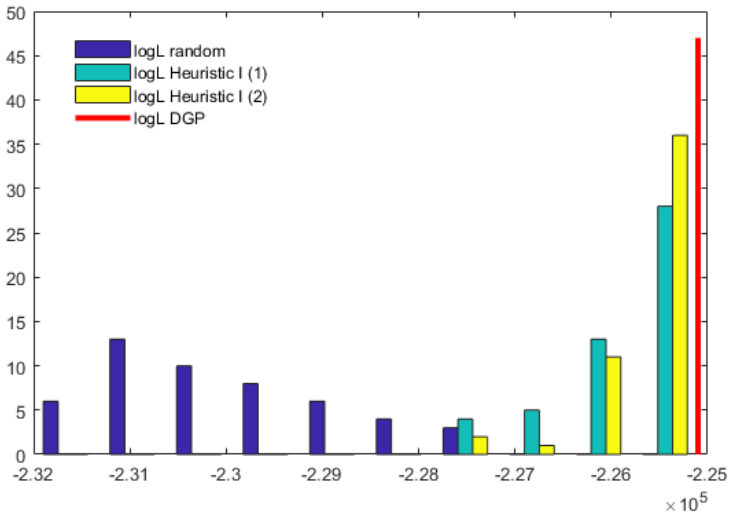

Figure 2: Heuristic I: log-likelihood values of the $\mathcal{R}^I$ distribution
This figure shows a histogram of $\log \mathcal{L}$ values of the 15-variate $\mathcal{R}^I$ distribution. The blue bars correspond to the likelihood values of 50 randomly chosen order *before* applying Heuristic I, while the green (yellow) bars show $\log \mathcal{L}$ values *after* applying the heuristic (twice). The red bar denotes the log-likelihood with the correct ordering from the DGP.

$\tilde{\boldsymbol{\nu}} = \bar{\nu}\boldsymbol{\iota}_k + \lambda \boldsymbol{\nu}_{range}$ for $\lambda = (0, 0.02, \ldots, 0.08, 0.10)$. Note that if $\lambda = 0$, the $\mathcal{FR}^I$ distribution collapses to a matrix-F distribution with $\bar{\mu}$ and $\bar{\nu}$ degrees of freedom. Second, we estimate $\boldsymbol{\Sigma}$ (using the targeting approach) and the degrees of freedom parameters assuming a matrix F or $\mathcal{FR}^I$ distribution. For each λ we test the null-hypotheses $\boldsymbol{\mu} = \bar{\mu}\boldsymbol{\iota}_k$ and $\boldsymbol{\nu} = \bar{\nu}\boldsymbol{\iota}_k$. This boils down to Likelihood-Ratio test with $2*k-2$ degrees of freedom. We repeat this exercise 1000 times.

Table 3 shows the results. In Panel A, we see that if we simulate from a matrix-F distribution (i.e. $\lambda = 0$), the likelihood ratio test has been rejected in 8.4 % of all cases. Further, when we deviate slightly from the matrix-F setting, we immediately reject the null-hypothesis of a scalar μ and ν in all cases. Panel B lists that the correct matrix-F parameters are indeed estimated back on average. Also the average parameter estimates of the F-Riesz I corresponds to the simulated values of 69.2 and 23.25.

5 Empirical application

In this section, we apply the F-Riesz distribution to an empirical data set of 30 U.S. equities from the S&P 500 index over the period January 2, 2001, until July 31, 2014, a total of 3415

Table 3: The matrix F vs the F-Riesz distributions

This table shows Monte Carlo results on the difference between the F-Riesz and the matrix-F distribution. Panel A lists results on simulating 1000 matrices from a $\mathcal{FR}^I(\boldsymbol{\Sigma}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\nu}})$ distribution with $\tilde{\boldsymbol{\mu}} = \bar{\mu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\mu}_{range}$ and $\tilde{\boldsymbol{\nu}} = \bar{\nu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\nu}_{range}$ for $\lambda = (0, 0.02, \ldots, 0.08, 0.10)$ with $\bar{\mu} = 69.2$, $\bar{\nu} = 23.3$, $\boldsymbol{\mu}_{range} = \boldsymbol{\mu} - \bar{\mu}\boldsymbol{\iota}_k$ and $\boldsymbol{\nu}_{range} = \boldsymbol{\nu} - \bar{\nu}\boldsymbol{\iota}_k$ and $\boldsymbol{\mu} = (18.7, 35.8, 58.2, 89.4, 143.9)^\top$ and $\boldsymbol{\nu} = (22.8, 24.3, 28.6, 22.3, 18.2)^\top$. We estimate the parameters assuming a matrix-F or $\mathcal{FR}^I$ distribution. For each value of λ we perform a likelihood ratio test on the null-hypothesis $\boldsymbol{\mu} = \bar{\mu}\boldsymbol{\iota}_k$ and $\boldsymbol{\nu} = \bar{\nu}\boldsymbol{\iota}_k$. Panel A lists the percentage rejections of this hypothesis for different values of λ. Further, Panel B reports results on the estimated degrees-of-freedom parameters of the matrix- F and/or F-Riesz I distribution for the case $\lambda = 0$. The panel reports the true values, the mean and standard deviation of the estimated coefficients. All results are based on 1000 Monte Carlo replications.

Panel A: Matrix F vs F-Riesz I						
λ	0	0.02	0.04	0.06	0.08	0.10
rejection rate	0.084	0.126	0.311	0.594	0.839	0.980

Panel B: DoF parameters when $\lambda = 0$						
matrix-F		$\bar{\mu}$	$\bar{\nu}$			
	true	69.20	23.25			
	mean	69.25	23.33			
	sd	5.72	0.63			
F-Riesz I		$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	$\boldsymbol{\mu}_3$	$\boldsymbol{\mu}_4$	$\boldsymbol{\mu}_5$
	true	69.20	69.20	69.20	69.20	69.20
	mean	69.60	69.47	69.54	69.44	69.42
	sd	7.26	6.52	6.29	6.05	5.83
		$\boldsymbol{\nu}_1$	$\boldsymbol{\nu}_2$	$\boldsymbol{\nu}_3$	$\boldsymbol{\nu}_4$	$\boldsymbol{\nu}_5$
	true	23.25	23.25	23.25	23.25	23.25
	mean	23.36	23.32	23.34	23.40	23.42
	sd	0.99	0.91	0.99	1.17	1.52

trading days. We extract transaction prices from the Trade and Quote (TAQ) database and clean the high-frequency data in line with Brownlees and Gallo (2006) and Barndorff-Nielsen et al. (2009). After this cleaning procedure we construct realized covariance matrices $\boldsymbol{X}_t$ using 5 minute returns with subsampling.

We consider six different matrix distributions with a time-varying mean $\boldsymbol{V}_t$ for the realized covariance matrices: the Wishart, the Riesz, the inverse Wishart, the inverse Riesz, the Matrix-F, and the F-Riesz distribution. For the Riesz related distributions, we only consider the type I version. The type II versions of these distributions yield very similar results. The dynamic specification is as in (7)–(9), with only equation (7) modified for the distribution at hand. We use the two-step targeting approach from Section 2.2 to estimate $\boldsymbol{\Omega}$, and the algorithm from Section 2.3 with $p = 50$ on the CAFr model to determine the order of the different stocks in the system. In order to effectively compare the likelihoods, we use the same optimal order from this model to the conditional autoregressive Wishart

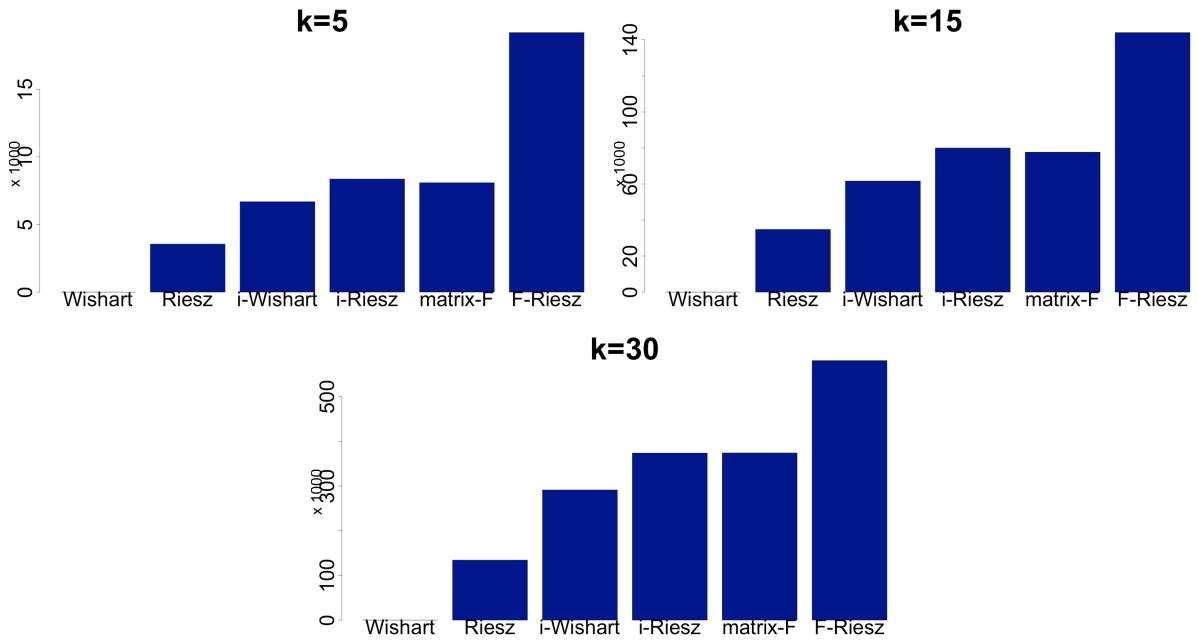

Figure 3: AIC improvements
The figure shows the difference between the AIC of the Wishart and that of the other distributions for the models in Table 4.

(CAW) and inverse Wishart (CAiW) models. Table 4 and Figure 3 report the results for full-sample estimation.

The results provide three main take-aways. First, the maximized log-likelihood shows that the model with the F-Riesz distribution is superior against all the other specifications. This is most clearly seen in Figure 3, which shows the AIC improvements of all models compared to the Wishard specification. The gain of the F-Riesz specification increases substantially with the dimension of the system as can be seen from the scales of the vertical axes of the different panels. For example the difference between the F-Riesz and the matrix-F distribution equals 5000, 30,000 and 90,000 log-likelihood points for dimension 5, 15 and 30 respectively. This result is striking and suggests substantial heterogeneity in the tails. The AIC values further support the usefulness of the F-Riesz distribution: the large differences in likelihoods persist if we correct for the number of estimated parameters.

Second, tail heterogeneity and tail fatness both play an important role at all levels of the analysis. When relaxing the Wishart to the Riesz specification, the AIC improves substantially for all dimensions considered, thus underlining the importance of tail-heterogeneity; see Figure 3. The same holds when relaxing the inverse Wishart to the

Table 4: Parameter estimates, likelihoods and information criteria

This table reports maximum likelihood parameter estimates of the conditional autoregressive model (7)–(9), assuming a Wishart, Riesz I, Inverse Wishart, Inverse Riesz I, matrix-F, or F-Riesz I distribution in (7). Data consist of realized covariance matrices observed over the period January 2, 2001 until July 31, 2014 ($T = 3415$ trading days). Panels A and B list results for a randomly chosen subset of 5 and 15 different assets, respectively. Panel C shows the results for the full set of 30 assets. Standard errors are provided in parentheses. We report the likelihood $\mathcal{L}$, the AIC and the number of estimated parameters.

Distribution	A	B	$\mu_{\min}$	$\mu_{\max}$	$\nu_{\min}$	$\nu_{\max}$	$\mathcal{L}$	AIC	♯para
Panel A: AA/BA/CAT/GE/KO									
Wishart	0.313	0.669			21.17		-20,508	41,022	3
	(0.005)	(0.005)			(0.12)				
Riesz I	0.294	0.686			7.87	27.25	-18,727	37,468	7
	(0.005)	(0.005)			(0.18)	(0.31)			
iWishart	0.244	0.744			23.64		-17,164	34,335	3
	(0.004)	(0.004)			(0.11)				
iRiesz I	0.240	0.748			12.18	28.01	-16,320	32,654	7
	(0.004)	(0.004)			(0.24)	(0.28)			
F	0.259	0.729	74.51		33.72		-16,461	32,930	4
	(0.004)	(0.005)	(2.22)		(0.45)				
FRiesz I	0.205	0.783	16.94	143.23	15.26	32.61	-10,900	21,824	12
	(0.004)	(0.004)	(0.62)	(4.01)	(0.24)	(0.94)			
Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM									
Wishart	0.235	0.753			39.32		104,666	-209,327	3
	(0.001)	(0.001)			(0.07)				
Riesz I	0.209	0.777			7.46	53.60	122,079	-244,123	17
	(0.001)	(0.001)			(0.17)	(0.28)			
iWishart	0.167	0.826			44.52		135,490	-270,974	3
	(0.001)	(0.001)			(0.06)				
iRiesz I	0.158	0.834			11.23	51.04	144,673	-289,313	17
	(0.001)	(0.001)			(0.22)	(0.30)			
F	0.179	0.814	130.38		63.09		143,506	-287,004	4
	(0.001)	(0.001)	(1.19)		(0.26)				
FRiesz I	0.132	0.862	15.36	223.15	14.35	57.89	176,713	-353,361	32
	(0.001)	(0.001)	(0.39)	(2.28)	(0.32)	(0.71)			
Panel C: All equities ($k = 30$)									
Wishart	0.173	0.821			57.04		725,787	-1,451,568	3
	(0.001)	(0.001)			(0.05)				
Riesz I	0.157	0.835			4.48	78.74	792,880	-1,585,696	32
	(0.001)	(0.001)			(0.10)	(0.28)			
iWishart	0.117	0.880			66.35		871,345	-1,742,683	3
	(0.000)	(0.000)			(0.04)				
iRiesz I	0.107	0.888			7.35	76.06	912,590	-1,825,117	4
	(0.000)	(0.000)			(0.12)	(0.23)			
F	0.123	0.875	180.14		92.405		912,732	-1,825,400	32
	(0.000)	(0.000)	(0.72)		(0.17)				
FRiesz I	0.087	0.910	10.55	277.05	8.98	83.14	1,016,003	-2,031,882	62
	(0.000)	(0.000)	(0.27)	(1.41)	(0.18)	(0.67)			

inverse Riesz. Tail fatness is also clearly important: the AIC improvement for the matrix-F is large compared to the Wishart. With only 2 parameters, the matrix-F succeeds in having a similar AIC improvement as the inverse Riesz with 5, 15, and 30 parameters for the different values of k, respectively. Interestingly, it seems that both the heterogeneity effects from the Wishart to the Riesz and from the inverse Wishart to the inverse Riesz accumulate when generalizing the matrix-F to the F-Riesz distribution: the jump in likelihood from the matrix-F to the F-Riesz is in all cases more than the jump from the Wishart to the Riesz plus that of the inverse Wishart to the inverse Riesz. This is the more interesting result given that the matrix-F already heavily outperforms the Wishart, inverse Wishart, Riesz, and to a lesser extent also the inverse Riesz distributions. The F-Riesz distribution thus seems a major step forward in capturing the empirical features of realized covariance matrices by allowing for tail heterogeneity via both $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$.

Third, the importance of allowing for tail heterogeneity is confirmed by looking at the estimates of the degrees of freedom parameters. For conciseness, the table only reports the minima and maxima of the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. Still, the picture is clear. For example, the estimate of μ in Panel A for the matrix-F is around 75, while the elements of $\boldsymbol{\mu}$ of the F-Riesz distribution vary from around 17 to 144. The pattern persists for the other panels in the table, as well as for the $\boldsymbol{\nu}$ parameters. The Riesz and F-Riesz distributions also solve an empirical puzzle for the (inverse) Wishart and matrix-F distributions. As we can see in the table, increasing the dimension of the system from 5 to 15 to 30 augments the estimated degrees of freedom for the (inverse) Wishart and matrix-F. We can understand this by looking at the spreads of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ for the F-Riesz distribution. These reveal that the tail fatness (low $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ values) persists across dimensions, as $\mu_{\min}$ and $\nu_{\min}$ remain relatively constant across panels A, B, and C. By contrast, $\mu_{\max}$ and $\nu_{\max}$ increase if we consider more stocks, indicating that some of the realized volatilities exhibit thinner tail behavior. As the (inverse) Wishart and matrix-F because of their one and two parameter set-up can only accommodate this by using some sort of average degrees of freedom value across assets, we see the empirical increase in the degrees of freedom for these distributions. By contrast, the F-Riesz (and also the (inverse) Riesz) distributions do not show this behavior.

Fourth, we see that the heterogeneity biases discussed in the previous point spill over into biases in the estimated persistence of $\boldsymbol{X}_t$. The B of the F-Riesz distribution is higher

across all dimensions than that of the other models, while its A parameter is lower. This results in a much smoother pattern of $\boldsymbol{V}_t$ for the F-Riesz. Again, this is the accumulation of two effects: fat tails of $\boldsymbol{X}_t$, and tail heterogeneity. Fatter tails for $\boldsymbol{X}_t$ in the model imply the dynamics of $\boldsymbol{V}_t$ react less violently to incidental outliers in $\boldsymbol{X}_t$, similar to the effect of using a Student's t distribution in a GARCH model. This explains why the F-Riesz results in more persistence than the Riesz or inverted Riesz, for instance. The second effect is that of tail heterogeneity. Because the (inverse) Wishart and matrix-F only have one or two degrees of freedom parameters, they miss heavy-tailed behavior in part of the assets (see also the previous point). This missed out tail-fatness in some dimension makes the dynamics less persistent due to in increased sensitivity to outliers; compare for instance for the univariate volatility setting. This explains why the F-Riesz and Riesz have higher persistence compared to the matrix-F and Wishart, respectively.

Beyond a full-sample analysis, we also apply our newly developed distributions in an out-of-sample exercise, by performing 1-step ahead density forecasts, using again the model of (7). These density forecasts immediately depend on the 1-step ahead forecasts of $\boldsymbol{V}_t$.

We use a moving-window approach in the forecasting exercise with an in-sample period of 1,000 observations. This corresponds roughly to four calendar years. The out-of-sample period contains $P = 2415$ observations including the Great Financial Crisis, which therefore constitutes an important test for the robustness of the model. We re-estimate our model repeatedly after each 50 observations, which roughly corresponds to bi-monthly updating of the parameters. In case of the (i)Riesz and F-Riesz I distributions, we apply Heuristic I once on the first moving window to the CAFr model, and keep this ordering fixed throughout the whole exercise and for all different distributions.

We use the log scoring rule (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007) to differentiate between the density forecasts of the models. Define the difference in log score between the two density forecasts M_1 and M_2 corresponding to the realized covariance matrix $\boldsymbol{X}_t$ as

$$d_{ls,t} = S_{ls,t}(\boldsymbol{X}_t, M_1) - S_{ls,t}(\boldsymbol{X}_t, M_2), \tag{12}$$

for $t = R, R+1, \ldots T-1$ with R the length of the estimation window and $S_{ls,t}(\boldsymbol{X}_t, M_j)$

Table 5: Out-of-sample log-scores

This table shows the mean of log scores, defined in (13), based on 1-step ahead predictions of the covariance matrix, according to the Conditional Autoregressive model, assuming a Wishart, Riesz I, Inverse Wishart, Inverse Riesz I, matrix-F, or a F-Riesz I distribution. Panel A shows results of the model applied to five assets, Panel B and C show results corresponding to 15 and to all 30 equities. The highest value of the predictive log-score across the models are marked bold. In addition, we report HAC based test-statistics on the difference in predictive ability (DM_{DF}) between the CAFr model and the other considered models. Positive statistics indicate that the CAFr model has superior density forecasts. The out-of-sample period goes from January 2005 until July 2014 and contains 2514 observations.

	Wishart	Riesz I	iWishart	iRiesz I	matrix-F	Friesz I
Panel A: AA/BA/CAT/GE/KO						
$S^{ls}(\boldsymbol{X}_t)$	-4.324	-3.843	-3.083	-2.845	-2.921	**-1.363**
DM_{DF}	(13.22)	(11.36)	(23.36)	(21.18)	(25.44)	
Panel B: k = 15						
$S^{ls}(\boldsymbol{X}_t)$	49.38	54.33	59.75	62.28	61.42	**71.01**
DM_{DF}	(13.56)	(13.14)	(20.33)	(17.91)	(31.89)	
Panel C: All equities (k = 30)						
$S^{ls}(\boldsymbol{X}_t)$	277.4	297.1	325.6	337.8	334.8	**366.2**
DM_{DF}	(21.68)	(20.11)	(25.85)	(22.66)	(31.82)	

$(j = 1, 2)$ the log score of the density forecast corresponding to model M_j at time t,

$$S_{ls,t}(\boldsymbol{X}_t, M_j) = \log p_t(\boldsymbol{X}_t | \boldsymbol{V}_t, \mathcal{F}_{t-1}, M_j), \tag{13}$$

where $p_t(\cdot)$ is the probability distribution function of our six considered distributions. The null hypothesis of equal predictive ability is given by $H_0 : \mathbb{E}[d_{ls}] = 0$ for all P out-of-sample forecasts. This null can be tested by means of a Diebold and Mariano (1995) (DM) statistic given by

$$DM_{ls} = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2/N}}, \tag{14}$$

with $\bar{d}$ the out-of-sample average of the log score differences and $\hat{\sigma}^2$ a HAC-consistent variance estimator of the true variance σ^2 of $d_{ls,t}$. Under the assumptions of the framework of Giacomini and White (2006), DM_{ls} asymptotically follows a standard normal distribution. A significantly positive value means that model M_1 has a superior forecast performance over model M_2.

Table 5 shows the average values of the log score over the out-of-sample period for a

set of five, 15, and the complete set of 30 assets. In addition, we provide corresponding t-statistics for the difference in the log predictive density scores of the realized covariance matrix between the CAFr model and competing models. Positive values means that the density forecasts of the former model is superior against its competitor. The results heavily reinforce our earlier full-sample analysis as the F-Riesz distribution again clearly outperform the other distributions, even at a 1% significance level. This result holds for all three considered dimensions. The differences in the mean log-score increases by dimension. In sum, the F-Riesz distribution outperforms both in-sample and out-of-sample the other competitors, indicating that it is able - in contrast to the (inverse) Riesz distributions - to capture the tail heterogeneity of realized covariance matrices.

6 Conclusions

We introduce the new F-Riesz distributions for econometric modeling of matrix valued random variables. The distribution is obtained by mixing the Riesz distribution (Hassairi and Lajmi, 2001) with an inverse Riesz distribution (Tounsi and Zine, 2012) and allows for much more heterogeneity in tail behavior compared to well-known fat-tailed distributions like the Wishart, inverse Wishart, or the matrix-F distribution. While the latter distributions depend on one or two degrees-of-freedom parameters, our new distribution allows the degrees of freedom parameters to be vectors. These parameters can easily be estimated by a two-step targeted maximum likelihood approach, even in high dimensional settings.

An empirical application to realized covariance matrices of 30 U.S. stocks over 14 years shows a remarkably high increase of the likelihood of the developed F-Riesz distribution compared to the (inverse) Riesz distributions. The degrees of freedom parameters are significantly over the different coordinates. The new distribution even outperforms the fat-tailed matrix-F distribution, applied in financial econometrics by Opschoor et al. (2018). The margin of outperformance is as wide as that of switching from a Wishart to the matrix-F, and therefore empirically highly significant, essentially doubling the gain already obtained by using the fat-tailed matrix-F. Out-of-sample we confirm these results by means of one-step ahead density forecasts of the time-varying realized covariance matrix. Again the

F-Riesz distribution is superior against the competitors by a very wide margin. Overall these results show that there is strong heterogeneity of tail behavior of realized covariance matrices, and that the F-Riesz distribution is a helpful vehicle to obtain better empirical models.

References

Abadir, K. and J. Magnus (2005). *Matrix Algebra*. Cambridge University Press.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics 25*(2), 177–190.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 529–626.

Anderson, T. W. (1962). An introduction to multivariate statistical analysis. Technical report, Wiley New York.

Andersson, S. A. and T. Klein (2010). On Riesz and Wishart distributions associated with decomposable undirected graphs. *Journal of Multivariate Analysis 101*(4), 789–810.

Asai, M. and M. K. So (2021). Quasi-maximum likelihood estimation of conditional autoregressive Wishart models. *Journal of Time Series Analysis 88*, in press.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2009). Realized kernels in practice: trades and quotes. *Econometrics Journal 12*(3), 1–32.

Barndorff-Nielsen, O. E. and N. Shephard (2004, 2020/12/29). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica 72*(3), 885–925.

Bollerslev, T., J. Li, A. J. Patton, and R. Quaedvlieg (2020). Realized semicovariances. *Econometrica 88*(4), 1515–1551.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2018). Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions. *Journal of Econometrics 207*(1), 71–91.

Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications 121*(10), 2331–2360.

Brownlees, C. and G. Gallo (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics and Data Analysis 51*(4), 2232–2245.

Callot, L. A., A. B. Kock, and M. C. Medeiros (2017). Modeling and forecasting large realized covariance matrices and portfolio choice. *Journal of Applied Econometrics 32*(1), 140–158.

Chen, R., D. Yang, and C.-H. Zhang (2021). Factor models for high-dimensional tensor time series. *Journal of the American Statistical Association*, 1–23.

Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics 26*(6), 922–947.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Díaz-García, J. A. (2013). A note on the moments of the Riesz distribution. *Journal of Statistical Planning and Inference 143*(11), 1880–1886.

Díaz-García, J. A. (2016). Riesz and beta-Riesz distributions. *Austrian Journal of Statistics 45*(2), 35–51.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*(2), 363–375.

Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica 74*(6), 1545–1578.

Golosnoy, V., B. Gribisch, and R. Liesenfeld (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics 167*(1), 211–223.

Gribisch, B. and J. Hartkopf (2022). Modeling realized covariance measures with heterogeneous liquidity: A generalized matrix-variate wishart state-space model. *Journal of Econometrics*, in press.

Gross, K. I. and D. S. P. Richards (1987). Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions. *Transactions of the American Mathematical Society 301*(2), 781–811.

Hassairi, A. and S. Lajmi (2001). Riesz exponential families on symmetric cones. *Journal of Theoretical Probability 14*, 927–948.

Hidot, S. and C. Saint-Jean (2010). An expectation–maximization algorithm for the wishart mixture model: Application to movement clustering. *Pattern Recognition Letters 31*(14), 2318–2324.

Jin, X. and J. Maheu (2013). Modeling realized covariances and returns. *Journal of Financial Econometrics 11*(2), 335–369.

Jin, X. and J. Maheu (2016). Bayesian semiparametric modeling of realized covariance matrices. *Journal of Econometrics 192*(1), 19–39.

Konno, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution. *Annals of the Institute of Statistical Mathematics 43*, 157–165.

Louati, M. and A. Masmoudi (2015). Moment for the inverse Riesz distributions. *Statistics & Probability Letters 102*, 30–37.

Lunde, A., N. Shephard, and K. Sheppard (2016). Econometric analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice. *Journal of Business & Economic Statistics 34*(4), 504–518.

Markowitz, H. M. (1991). Foundations of portfolio theory. *The journal of finance 46*(2), 469–477.

Mitchell, J. and S. Hall (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESR fan-charts of inflation. *Oxford Bulletin of Economics and Statistics 67*(s1), 995–1033.

Noureldin, D., N. Shephard, and K. Sheppard (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 27*(6), 907–933.

Oh, D. and A. Patton (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business and Economic Statistics 35*(1), 139–154.

Oh, D. and A. Patton (2018). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business and Economic Statistics 36*(2), 181–195.

Olkin, I., H. Rubin, et al. (1964). Multivariate Beta distributions and independence properties of the Wishart distribution. *The Annals of Mathematical Statistics 35*(1), 261–269.

Opschoor, A., P. Janus, A. Lucas, and D. van Dijk (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics 36*(4), 643–657.

Opschoor, A., A. Lucas, I. Barra, and D. J. van Dijk (2020). Closed-Form Multi-Factor Copula Models with Observation-Driven Dynamic Factor Loadings. *Journal of Business & Economic Statistics*. 10.1080/07350015.2020.1763806.

Patton, A. (2009). Copula-based models for financial time series. In *Handbook of financial time series*, pp. 767–785. Springer.

Saint-Jean, C. and F. Nielsen (2014). Hartigan's method for k-mle: Mixture modeling with wishart distributions and its application to motion retrieval. In *Geometric theory of information*, pp. 301–330. Springer.

Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics 34*(5), 2449–2495.

Tounsi, M. and R. Zine (2012). The inverse Riesz probability distribution on symmetric matrices. *Journal of Multivariate Analysis 111*, 174–182.

van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

Wang, D., X. Liu, and R. Chen (2019). Factor models for matrix-valued high-dimensional time series. *Journal of econometrics 208*(1), 231–248.

Wintenberger, O. (2013). Continuous invertibility and stable QML estimation of the EGARCH(1, 1) model. *Scandinavian Journal of Statistics 40*(4), 846–867.

Online Appendix to: Tail Heterogeneity for Dynamic Covariance-Matrix-Valued Random Variables Using the F-Riesz Distribution*

Francisco Blasques[a,b] Andre Lucas[a,b] Anne Opschoor[a,b] Luca Rossini[c]

[a]Vrije Universiteit Amsterdam, The Netherlands

[b]Tinbergen Institute, The Netherlands

[c]Queen Mary University of London, United Kingdom

A Proofs

We start with a simple lemma that states some of the manipulation rules for power weighted determinants. We note these rules can differ from those applicable to standard determinants. For instance, while properties (i)–(iii) and (v) are intuitive, property (iv) is an important difference with the standard determinant case. In particular, $|\boldsymbol{Y}|_{\boldsymbol{\nu}} \neq |\boldsymbol{Y}^{-1}|_{-\boldsymbol{\nu}}$ in general, whereas for a positive definite $\boldsymbol{Y}$, non-zero ν, and a regular determinant we have $|\boldsymbol{Y}|^{\nu} = |\boldsymbol{Y}^{-1}|^{-\nu}$.

Lemma 20. *Given a scalar ν, a vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top \in \mathbb{R}^{k\times 1}$, a vector of ones $\boldsymbol{\iota}_k \in \mathbb{R}^{k\times 1}$, and $\boldsymbol{Y} \in \mathbb{R}^{k\times k}$ a positive definite matrix, then we have the following identities.*

(i) If $\boldsymbol{\nu} = \nu \cdot \boldsymbol{\iota}_k$, then $|\boldsymbol{Y}|_{\nu\cdot\boldsymbol{\iota}_k} = {}_U|\boldsymbol{Y}|_{\nu\cdot\boldsymbol{\iota}_k} = |\boldsymbol{Y}|^{\nu}$. As special case, when $\nu = 1$, we have $|\boldsymbol{Y}|_{\boldsymbol{\iota}_k} = {}_U|\boldsymbol{Y}|_{\boldsymbol{\iota}_k} = |\boldsymbol{Y}|$.

(ii) Let $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \in \mathbb{R}^{k\times 1}$ be two vectors of constants, then we have $|\boldsymbol{Y}|_{\boldsymbol{\nu}_1} \cdot |\boldsymbol{Y}|_{\boldsymbol{\nu}_2} = |\boldsymbol{Y}|_{\boldsymbol{\nu}_1+\boldsymbol{\nu}_2}$, and ${}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}_1} \cdot {}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}_2} = {}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}_1+\boldsymbol{\nu}_2}$.

(iii) $(|\boldsymbol{Y}|_{\boldsymbol{\nu}})^{-1} = |\boldsymbol{Y}|_{-\boldsymbol{\nu}}$, and $({}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}})^{-1} = {}_U|\boldsymbol{Y}|_{-\boldsymbol{\nu}}$.

(iv) $|\boldsymbol{Y}|_{\boldsymbol{\nu}} = {}_U|\boldsymbol{Y}^{-1}|_{-\boldsymbol{\nu}}$.

(v) If $\boldsymbol{L}, \boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$, where $\boldsymbol{\Sigma}$ is positive definite with lower triangular Cholesky decomposition $\boldsymbol{L}$ such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, then $|\boldsymbol{L}^{-1}\boldsymbol{Y}(\boldsymbol{L}^{-1})^\top|_{\boldsymbol{\nu}} = |\boldsymbol{Y}|_{\boldsymbol{\nu}} \cdot |\boldsymbol{\Sigma}|_{-\boldsymbol{\nu}}$.
Similarly, if $\boldsymbol{U}$ is the upper triangular Cholesky decomposition of $\boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{U}^\top$, then ${}_U|\boldsymbol{U}^{-1}\boldsymbol{Y}(\boldsymbol{U}^{-1})^\top|_{\boldsymbol{\nu}} = {}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}} \cdot {}_U|\boldsymbol{\Sigma}|_{-\boldsymbol{\nu}}$.

Proof of Lemma 20:

[1]*: Blasques thanks the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.195.099.

Part (i): Note that in this case

$$|\boldsymbol{Y}|_{\nu\cdot\boldsymbol{\iota}_k} = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu} = \left(\prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2}\right)^{\nu} = \left(|\boldsymbol{L}|^2\right)^{\nu} = |\boldsymbol{Y}|^{\nu}, \tag{A.1}$$

where the last two equalities follow from the triangularity of $\boldsymbol{L}$ and the fact that $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{L}^\top$, and thus $|\boldsymbol{Y}| = |\boldsymbol{L}\boldsymbol{L}^\top| = |\boldsymbol{L}| \cdot |\boldsymbol{L}^\top| = |\boldsymbol{L}|^2$. The proof for the UPWD is similar. The special case $\nu = 1$ follows directly.

Part (ii): Note that

$$|\boldsymbol{Y}|_{\boldsymbol{\nu}_1}\ |\boldsymbol{Y}|_{\boldsymbol{\nu}_2} = \left(\prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_{1,i}}\right)\left(\prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_{2,i}}\right) = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2(\nu_{1,i}+\nu_{2,i})} = |\boldsymbol{Y}|_{\boldsymbol{\nu}_1+\boldsymbol{\nu}_2}\,. \tag{A.2}$$

The proof for the UPWD is similar.

Part (iii): Note that

$$(\,|\boldsymbol{Y}|_{\boldsymbol{\nu}}\,)^{-1} = \left(\prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_i}\right)^{-1} = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{-2\nu_i} = |\boldsymbol{Y}|_{-\boldsymbol{\nu}}\,. \tag{A.3}$$

The proof for the UPWD is similar.

Part (iv): Note that for positive definite $\boldsymbol{Y}$ we have $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{L}^\top$ where $\boldsymbol{L}$ is unique and lower triangular. Therefore, $\boldsymbol{Y}^{-1} = (\boldsymbol{L}^\top)^{-1}\boldsymbol{L}^{-1} = \boldsymbol{U}\boldsymbol{U}^\top$, with $\boldsymbol{U} = (\boldsymbol{L}^\top)^{-1}$ upper triangular and unique. We also note that the diagonal elements of $\boldsymbol{U} = (\boldsymbol{L}^\top)^{-1}$ are the inverse of the diagonal elements of $\boldsymbol{L}$. To see this, note that the diagonal elements of $\boldsymbol{U}$ and of $\boldsymbol{L}^{-1}$ are the same, as the first is the transpose of the latter. Also, we have

$$1 = \boldsymbol{L}_{i,\cdot}\,\boldsymbol{L}^{-1}_{\cdot,i} = \boldsymbol{L}_{i,i}(\boldsymbol{L}^{-1})_{i,i} \quad \Leftrightarrow \quad (\boldsymbol{L}^{-1})_{i,i} = \boldsymbol{U}_{i,i} = \frac{1}{\boldsymbol{L}_{i,i}}, \tag{A.4}$$

where $\boldsymbol{L}_{i,\cdot}$ and $\boldsymbol{L}_{\cdot,i}$ denote the ith row and column of $\boldsymbol{L}$, respectively, and where the first equality follows from the fact that $\boldsymbol{L}^{-1}$ is the inverse of $\boldsymbol{L}$, and second equality from the fact that both $\boldsymbol{L}$ and $\boldsymbol{L}^{-1}$ are lower triangular. As a result and using these definitions, we obtain

$$|\boldsymbol{Y}|_{\boldsymbol{\nu}} = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_i} = \prod_{i=1}^{k} \left(\frac{1}{\boldsymbol{L}_{i,i}}\right)^{-2\nu_i} = \prod_{i=1}^{k} \boldsymbol{U}_{i,i}^{-2\nu_i} = {}_U|\boldsymbol{U}\boldsymbol{U}^\top|_{-\boldsymbol{\nu}} = {}_U|\boldsymbol{Y}^{-1}|_{-\boldsymbol{\nu}}\,. \tag{A.5}$$

Part (v): As in the proof of part (iv), we first note that if $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are to lower triangular matrices, then the ith diagonal element of $\boldsymbol{L}_1\boldsymbol{L}_2$ equals $\boldsymbol{L}_{1,i,i}\boldsymbol{L}_{2,i,i}$, and the ith diagonal element of $\boldsymbol{L}^{-1}$ equals $1/\boldsymbol{L}_{i,i}$. Let $\boldsymbol{L}_{\boldsymbol{Y}}$ denote the unique lower triangular decomposition of $\boldsymbol{Y}$, and note that the inverse of a lower triangular

matrix is again lower triangular. We then obtain

$$
\begin{aligned}
|\boldsymbol{L}^{-1}\boldsymbol{Y}(\boldsymbol{L}^{-1})^\top|_{\boldsymbol{\nu}} &= |\boldsymbol{L}^{-1}\boldsymbol{L}_{\boldsymbol{Y}}\boldsymbol{L}_{\boldsymbol{Y}}^\top(\boldsymbol{L}^{-1})^\top|_{\boldsymbol{\nu}} = \prod_{i=1}^{k}\left((\boldsymbol{L}^{-1})_{i,i}\boldsymbol{L}_{\boldsymbol{Y},i,i}\right)^{2\nu_i} \\
&= \left(\prod_{i=1}^{k}\left((\boldsymbol{L}^{-1})_{i,i}\right)^{2\nu_i}\right)\left(\prod_{i=1}^{k}\boldsymbol{L}_{\boldsymbol{Y},i,i}^{2\nu_i}\right) = \left(\prod_{i=1}^{k}(\boldsymbol{L}_{i,i})^{-2\nu_i}\right)\left(\prod_{i=1}^{k}\boldsymbol{L}_{\boldsymbol{Y},i,i}^{2\nu_i}\right) \\
&= |\boldsymbol{\Sigma}|_{-\boldsymbol{\nu}} \cdot |\boldsymbol{Y}|_{\boldsymbol{\nu}}\,. \qquad \text{(A.6)}
\end{aligned}
$$

The proof for the UPWD is similar. ∎

Proof of Theorem 3:

Part (i): Let us define the pdf of $\boldsymbol{X}$ as

$$
p(\boldsymbol{X}) = \int p(\boldsymbol{X}|\boldsymbol{Y})p(\boldsymbol{Y})\,\mathrm{d}\boldsymbol{Y}, \qquad \text{(A.7)}
$$

where $p(\boldsymbol{X}|\boldsymbol{Y})$ follows a Riesz type I distribution with parameters $\boldsymbol{Y}$ and $\boldsymbol{\mu}$ as in (B.5), and $p(\boldsymbol{Y})$ follows an inverse Riesz type II distribution with pdf as in (B.9). Then (A.7) can be written as

$$
p(\boldsymbol{X}) = \frac{|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}}{{}_U|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}}\cdot\overline{\Gamma}\left(\boldsymbol{\mu}/2\right)\cdot\overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right)\cdot 2^{(\boldsymbol{\mu}+\boldsymbol{\nu})^\top\boldsymbol{\iota}_k/2}}\cdot\int\frac{{}_U|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}}{|\boldsymbol{Y}|_{0.5\boldsymbol{\mu}}}\mathrm{etr}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)\boldsymbol{Y}^{-1}\right)\mathrm{d}\boldsymbol{Y} \qquad \text{(A.8)}
$$

We focus on the integral in (A.8). By applying Lemma 20, we obtain

$$
\begin{aligned}
&\int\frac{{}_U|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}}{|\boldsymbol{Y}|_{0.5\boldsymbol{\mu}}}\mathrm{etr}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)\boldsymbol{Y}^{-1}\right)\mathrm{d}\boldsymbol{Y} \\
&\qquad = \int {}_U|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}\;|\boldsymbol{Y}|_{-0.5\boldsymbol{\mu}}\,\mathrm{etr}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)\boldsymbol{Y}^{-1}\right)\mathrm{d}\boldsymbol{Y} \\
&\qquad = \int {}_U|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}\;{}_U|\boldsymbol{Y}^{-1}|_{0.5\boldsymbol{\mu}}\,\mathrm{etr}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)\boldsymbol{Y}^{-1}\right)\mathrm{d}\boldsymbol{Y} \\
&\qquad = \int {}_U|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu}+k+1)}\,\mathrm{etr}\left(-\tfrac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)\boldsymbol{Y}^{-1}\right)\mathrm{d}\boldsymbol{Y}, \\
&\qquad = {}_U|(\boldsymbol{\Sigma}+\boldsymbol{X})^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}\,\overline{\Gamma}_U\left(\left(\boldsymbol{\mu}+\boldsymbol{\nu}\right)/2\right)\;2^{(\boldsymbol{\mu}+\boldsymbol{\nu})^\top\boldsymbol{\iota}_k/2}.
\end{aligned}
$$

Equation (A.8) thus becomes

$$
\begin{aligned}
p(\boldsymbol{X}) &= \frac{|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}}{{}_U|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}}\cdot\overline{\Gamma}\left(\boldsymbol{\mu}/2\right)\cdot\overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right)}\cdot{}_U|\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}\,\overline{\Gamma}_U\left(\left(\boldsymbol{\mu}+\boldsymbol{\nu}\right)/2\right) \\
&\overset{\text{Lemma 20(iii)}}{=} \frac{\overline{\Gamma}_U\left(\left(\boldsymbol{\mu}+\boldsymbol{\nu}\right)/2\right)}{\overline{\Gamma}\left(\boldsymbol{\mu}/2\right)\cdot\overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right)}\cdot|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}\cdot{}_U|\boldsymbol{\Sigma}^{-1}|_{-0.5\boldsymbol{\nu}}\cdot{}_U|\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \\
&\overset{\text{Lemma 20(iv)}}{=} \frac{\overline{\Gamma}_U\left(\left(\boldsymbol{\mu}+\boldsymbol{\nu}\right)/2\right)}{\overline{\Gamma}\left(\boldsymbol{\mu}/2\right)\cdot\overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right)}\cdot|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}\cdot|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}}\cdot|\left(\boldsymbol{\Sigma}+\boldsymbol{X}\right)|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}\,.
\end{aligned}
$$

Part (ii): Using (A.7), with $p(\boldsymbol{X}|\boldsymbol{Y})$ a Riesz type II distribution with parameters $\boldsymbol{Y}$ and $\boldsymbol{\mu}$ as in (B.7) and

$p(\boldsymbol{Y})$ an inverse Riesz type I distribution as in (B.8), we have the following pdf for $\boldsymbol{X}$:

$$p(\boldsymbol{X}) = \frac{{}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}(\boldsymbol{\nu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\mu}/2) \cdot 2^{(\boldsymbol{\nu}^\top+\boldsymbol{\mu}^\top)\iota_k/2}} \cdot \int \frac{|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}}{{}_U|\boldsymbol{Y}|_{0.5\boldsymbol{\mu}}} \operatorname{etr}\left(-\tfrac{1}{2}(\boldsymbol{\Sigma}+\boldsymbol{X})\boldsymbol{Y}^{-1}\right) \mathrm{d}\boldsymbol{Y}. \tag{A.9}$$

Focusing on the integral in (A.9) and by using Lemma 20(ii)-(iv), we obtain

$$\begin{aligned}
&\int \frac{|\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}}{{}_U|\boldsymbol{Y}|_{0.5\boldsymbol{\mu}}} \operatorname{etr}\left(-\tfrac{1}{2}(\boldsymbol{\Sigma}+\boldsymbol{X})\boldsymbol{Y}^{-1}\right) \mathrm{d}\boldsymbol{Y} \\
&\qquad = \int |\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \; {}_U|\boldsymbol{Y}|_{-0.5\boldsymbol{\mu}} \operatorname{etr}\left(-\tfrac{1}{2}(\boldsymbol{\Sigma}+\boldsymbol{X})\boldsymbol{Y}^{-1}\right) \mathrm{d}\boldsymbol{Y} \\
&\qquad = \int |\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \; |\boldsymbol{Y}^{-1}|_{0.5\boldsymbol{\mu}} \operatorname{etr}\left(-\tfrac{1}{2}(\boldsymbol{\Sigma}+\boldsymbol{X})\boldsymbol{Y}^{-1}\right) \mathrm{d}\boldsymbol{Y} \\
&\qquad = \int |\boldsymbol{Y}^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu}+k+1)} \operatorname{etr}\left(-\tfrac{1}{2}(\boldsymbol{\Sigma}+\boldsymbol{X})\boldsymbol{Y}^{-1}\right) \mathrm{d}\boldsymbol{Y} \\
&\qquad = |(\boldsymbol{\Sigma}+\boldsymbol{X})^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \, \overline{\Gamma}\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right) \; 2^{(\boldsymbol{\mu}+\boldsymbol{\nu})^\top \iota_k/2}.
\end{aligned}$$

Equation (A.9) can now be rewritten as

$$\begin{aligned}
p(\boldsymbol{X}) &= \frac{{}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}(\boldsymbol{\nu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\mu}/2)} \cdot |(\boldsymbol{\Sigma}+\boldsymbol{X})^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \, \overline{\Gamma}\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right) \\
&\overset{\text{Lemma 20(iii)}}{=} \frac{\overline{\Gamma}\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\nu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\mu}/2)} \cdot {}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}^{-1}|_{-0.5\boldsymbol{\nu}} \cdot |(\boldsymbol{\Sigma}+\boldsymbol{X})^{-1}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \\
&\overset{\text{Lemma 20(iv)}}{=} \frac{\overline{\Gamma}\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\nu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\mu}/2)} \cdot {}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot {}_U|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot {}_U|(\boldsymbol{\Sigma}+\boldsymbol{X})|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}.
\end{aligned}$$

Part (iii): Let $\boldsymbol{X} \sim \mathcal{FR}^I(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ with $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$ and $\boldsymbol{L}$ lower triangular, and consider the transformation $\boldsymbol{Y} = \boldsymbol{L}^{-1}\boldsymbol{X}(\boldsymbol{L}^\top)^{-1}$ with Jacobian $|\boldsymbol{L}|^{k+1} = |\boldsymbol{\Sigma}|^{0.5(k+1)} = |\boldsymbol{\Sigma}|_{0.5(k+1)}$, then the pdf of $\boldsymbol{Y}$ becomes

$$\begin{aligned}
p^{\boldsymbol{Y}}(\boldsymbol{Y}) &= p_{\mathcal{FR}^I}(\boldsymbol{L}\boldsymbol{Y}\boldsymbol{L}^\top;\ \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}) \\
&= \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\mu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\nu}/2)} \cdot |\boldsymbol{L}\boldsymbol{Y}\boldsymbol{L}^\top|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot |\boldsymbol{\Sigma}+\boldsymbol{L}\boldsymbol{Y}\boldsymbol{L}^\top|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \cdot |\boldsymbol{\Sigma}|_{0.5(k+1)} \\
&\overset{\text{Lemma 20(v)}}{=} \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\mu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\nu}/2)} \cdot |\boldsymbol{Y}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot |\boldsymbol{\Sigma}+\boldsymbol{L}\boldsymbol{Y}\boldsymbol{L}^\top|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \cdot |\boldsymbol{\Sigma}|_{0.5(k+1)} \\
&= \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\mu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\nu}/2)} \cdot |\boldsymbol{Y}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \cdot |\boldsymbol{L}(\mathbf{I}_k+\boldsymbol{Y})\boldsymbol{L}^\top|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \\
&\overset{\text{Lemma 20(v)}}{=} \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\mu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\nu}/2)} \cdot |\boldsymbol{Y}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \cdot |\mathbf{I}_k+\boldsymbol{Y}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \cdot |\boldsymbol{\Sigma}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \\
&= \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}(\boldsymbol{\mu}/2) \cdot \overline{\Gamma}_U(\boldsymbol{\nu}/2)} \cdot |\boldsymbol{Y}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\mathbf{I}_k+\boldsymbol{Y}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})}.
\end{aligned}$$

The proof for the F-Riesz type II is completely similar. ∎

Proof of Corollary 4:

For $\boldsymbol{\mu} = \mu \cdot \boldsymbol{\iota}_k$ and $\boldsymbol{\nu} = \nu \cdot \boldsymbol{\iota}_k$, we have the following sequence of equalities:

$$\begin{aligned}
p_{\mathcal{FR}^I}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}) &= \frac{\overline{\Gamma}_U\left((\boldsymbol{\mu}+\boldsymbol{\nu})/2\right)}{\overline{\Gamma}\left(\boldsymbol{\mu}/2\right) \cdot \overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right)} \cdot |\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)} \cdot |\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot |\boldsymbol{\Sigma}+\boldsymbol{X}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})} \\
&= \frac{\overline{\Gamma}_U\left((\mu+\nu)\cdot\boldsymbol{\iota}_k/2\right)}{\overline{\Gamma}\left(\mu\cdot\boldsymbol{\iota}_k/2\right) \cdot \overline{\Gamma}_U\left(\nu\cdot\boldsymbol{\iota}_k/2\right)} \cdot |\boldsymbol{X}|_{0.5(\mu-k-1)\boldsymbol{\iota}_k} \cdot |\boldsymbol{\Sigma}|_{0.5\nu\boldsymbol{\iota}_k} \cdot |\boldsymbol{\Sigma}+\boldsymbol{X}|_{-0.5(\mu+\nu)\boldsymbol{\iota}_k} \\
&= \frac{\Gamma_k\left((\mu+\nu)/2\right)}{\Gamma_k\left(\mu/2\right) \cdot \Gamma_k\left(\nu/2\right)} \cdot |\boldsymbol{X}|^{0.5(\mu-k-1)} \cdot |\boldsymbol{\Sigma}|^{0.5\nu} \cdot |\boldsymbol{\Sigma}+\boldsymbol{X}|^{-0.5(\mu+\nu)} \\
&= \frac{\Gamma_k\left((\mu+\nu)/2\right)}{\Gamma_k\left(\mu/2\right) \cdot \Gamma_k\left(\nu/2\right)} \cdot |\boldsymbol{X}|^{0.5(\mu-k-1)} \cdot |\boldsymbol{\Sigma}|^{0.5\nu} \cdot |\boldsymbol{\Sigma}+\boldsymbol{X}|^{-0.5(\mu+\nu)} \\
&= p_{\mathcal{F}}(\boldsymbol{X}; \boldsymbol{\Sigma}, \mu, \nu).
\end{aligned}$$

■

Proof of Theorem 5:

Part (i): As in the proof of Theorem 27, first note that we can write $\boldsymbol{X} \sim \mathcal{FR}^I(\mathbf{I}_k, \boldsymbol{\mu}, \boldsymbol{\nu})$ as $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{X}_1\boldsymbol{L}^\top$ for lower triangular $\boldsymbol{L}$, with $\boldsymbol{X}_1 \sim \mathcal{R}^I(\mathbf{I}, \boldsymbol{\mu})$ and $\boldsymbol{X}_2 = \boldsymbol{L}\boldsymbol{L}^\top \sim i\mathcal{R}^{II}(\mathbf{I}, \boldsymbol{\nu})$ and $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ independent. Moreover, as $\boldsymbol{X}_2^{-1} \sim \mathcal{R}^{II}(\mathbf{I}, \boldsymbol{\nu})$, we have $\boldsymbol{L} = (\boldsymbol{U}^\top)^{-1}$, with $\boldsymbol{U}\boldsymbol{U}^\top \sim \mathcal{R}^{II}(\mathbf{I}, \boldsymbol{\nu})$ due to the uniqueness of the choleski decompositions.

We define the $(i \times i)$ matrices $\boldsymbol{U}_i$ as the upper left $(i \times i)$ block of the matrix $\boldsymbol{U}$, where the elements and random structure of $\boldsymbol{U}$ were defined earlier as $\boldsymbol{H}$ in equation (B.6). We also define u_i as the ith diagonal element of $\boldsymbol{U}$. We have

$$\boldsymbol{U}_i = \begin{pmatrix} \boldsymbol{U}_{i-1} & \mathcal{N}^\top \\ 0 & u_i \end{pmatrix}, \qquad (\boldsymbol{U}_i^\top)^{-1} = \begin{pmatrix} (\boldsymbol{U}_{i-1}^\top)^{-1} & 0 \\ -u_i^{-1}\,\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1} & u_i^{-1} \end{pmatrix},$$

where $\mathcal{N} \in \mathbb{R}^{k\times 1}$ is a vector with independent (from $\boldsymbol{U}_{i-1}$ and u_i) standard normal random variables.

Define $\boldsymbol{A}_i$ as the upper $(i \times i)$ block of $\mathbb{E}[\boldsymbol{X}] = \mathbb{E}[(\boldsymbol{U}^\top)^{-1}\boldsymbol{X}_1\boldsymbol{U}^{-1}]$, and define $\boldsymbol{D}_i$ as a diagonal matrix with $(\mu_1, \ldots, \mu_i)$ on the diagonal. Due to the independence of $\boldsymbol{X}_1$ and $\boldsymbol{U}$, we have $\boldsymbol{A}_k = \mathbb{E}[(\boldsymbol{U}^\top)^{-1}\boldsymbol{D}_k\boldsymbol{U}^{-1}]$. Furthermore, due to the lower triangular structure of $(\boldsymbol{U}^\top)^{-1}$ and the independence of u_i, $\boldsymbol{U}_{i-1}$, and $\mathcal{N}$,

we have

$$
\begin{aligned}
\boldsymbol{A}_i &= \mathbb{E}\left[\begin{pmatrix}\mathbf{I}_i & \mathbf{0}_{i\times(k-i)}\end{pmatrix}(\boldsymbol{U}^\top)^{-1}\,\boldsymbol{D}\,\boldsymbol{U}^{-1}\begin{pmatrix}\mathbf{I}_i\\ \mathbf{0}_{(k-i)\times i}\end{pmatrix}\right]\\
&= \mathbb{E}\left[(\boldsymbol{U}_i^\top)^{-1}\,\boldsymbol{D}_i\,\boldsymbol{U}_i^{-1}\right]\\
&= \mathbb{E}\left[\begin{pmatrix}(\boldsymbol{U}_{i-1}^\top)^{-1} & 0\\ -u_i^{-1}\,\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1} & u_i^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{D}_{i-1} & 0\\ 0 & \mu_i\end{pmatrix}\begin{pmatrix}\boldsymbol{U}_{i-1}^{-1} & -\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}\,u_i^{-1}\\ 0 & u_i^{-1}\end{pmatrix}\right]\\
&= \mathbb{E}\left[\begin{pmatrix}(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1} & -(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}u_i^{-1}\\ -u_i^{-1}\,\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1} & \frac{\mu_i+\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}}{u_i^2}\end{pmatrix}\right]\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[\frac{\mu_i+\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}}{u_i^2}\right]\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[\frac{\mu_i+\mathrm{trace}\left(\mathcal{N}^\top\,(\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}\right)}{u_i^2}\right]\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[\frac{\mu_i+\mathrm{trace}\left((\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\,\mathcal{N}\mathcal{N}^\top\right)}{u_i^2}\right]\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[\frac{\mu_i+\mathrm{trace}\left((\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\right)}{u_i^2}\right]\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[u_i^{-2}\right]\mathbb{E}\left[\mu_i+\mathrm{trace}\left((\boldsymbol{U}_{i-1}^\top)^{-1}\boldsymbol{D}_{i-1}\boldsymbol{U}_{i-1}^{-1}\right)\right]\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \mathbb{E}\left[u_i^{-2}\right]\left(\mu_i+\mathrm{trace}\left(\boldsymbol{A}_{i-1}\right)\right)\end{pmatrix}\\
&= \begin{pmatrix}\boldsymbol{A}_{i-1} & 0\\ 0 & \frac{\mu_i+\mathrm{trace}(\boldsymbol{A}_{i-1})}{\nu_i-k+i-2}\end{pmatrix}.
\end{aligned}
$$

To start these recursions, it is easy to check that $\boldsymbol{A}_1 = \mu_1/(\nu_1 - k - 1)$. The proof for the type II F-Riesz is completely similar, with the only difference we partition all matrices from the bottom right rather than from the top left. ∎

Proof of Proposition 9: We note that under assumptions 7 and 8, we have $\mathbb{E}\|\boldsymbol{X}_t\| < \infty$ using Theorem 11, and thus $\|\boldsymbol{V}_0\| < \infty$. In particular, the bounded moment $\mathbb{E}\|\boldsymbol{X}_t\| < \infty$ follows from Theorem 5 and the finite $\|\boldsymbol{\Sigma}_0\| < \infty$ imposed in Assumption 8. The consistency $\hat{\boldsymbol{V}}_T = \frac{1}{T}\sum_{t=1}^T \boldsymbol{X}_t \overset{a.s.}{\to} \boldsymbol{V} = \mathbb{E}[\boldsymbol{X}_t]$ follows by application of Bernoulli's law of large numbers for i.i.d. data. ∎

Proof of Theorem 10: Following Theorem 5.14 in van der Vaart (2000) for i.i.d. data, we obtain the consistency of the MLE from

(i) the upper semi-continuity of the criterion function $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_T, \cdot, \cdot)$; and
(ii) the upper boundedness of the limit criterion in $(\boldsymbol{\mu}, \boldsymbol{\nu})$,

$$\mathbb{E} \sup_{(\boldsymbol{\mu},\boldsymbol{\nu})\in\tilde{\mathcal{U}}\times\tilde{\mathcal{V}})} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_0, \boldsymbol{\mu}, \boldsymbol{\nu}) < \infty$$

for every sufficiently small $\tilde{\mathcal{U}} \times \tilde{\mathcal{V}} \subseteq \mathcal{U} \times \mathcal{V}$. The upper semi-continuity of $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_T, \cdot, \cdot)$ follows easily as our pdf $\tilde{p}_{\mathcal{FR}^I}$ is continuous in all its parameters. The upper boundedness of $\mathbb{E}\sup_{(\boldsymbol{\mu},\boldsymbol{\nu})\in(\mathcal{U},\mathcal{V})} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_T, \boldsymbol{\mu}, \boldsymbol{\nu}) < \infty$ follows trivially from the fact that $\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}, \boldsymbol{\mu}, \boldsymbol{\nu})$ is uniformly bounded over the parameters $(\boldsymbol{V}, \boldsymbol{\mu}, \boldsymbol{\nu})$ in the admissible parameter space detailed in Assumption 8. ∎

Proof of Corollary 11: Consistency of $\hat{\boldsymbol{L}}_{\boldsymbol{V},T}$ follows from the consistency of $\hat{\boldsymbol{V}}_T \xrightarrow{p} \boldsymbol{V}_0 = \boldsymbol{L}_{\boldsymbol{V},0}\boldsymbol{L}_{\boldsymbol{V},0}^\top$ as $T \to \infty$, established in Proposition 9, since the Cholesky decomposition is unique and continuous for positive definite matrices. Consistency of $\boldsymbol{A}_k^{-1}(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ follows by the consistency of $(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) \xrightarrow{p} (\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ as $T \to \infty$ established in Theorem 10, and the continuity and non-singularity of $\boldsymbol{A}_k(\boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ under the maintained assumptions. Note that $\hat{\boldsymbol{\Sigma}}_T = \hat{\boldsymbol{\Sigma}}_T(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T) = \hat{\boldsymbol{L}}_T\hat{\boldsymbol{L}}_T^\top = \hat{\boldsymbol{L}}_{\boldsymbol{V},T}\boldsymbol{A}_k^{-1}(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)\hat{\boldsymbol{L}}_{\boldsymbol{V},T}^\top$, and thus $\hat{\boldsymbol{L}}_T = \hat{\boldsymbol{L}}_{\boldsymbol{V},T}\boldsymbol{A}_k^{-1/2}(\hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$ given the uniqueness of the Cholesky decomposition and the diagonality of $\boldsymbol{A}_k$. The consistency of $\hat{\boldsymbol{\Sigma}}_T$ to $\boldsymbol{\Sigma}_0$ now follows naturally by application of the continuous mapping theorem ∎

Proof of Proposition 15: Note that the set-up of the autoregressive specification for $\boldsymbol{V}_t$ in Boussama et al. (2011) is the same as equation (9), where Boussama et al. use a multivariate GARCH specification, thus replacing $\boldsymbol{X}_t$ by a rank one matrix $\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^\top$ for a vector valued random variable $\boldsymbol{\varepsilon}_t$. Under assumptions 12-13, we ensure that the regularity conditions required by Theorem 2.4 of Boussama et al. (2011) hold, which is most easily seen by subtracting the conditional mean of $\boldsymbol{X}_t$; compare the line of proof in Asai and So (2021). As a result, the random sequences $\{\boldsymbol{V}_t\}_{t\in\mathbb{Z}}$ and $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ are strictly stationary and ergodic and satisfy $\mathbb{E}|\boldsymbol{V}_t| < \infty$ and $\mathbb{E}|\boldsymbol{X}_t|^2 < \infty$. Finally, the almost sure convergence of $\hat{\boldsymbol{\Omega}}$ follows immediately by application of the ergodic theorem. ∎

Proof of Proposition 17: Define the unfolded limit process as

$$\boldsymbol{V}_t = \frac{(1-A-B)}{1-B}\boldsymbol{\Omega} + A\sum_{j=0}^{\infty} B^j\, \boldsymbol{X}_{t-j}.$$

Unfold the filter $\hat{\boldsymbol{V}}_t$ back to its initial condition $\hat{\boldsymbol{V}}_1$ at time $t = 1$,

$$\hat{\boldsymbol{V}}_t = \sum_{j=0}^{t-1} B^j(1-A-B)\boldsymbol{\Omega} + A\sum_{j=0}^{t-2} B^j\, \boldsymbol{X}_{t-1-j} + B^{t-1}\hat{\boldsymbol{V}}_1$$

Next, we note that

$$\lim_{t\to\infty}\sup_{B\in\mathcal{B}}\|\hat{\boldsymbol{V}}_t-\boldsymbol{V}_t\|\leq\lim_{t\to\infty}\sup_{B\in\mathcal{B}}\Big|\sum_{j=0}^{t-1}B^j(1-A-B)-\frac{(1-A-B)}{1-B}\Big|\|\boldsymbol{\Omega}\|$$
$$+\lim_{t\to\infty}|A|\sup_{B\in\mathcal{B}}\Big\|\sum_{j=0}^{t-2}B^j\,\boldsymbol{X}_{t-1-j}-\sum_{j=0}^{\infty}B^j\,\boldsymbol{X}_{t-j}\Big\|+\lim_{t\to\infty}\|\boldsymbol{X}_1\|\sup_{B\in\mathcal{B}}|B^{t-1}|\;=\;0.$$

It is clear that

$$\lim_{t\to\infty}\sup_{B\in\mathcal{B}}\Big|\sum_{j=0}^{t-1}B^j(1-A-B)-\frac{(1-A-B)}{1-B}\Big|=0$$

as the series is geometrically declining to zero for $|B|<1$. Furthermore,

$$\lim_{t\to\infty}|A|\sup_{B\in\mathcal{B}}\Big\|\sum_{j=0}^{t-2}B^j\,\boldsymbol{X}_{t-1-j}-\sum_{j=0}^{\infty}B^j\,\boldsymbol{X}_{t-j}\Big\|=0$$

holds as $|B|<1$ and $\{\boldsymbol{X}_t\}_{t\in\mathbb{Z}}$ is strictly stationary with a logarithmic moment (see e.g. Lemma 2.1 in Straumann and Mikosch, 2006). Finally, $\lim_{t\to\infty}\|\boldsymbol{X}_1\|\sup_{B\in\mathcal{B}}|B^{t-1}|=0$ as $\boldsymbol{X}_1$ has a logarithmic moment and $|B|<1$ (see also Lemma 2.1 in Straumann and Mikosch, 2006). ∎

Proof of Theorem 18: Application of Theorem 5.14 in van der Vaart (2000) to strictly stationary data follows as long as a pointwise law of large numbers still holds for the log likelihood. We thus proceed to obtain the consistency of our estimator by verifying:
(i) the upper semi-continuity of the criterion function $\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T,\cdot,\cdot),\cdot,\cdot)$;
(ii) the upper boundedness of the limit criterion

$$\mathbb{E}\sup_{(A,B,\boldsymbol{\mu},\boldsymbol{\nu})\in\mathcal{A}\times\mathcal{B}\times\mathcal{U}\times\mathcal{V}}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A,B),\boldsymbol{\mu},\boldsymbol{\nu})<\infty$$

for every sufficiently small $\mathcal{A}\times\mathcal{B}\times\mathcal{U}\times\mathcal{V}$, and
(iii) the pointwise convergence in probability of the sample criterion

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T,A,B),\boldsymbol{\mu},\boldsymbol{\nu})\;=\;\mathbb{E}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A,B),\boldsymbol{\mu},\boldsymbol{\nu}),$$

for every $(A,B,\boldsymbol{\mu},\boldsymbol{\nu})\in\mathcal{A}\times\mathcal{B}\times\mathcal{U}\times\mathcal{V}$ such that $\mathbb{E}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A,B),\boldsymbol{\mu},\boldsymbol{\nu})>-\infty$.[2]

The upper semi-continuity of $\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T,\cdot,\cdot),\cdot,\cdot)$ in (i) follows easily from the continuity of the filter $\hat{\boldsymbol{V}}_t$ on all parameters.

[2]We highlight that the existence of parameter values $(A,B,\boldsymbol{\mu},\boldsymbol{\nu})$ for which the loglikelihood is not $-\infty$, can be taken as given. As noted in van der Vaart (2000), if $\mathbb{E}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A,B),\boldsymbol{\mu},\boldsymbol{\nu})=-\infty$ uniformly in $(A,B,\boldsymbol{\mu},\boldsymbol{\nu})$, then $\mathcal{A}_0\times\mathcal{B}_0\times\mathcal{U}_0\times\mathcal{V}_0=\mathcal{A}\times\mathcal{B}\times\mathcal{U}\times\mathcal{V}$ and there is nothing to prove. So we can proceed under the assumption that there exists a $(A_0,B_0),\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)$ such that $\mathbb{E}\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A_0,B_0),\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)>-\infty$, and hence that $\mathbb{E}|\log\tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t;\boldsymbol{V}_t(\boldsymbol{\Omega}_0,A_0,B_0),\boldsymbol{\mu}_0,\boldsymbol{\nu}_0)|<\infty$ by the one-sided bound.

The upper boundedness of the limit criterion in (ii) follows from the uniform boundedness of $\tilde{p}_{\mathcal{FR}^I}$ over the parameters in the admissible parameter space as detailed in Assumption 13.

Note that the convergence in (iii) holds for parameter values $(A, B, \boldsymbol{\mu}, \boldsymbol{\nu})$ for which

$$\mathbb{E}|\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})| < \infty.$$

Hence, the desired convergence is obtained by noting that,

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) \;= \\
&\quad \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Big[\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) - \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})\Big] \\
&\quad + \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Big[\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) - \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})\Big] \\
&\quad + \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) \\
&\quad = 0 + 0 + \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) \\
&\quad = \mathbb{E} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}),
\end{aligned}$$

since

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Big[\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) - \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})\Big] = 0$$

by the continuity of the filter $\hat{\boldsymbol{V}}_t$ in the parameters and the consistency of $\hat{\boldsymbol{\Omega}}_T$ to $\boldsymbol{\Omega}_0$, and

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Big[\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) - \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})\Big] = 0,$$

holds by the invertibility of the filter as established in Proposition 17, and

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \boldsymbol{V}_t(\boldsymbol{\Omega}_0, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}) = 0,$$

holds by the ergodic theorem. Identification of the parameter of interest holds trivially in this setting given the linearity of the updating equation. ∎

B More detailed introduction to the Riesz distribution

This section starts with a brief introduction of the Riesz distribution as an extension of the Wishart distribution that allows for more tail heterogeneity. We discuss the Riesz distribution's most important

properties and review some key notation that is also required for the definition of the F-Riesz distribution in the next section.

The Riesz distribution (Hassairi and Lajmi, 2001) is defined over the space of positive definite matrices. It generalizes the well-known Wishart distribution, which has probability density function (pdf)

$$p_{\mathcal{W}}(\boldsymbol{Y}; \boldsymbol{\Sigma}, \nu) = \frac{|\boldsymbol{Y}|^{0.5(\nu-k-1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right)}{|\boldsymbol{\Sigma}|^{0.5\nu} \cdot \Gamma_k\left(\nu/2\right) \cdot 2^{k\cdot\nu/2}}, \tag{B.1}$$

for a positive definite matrix random variable $\boldsymbol{Y} \in \mathbb{R}^{k\times k}$, a positive definite scaling matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{k\times k}$, and a positive scalar degrees of freedom parameter ν, where $\operatorname{etr}(\,\cdot\,) = \exp(\operatorname{trace}(\,\cdot\,))$ denotes the exponential trace operator, and $\Gamma_k(\,\cdot\,)$ is the multivariate gamma function,

$$\Gamma_k\left(\nu\right) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\nu + \tfrac{1-i}{2}\right). \tag{B.2}$$

A Wishart distributed random variable, denoted as $\mathcal{W}(\boldsymbol{\Sigma}, \nu)$, thus has two key parameters: one matrix-valued, and one scalar. Interestingly, the Wishart distribution can be constructed using the so-called Bartlett decomposition; see Anderson (1962). Define the lower triangular matrix $\boldsymbol{G} \in \mathbb{R}^{k\times k}$ with all its elements independent random variables with $\boldsymbol{G}_{ii}^2 \sim \chi^2_{\nu-i+1}$ and $\boldsymbol{G}_{ij} \sim \mathcal{N}(0,1)$ for $i > j$, i.e.,

$$\boldsymbol{G} = \begin{pmatrix} \sqrt{\chi^2_\nu} & 0 & \cdots & 0 \\ \mathcal{N}(0,1) & \ddots & 0 & \vdots \\ \vdots & \mathcal{N}(0,1) & \ddots & 0 \\ \mathcal{N}(0,1) & \cdots & \mathcal{N}(0,1) & \sqrt{\chi^2_{\nu-k+1}} \end{pmatrix}. \tag{B.3}$$

Then $\boldsymbol{Y} = \boldsymbol{G}\boldsymbol{G}^\top \sim \mathcal{W}(\mathbf{I}_k, \nu)$, and $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{G}^\top\boldsymbol{L}^\top \sim \mathcal{W}(\boldsymbol{\Sigma}, \nu)$ for a matrix $\boldsymbol{L}$ such that $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$. A key property of the Bartlett decomposition is that the same degrees of freedom parameter ν plays a role in all the diagonal elements of $\boldsymbol{G}$ in (B.3). The Riesz distribution generalizes the Wishart by instead introducing a vector $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top$ of degrees of freedom parameters, and inserting it into (B.3). This is done in Definition 21 below. A Riesz distribution $\mathcal{R}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ is thus characterized by a scaling matrix $\boldsymbol{\Sigma}$ and a vector $\boldsymbol{\nu}$.

Using the definitions of the generalized gamma function and the power weighted determinant from the main text, we can now introduce the lower triangular (type-I) and upper triangular (type-II) version of the Riesz distribution; see also for instance Díaz-García (2013) and Louati and Masmoudi (2015).

Theorem 21 (**Riesz distribution type I and II**).

(i) Consider the Bartlett decomposition $\boldsymbol{G} \in \mathbb{R}^{k\times k}$*, defined as*

$$\boldsymbol{G} = \begin{pmatrix} \sqrt{\chi^2_{\nu_1}} & 0 & \cdots & 0 \\ \mathcal{N}(0,1) & \ddots & 0 & \vdots \\ \vdots & \mathcal{N}(0,1) & \ddots & 0 \\ \mathcal{N}(0,1) & \cdots & \mathcal{N}(0,1) & \sqrt{\chi^2_{\nu_k-k+1}} \end{pmatrix}, \tag{B.4}$$

for $\nu_i > i-1$ *for* $i = 1,\ldots,k$*, and let* $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{G}^\top\boldsymbol{L}^\top$*, where* $\boldsymbol{L}$ *is the lower triangular Cholesky decomposition of* $\boldsymbol{\Sigma}$*, such that* $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$*. Then* $\boldsymbol{Y}$ *has density function*

$$p_{\mathcal{R}^I}(\boldsymbol{Y};\boldsymbol{\Sigma},\boldsymbol{\nu}) = \frac{|\boldsymbol{Y}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right)}{|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top\boldsymbol{\iota}_k/2}}, \tag{B.5}$$

also known as a Riesz type-I density, $\mathcal{R}^I(\boldsymbol{\Sigma},\boldsymbol{\nu})$*, where the generalized multivariate Gamma function* $\overline{\Gamma}(\cdot)$ *and the Lower Power Weighted Determinant* $|\cdot|_{\boldsymbol{\nu}}$ *were defined in Definitions 1 and 2, respectively.*

(ii) Let the Bartlett decomposition $\boldsymbol{H} \in \mathbb{R}^{k\times k}$ *be defined as*

$$\boldsymbol{H} = \begin{pmatrix} \sqrt{\chi^2_{\nu_1-k+1}} & \mathcal{N}(0,1) & \cdots & \mathcal{N}(0,1) \\ 0 & \ddots & \mathcal{N}(0,1) & \vdots \\ \vdots & 0 & \ddots & \mathcal{N}(0,1) \\ 0 & \cdots & 0 & \sqrt{\chi^2_{\nu_k}} \end{pmatrix}, \tag{B.6}$$

for $\nu_i > k-i$ *for* $i = 1,\ldots,k$*, and let* $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{H}\boldsymbol{H}^\top\boldsymbol{U}^\top$*, where* $\boldsymbol{U}$ *is the upper triangular Cholesky decomposition of* $\boldsymbol{\Sigma}$*, such that* $\boldsymbol{\Sigma} = \boldsymbol{U}\boldsymbol{U}^\top$*. Then* $\boldsymbol{X}$ *has density function*

$$p_{\mathcal{R}^{II}}(\boldsymbol{X};\boldsymbol{\Sigma},\boldsymbol{\nu}) = \frac{{}_U|\boldsymbol{X}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)}{{}_U|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top\boldsymbol{\iota}_k/2}}, \tag{B.7}$$

also known as a Riesz type-II density, $\mathcal{R}^{II}(\boldsymbol{\Sigma},\boldsymbol{\nu})$*, with* $\overline{\Gamma}_U(\cdot)$ *and* ${}_U|\cdot|_{\boldsymbol{\nu}}$ *as defined in Definitions 1 and 2.*

The Riesz distributions of type I and II bear a close resemblance to the Wishart distribution. Using manipulation rule (i) from Lemma 20, we directly establish the following corollary.

Corollary 22. *If* $\boldsymbol{\nu} = (\nu,\ldots,\nu)^\top$ *for some positive scalar* $\nu > k-1$*, then the Wishart, Riesz-I, and Riesz-II densities from (B.1), (B.5), and (B.7), respectively, all coincide.*

The Wishart distribution is thus a special case of the Riesz. The Bartlett decompositions in (B.4) and (B.6), moreover, provide a direct way to simulate from the Riesz-I and Riesz-II distribution. Also note that the density expressions in Theorem 21 are easy to implement for numerical maximization of a likelihood function to estimate $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}$. They only require determinants and Cholesky decompositions.

Finally, like the Wishart, the Riesz distribution also allows for an inverse version, called the inverse Riesz type-I $i\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ and type II $i\mathcal{R}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$. This will be important for constructing F-Riesz distributions in the next section. The definition of the type I and II inverse Riesz distributions is given in the following definition and theorem.

Definition 23.

(i) Let $\boldsymbol{Y}$ be a Riesz distribution of type I, $\mathcal{R}^I(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\nu})$, and let $\boldsymbol{X} = \boldsymbol{Y}^{-1}$, then $\boldsymbol{X}$ is inverse Riesz distributed of type I, $i\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$.

(ii) Similarly, if $\boldsymbol{Y} \sim \mathcal{R}^{II}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\nu})$, then $\boldsymbol{X} = \boldsymbol{Y}^{-1}$ is inverse Riesz type II, $i\mathcal{R}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$.

Theorem 24. *The pdf of an $i\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ distributed random variable $\boldsymbol{X}$ is given by*

$$p_{i\mathcal{R}^I}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \frac{|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{X}^{-1}\right)}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}. \tag{B.8}$$

The pdf of an $i\mathcal{R}^{II}(\boldsymbol{\Sigma}, \boldsymbol{\nu})$ distributed random variable $\boldsymbol{X}$ is given by

$$p_{i\mathcal{R}^{II}}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \frac{{}_U|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{X}^{-1}\right)}{{}_U|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}_U\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}. \tag{B.9}$$

The first moments of the Riesz and inverse Riesz distributions have been derived by Díaz-García (2013) and Louati and Masmoudi (2015). For completeness, they are included in Appendix C.

C Additional proofs

Proof of Theorem 21:

Part (i): We first derive the expression for the Riesz distribution of type I with $\boldsymbol{\Sigma} = \mathbf{I}_k$. Let $\boldsymbol{Y}^\star = \boldsymbol{G}\boldsymbol{G}^\top$. Given the independence of the elements of $\boldsymbol{G}$, we can write its density as the product of the densities of the individual elements. We have

$$\begin{aligned} p\left(\operatorname{vech}(\boldsymbol{G})\right) &= \frac{2^k \cdot \left(\prod_{i=1}^k \boldsymbol{G}_{i,i}^{\nu_i - i}\right) \cdot \exp\left(-\frac{1}{2}\operatorname{vech}(\boldsymbol{G})^\top \operatorname{vech}(\boldsymbol{G})\right)}{(2\pi)^{k(k-1)/4}\,\Gamma\left(\frac{\nu_1}{2}\right)\cdots\Gamma\left(\frac{\nu_k-k+1}{2}\right)\,2^{(\nu_1+(\nu_2-1)+\cdots+(\nu_k-k+1))/2}} \\ &= \frac{2^k \cdot |\boldsymbol{G}\boldsymbol{G}^\top|_{0.5(\boldsymbol{\nu}-\boldsymbol{m})} \cdot \exp\left(-\frac{1}{2}\operatorname{trace}(\boldsymbol{G}^\top\boldsymbol{G})\right)}{\overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}, \end{aligned} \tag{C.1}$$

where $\boldsymbol{m} = (1, 2, \ldots, k)^\top$. We note that the Jacobian of the transformation from $\operatorname{vech}(\boldsymbol{G})$ to $\operatorname{vech}(\boldsymbol{Y}^\star)$ is

$$\left|\frac{\partial \operatorname{vech}(\boldsymbol{Y}^\star)}{\partial \operatorname{vech}(\boldsymbol{G})^\top}\right|^{-1} = 2^{-k} \cdot \boldsymbol{G}_{1,1}^{-k} \cdot \boldsymbol{G}_{2,2}^{-(k-1)} \cdots \boldsymbol{G}_{k,k}^{-1} = 2^{-k} \cdot |\boldsymbol{Y}^\star|_{0.5(\boldsymbol{m}-1-k)} \cdot$$

Multiplying this Jacobian with the density of vech($\boldsymbol{G}$) in (C.1), and substituting $\boldsymbol{G}\boldsymbol{G}^\top = \boldsymbol{Y}^\star$, we obtain

$$\begin{aligned} p\left(\text{vech}(\boldsymbol{Y}^\star)\right) &= \frac{2^k \cdot |\boldsymbol{Y}^\star|_{0.5(\boldsymbol{\nu}-\boldsymbol{m})} \cdot \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{Y}^\star)\right)}{\overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \times 2^{-k} \cdot |\boldsymbol{Y}^\star|_{0.5(\boldsymbol{m}-1-k)} \\ &= \frac{|\boldsymbol{Y}^\star|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{Y}^\star)\right)}{\overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}. \end{aligned}$$

Now consider $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{Y}^\star\boldsymbol{L}^\top$ for $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{\Sigma}$. By applying (11.33.c) from Abadir and Magnus (2005), we note that the Jacobian of the transformation from vech($\boldsymbol{Y}^\star$) to vech($\boldsymbol{Y}$) is given by

$$\left|\frac{\partial \text{vech}(\boldsymbol{Y}^\star)}{\partial \text{vech}(\boldsymbol{Y})^\top}\right| = |\boldsymbol{L}|^{-k-1} = |\boldsymbol{\Sigma}|_{-0.5(k+1)}.$$

Multiplying this Jacobian with the density of vech($\boldsymbol{Y}^\star$) and substituting $\boldsymbol{Y}^\star = \boldsymbol{L}^{-1}\boldsymbol{Y}(\boldsymbol{L}^{-1})^\top$, and using Lemma 20(v), we obtain

$$\begin{aligned} p\left(\text{vech}(\boldsymbol{Y})\right) &= \frac{|\boldsymbol{L}^{-1}\boldsymbol{Y}\boldsymbol{L}^{-1\top}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{L}^{-1}\boldsymbol{Y}\boldsymbol{L}^{-1\top})\right)}{\overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \times |\boldsymbol{\Sigma}|_{-0.5(k+1)} \\ &= \frac{|\boldsymbol{\Sigma}|_{-0.5(\boldsymbol{\nu}-k-1)} \cdot |\boldsymbol{Y}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Y})\right)}{\overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \times |\boldsymbol{\Sigma}|_{-0.5(k+1)} \\ &= \frac{|\boldsymbol{Y}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \exp\left(-\frac{1}{2}\text{trace}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Y})\right)}{|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}_k\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}. \end{aligned}$$

The proof for part (ii) of the Theorem is similar. ∎

Proof of Corollary 22:

If we assume $\boldsymbol{\nu} = (\nu, \ldots, \nu)^\top$ for some positive scalar $\nu > k-1$, then using Lemma 20(i) the Riesz-I distribution becomes

$$p_{\mathcal{R}^I}(\boldsymbol{Y}; \boldsymbol{\Sigma}, \nu) = \frac{|\boldsymbol{Y}|^{\,0.5(\nu-k-1)} \cdot \text{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right)}{|\boldsymbol{\Sigma}|^{\,0.5\nu} \cdot \overline{\Gamma}\left(\nu \cdot \boldsymbol{\iota}_k/2\right) \cdot 2^{k\nu/2}}.$$

Similarly the Riesz-II distribution becomes

$$p_{\mathcal{R}^{II}}(\boldsymbol{Y}; \boldsymbol{\Sigma}, \nu) = \frac{|\boldsymbol{Y}|^{\,0.5(\nu-k-1)} \cdot \text{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}\right)}{|\boldsymbol{\Sigma}|^{\,0.5\nu} \cdot \overline{\Gamma}_U\left(\nu \cdot \boldsymbol{\iota}_k/2\right) \cdot 2^{k\nu/2}}.$$

For $\boldsymbol{\nu} = \nu \cdot \boldsymbol{\iota}_k$ with scalar ν, it is easy to check that $\overline{\Gamma}_U(\nu \cdot \boldsymbol{\iota}_k) = \overline{\Gamma}(\nu \cdot \boldsymbol{\iota}_k) = \Gamma_k(\nu)$. It follows that in this case the Riesz-I, Riesz-II, and Wishart distribution all coincide. ∎

Proof of Theorem 24:

Let $\boldsymbol{Y}$ have a Riesz distribution of Type I with hyper-parameters $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\nu}$. Then following (B.5), we have that $\boldsymbol{Y}$ has density

$$p_{\mathcal{R}^I}(\boldsymbol{Y}; \boldsymbol{\Sigma}^{-1}, \boldsymbol{\nu}) = \frac{|\boldsymbol{Y}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \text{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{Y}\right)}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}.$$

We apply the transformation, $\boldsymbol{X} = \boldsymbol{Y}^{-1}$ with Jacobian $abs(|X|)^{-k-1}$; see (13.38b) of Abadir and Magnus (2005). Combining this Jacobian with the expression for $p_{\mathcal{R}^I}(\cdot)$ above, we obtain

$$\begin{aligned} p_{i\mathcal{R}^I}(\boldsymbol{X};\boldsymbol{\Sigma},\boldsymbol{\nu}) &= p_{\mathcal{R}^I}(\boldsymbol{X}^{-1};\boldsymbol{\Sigma}^{-1},\boldsymbol{\nu}) \cdot |X|^{-(k+1)} \\ &= \frac{|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}-k-1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{X}^{-1}\right)}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}(\boldsymbol{\nu}/2) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \cdot |X|^{-(k+1)} \\ &\overset{\text{Lemma 20(i)–(ii)}}{=} \frac{|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{X}^{-1}\right)}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}(\boldsymbol{\nu}/2) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}. \end{aligned}$$

Thus by applying the transformation $\boldsymbol{X} = \boldsymbol{Y}^{-1}$, we obtain the pdf of the inverse Riesz type I as in B.8. The proof for the type II inverse Riesz is similar. ∎

Theorem 25 (Expectation of (inverse) Riesz distributions; Díaz-García (2013) and Louati and Masmoudi (2015)).

(i) Let $\boldsymbol{Y} \sim \mathcal{R}^I(\mathbf{I},\boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \operatorname{diag}(\boldsymbol{\nu})$ *for* $\nu_i > i-1$ *for* $i = 1,\ldots,k$*.*

(ii) Let $\boldsymbol{Y} \sim \mathcal{R}^{II}(\mathbf{I},\boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \operatorname{diag}(\boldsymbol{\nu})$ *for* $\nu_i > k-i$ *for* $i = 1,\ldots,k$*.*

(iii) Let $\boldsymbol{Y} \sim i\mathcal{R}^I(\mathbf{I},\boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \sum_{i=1}^k a_i \boldsymbol{e}_i \boldsymbol{e}_i^\top$*, with* $\boldsymbol{e}_i$ *the* i*th column from* $\mathbf{I}_k$*, and*

$$\begin{aligned} a_i &= \frac{1}{\nu_i - (i+1)} \prod_{j=i+1}^{k} \frac{\nu_j - j}{\nu_j - (j+1)} \quad \text{if} \quad i = 1,\ldots,k-1, \\ a_i &= \frac{1}{\nu_i - (i+1)} \quad \text{if} \quad i = k. \end{aligned}$$

(iv) Let $\boldsymbol{Y} \sim i\mathcal{R}^{II}(\mathbf{I},\boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \sum_{i=1}^k a_i \boldsymbol{e}_i \boldsymbol{e}_i^\top$*, with* $\boldsymbol{e}_i$ *the* i*th column from* $\mathbf{I}_k$*, and*

$$\begin{aligned} a_i &= \frac{1}{\nu_i - (k+1)} \quad \text{if} \quad i = 1, \\ a_i &= \frac{1}{\nu_i - (k-i+2)} \prod_{j=1}^{i-1} \frac{\nu_j - (k-j+1)}{\nu_j - (k-j+2)} \quad \text{if} \quad i = 2,\ldots,k. \end{aligned}$$

Lemma 26. *If* $\boldsymbol{X} \sim \mathcal{R}^I(\mathbf{I}_k,\boldsymbol{\nu})$ *with lower triangular Bartlett generator* $\boldsymbol{L}$ *such that* $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{L}^\top$*, then* $\boldsymbol{L}^\top$ *is the upper triangular Bartlett generator of a* $\mathcal{R}^{II}(\mathbf{I},\boldsymbol{\nu}-2\tilde{\boldsymbol{\gamma}})$*.*
If $\boldsymbol{X} \sim \mathcal{R}^{II}(\mathbf{I}_k,\boldsymbol{\nu})$ *with upper triangular Bartlett generator* $\boldsymbol{U}$ *such that* $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{U}^\top$*, then* $\boldsymbol{U}^\top$ *is the lower triangular Bartlett generator of a* $\mathcal{R}^I(\mathbf{I},\boldsymbol{\nu}+2\tilde{\boldsymbol{\gamma}})$*.*

Proof: The second part follows directly by observing

$$\begin{aligned}\operatorname{diag}(\boldsymbol{U}^\top) &= \left(\sqrt{\chi^2_{\nu_1-k+1}}\,,\quad \cdots\quad ,\ \sqrt{\chi^2_{\nu_k}}\right)\\ &= \left(\sqrt{\chi^2_{\nu_1+2\cdot(-0.5(k-1))}}\,,\quad \cdots\quad ,\ \sqrt{\chi^2_{(\nu_k-k+1)+2\cdot(0.5(k-1))}}\right)\\ &= \left(\sqrt{\chi^2_{(\nu_1+2\tilde{\boldsymbol{\gamma}}_1)}}\,,\quad \cdots\quad ,\ \sqrt{\chi^2_{(\nu_k+2\tilde{\boldsymbol{\gamma}}_k)-k+1}}\right),\end{aligned}$$

whereas the strict lower diagonal part of $\boldsymbol{U}^\top$ is filled with independent standard normal random variables. The first part follows similarly from

$$\begin{aligned}\operatorname{diag}(\boldsymbol{L}^\top) &= \left(\sqrt{\chi^2_{\nu_1}}\,,\quad \cdots\quad ,\ \sqrt{\chi^2_{\nu_k-k+1}}\right)\\ &= \left(\sqrt{\chi^2_{(\nu_1-2\tilde{\boldsymbol{\gamma}}_1)-k+1}}\,,\quad \cdots\quad ,\ \sqrt{\chi^2_{(\nu_k-2\tilde{\boldsymbol{\gamma}}_k)}}\right).\end{aligned}$$

■

Theorem 27 below provides the correction for a result derived by Díaz-García (2016). The original result, sadly enough, is incorrect and contains a mistake in the density specification which results in severe biases in $\boldsymbol{\Sigma}$, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$ if $\boldsymbol{\Sigma}_t \neq \mathbf{I}_k$. This is particularly important in the setting of our paper, where $\boldsymbol{\Sigma}_t$ and $\boldsymbol{V}_t$ are the focal point. The next theorem provides the correct result, using Lemma 26 above.

Theorem 27. *Assume $\boldsymbol{X}_1 \sim \mathcal{R}^I(\mathbf{I}, \boldsymbol{\mu})$ and $\boldsymbol{X}_2 \sim \mathcal{R}^I(\mathbf{I}, \boldsymbol{\nu})$. Define the lower triangular matrices $\boldsymbol{G}$ and $\boldsymbol{L}$ such that $\boldsymbol{G}\boldsymbol{G}^\top = \boldsymbol{X}_2$ and $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$. Then $\boldsymbol{X} = \boldsymbol{L}\boldsymbol{G}^{-1}\boldsymbol{X}_1\boldsymbol{G}^{-\top}\boldsymbol{L}^\top$ follows a type I F-Riesz distribution with density function*

$$p(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \frac{\overline{\Gamma}\left(\frac{1}{2}(\boldsymbol{\mu}+\boldsymbol{\nu})\right)\cdot|\boldsymbol{\Sigma}|_{0.5\boldsymbol{\nu}-\tilde{\boldsymbol{\gamma}}}}{\overline{\Gamma}\left(\frac{1}{2}\boldsymbol{\mu}\right)\overline{\Gamma}\left(\frac{1}{2}\boldsymbol{\nu}\right)}\ |\boldsymbol{X}|_{0.5(\boldsymbol{\mu}-k-1)}\ |\boldsymbol{\Sigma}+\boldsymbol{X}|_{-0.5(\boldsymbol{\mu}+\boldsymbol{\nu})+\tilde{\boldsymbol{\gamma}}}\,, \tag{C.2}$$

where $\tilde{\boldsymbol{\gamma}}$ is defined in equation (4). Moreover,

$$p(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}+2\tilde{\boldsymbol{\gamma}}) = p_{\mathcal{FR}^I}(\boldsymbol{X}; \boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu}). \tag{C.3}$$

Proof of Theorem 27: We prove the result for $\boldsymbol{\Sigma} = \mathbf{I}_k$. For general $\boldsymbol{\Sigma}$, one can use the result of Theorem 3 part (iii).

To construct the F-Riesz type I, we generate a $\boldsymbol{X}_1 \sim \mathcal{R}(\mathbf{I}, \boldsymbol{\mu})$ and an independent $\boldsymbol{X}_2 \sim i\mathcal{R}(\mathbf{I}, \boldsymbol{\nu})$, and then use the lower triangular Choleski decomposition $\boldsymbol{L}_2$ of $\boldsymbol{X}_2$ to construct $\boldsymbol{X} \sim \mathcal{FR}^I(\mathbf{I}, \boldsymbol{\mu}, \boldsymbol{\nu})$ as $\boldsymbol{X} = \boldsymbol{L}_2\,\boldsymbol{X}_1\,\boldsymbol{L}_2^\top$ based on Theorem 3 part (iii).

As $\boldsymbol{X}_2$ is inverse Riesz type II, $\boldsymbol{X}_2^{-1}$ has a Riesz type II distribution, with upper triangular Bartlett generator $\boldsymbol{U}_2$ such that $\boldsymbol{U}_2\boldsymbol{U}_2^\top = \boldsymbol{X}_2^{-1} \sim \mathcal{R}^{II}(\mathbf{I}_k, \boldsymbol{\nu})$. As it also holds that $\boldsymbol{X}_2 = \boldsymbol{L}_2\boldsymbol{L}_2'$, it follows that $\boldsymbol{L}_2 = (\boldsymbol{U}_2^\top)^{-1}$ due to the uniqueness of the Choleski decompositions. Based on Lemma 26 below, $\boldsymbol{U}_2^\top$ is the Bartlet generator is of a $\mathcal{R}^I(\mathbf{I}, \boldsymbol{\nu}+2\tilde{\boldsymbol{\gamma}})$ distribution.

Díaz-García generates the F-Riesz type distribution as $\boldsymbol{L}_{DG}^{-1}\boldsymbol{X}_1(\boldsymbol{L}_{DG}^{\top})^{-1}$ with $\boldsymbol{X}_{DG} = \boldsymbol{L}_{DG}\boldsymbol{L}_{DG}^{\top}$ a draw from a $\mathcal{R}^I(\mathbf{I}_k, \boldsymbol{\nu}_{DG})$ distribution. It follows that $\boldsymbol{\nu}_{DG} = \boldsymbol{\nu} + 2\tilde{\boldsymbol{\gamma}}$. ∎