Journal of Financial Econometrics, 2024, 1–29
https://doi.org/10.1093/jjfinec/nbae023
Article

The Conditional Autoregressive *F*-Riesz Model for Realized Covariance Matrices

Anne Opschoor [1,2], André Lucas [2,3], and Luca Rossini[4,5]

[1]Department of Finance, Vrije Universiteit Amsterdam, 1081 HV, Amsterdam, The Netherlands
[2]Tinbergen Institute, Gustav Mahlerplein 117, Amsterdam, 1082 MS, The Netherlands
[3]Department of Econometrics, Vrije Universiteit Amsterdam, Amsterdam, 1081 HV, The Netherlands
[4]Department of Economics, Management, and Quantitative Methods, University of Milan, Milan, 20122, Italy
[5]Fondazione Eni Enrico Mattei, Italy

Address correspondence to Anne Opschoor, Department of Finance, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands, or e-mail: a.opschoor@vu.nl

Abstract

We introduce a new model for the dynamics of fat-tailed (realized) covariance-matrix-valued time-series using the *F*-Riesz distribution. The model allows for heterogeneous tail behavior across the coordinates of the covariance matrix via two vector-valued degrees of freedom parameters, thus generalizing the familiar Wishart and matrix-*F* distributions. We show that the filter implied by the new model is invertible and that a two-step targeted maximum likelihood estimator is consistent. Applying the new *F*-Riesz model to U.S. stocks, both tail heterogeneity and tail fatness turn out to be empirically relevant: they produce significant in-sample and out-of-sample likelihood increases, ex-post portfolio standard deviations that are in the global minimum variance model confidence set, and economic differences that are either in favor of the new model or competitive with a range of benchmark models.

Keywords: covariance matrix distributions, tail heterogeneity, (Inverse) Riesz distribution, fat-tails, realized covariance matrices

JEL classifications: C32, C58, G17

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, Ledoit, and Wolf 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, Shephard, and Sheppard (2016); Callot, Kock, and Medeiros (2017); Bollerslev, Patton, and Quaedvlieg (2018); Bollerslev et al. (2020); and the references cited therein.

Opschoor thanks the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.201.079. We thank the associate editor and two anonymous referees for their valuable comments, and Francisco Blasques for contributions on the asymptotic analysis of the model. This paper has previously circulated under the title "Tail Heterogeneity for Dynamic Covariance Matrices: the F-Riesz Distribution."

Received: June 27, 2023. **Revised:** September 11, 2024. **Editorial decision:** September 15, 2024. **Accepted:** September 17, 2024

An important challenge is to design parsimonious yet flexible time-series models for such series of realized covariance matrices that can be used for forecasting and decision purposes. 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, Liu, and Chen (2019) and Chen, Yang, and Zhang (2022) 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 are typically highly restrictive. For instance, the often-used Wishart or inverse Wishart distributions for matrix-valued time-series only feature two parameters: a matrix-valued mean, and a single scalar-valued degrees of freedom parameter to describe the tail behavior across all coordinates (Golosnoy, Gribisch, and Liesenfeld 2012; Jin and Maheu 2013, 2016). Similarly restrictive, the matrix-*F* distribution only features two tail parameters for any $k \times k$ realized covariance matrix (Konno 1991; Opschoor et al. 2018). While such distributions might be suitable for low-dimensional cases, in moderate to high dimensions the implied constraints on tail behavior in the cross-section are typically empirically too restrictive.

The typical approach found in the literature to flexibilize multivariate distributions by splitting them into the marginal distributions and a copula (see for instance Patton 2009; Oh and Patton 2017, 2018; Opschoor et al. 2021) cannot easily be applied here. Most copula methods relate to vector-valued observations and cannot deal with the positive definiteness of covariance-matrix-valued observations. Second, many copula structures available in the literature are also tightly parameterized, such as the Gaussian, (skewed) Student's *t*, and Archimedean copulas, with very little heterogeneity in the tail-dependence structure.

In this article, we, therefore, introduce the dynamic *F*-Riesz distribution to model sequences of realized covariance matrices. We do so by building on the beta type II Riesz distribution of Díaz-García (2016).[1] We introduce dynamics for the key scale parameter of the *F*-Riesz distribution and derive the invertibility of the filter and the consistency properties of the maximum likelihood estimator (MLE) for the static parameters of the model. A key property of the *F*-Riesz distribution is that it allows for 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. If each of these vectors is scalar (i.e., has the same elements), then the *F*-Riesz model reduces to the matrix-*F* model (see Konno 1991; Opschoor et al. 2018).

The *F*-Riesz distribution is constructed by mixing a Riesz distribution (Hassairi and Lajmi 2001; Díaz-García 2013) with an Inverse Riesz distribution (Tounsi and Zine 2012; Louati and Masmoudi 2015), both of which are generalizations of the Wishart and Inverse Wishart distributions. The Riesz distribution has thus far mainly been used in the physics literature (Andersson and Klein 2010). In economic statistics, Gribisch and Hartkopf (2023) 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 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 function in closed form and can stick to standard maximum likelihood rather than simulation-based or Bayesian techniques for the estimation of model parameters. Second, we use the more flexible and fat-tailed *F*-Riesz-distribution rather than the Riesz. This allows for additional

[1]We correct the expression for the pdf in the original paper of Díaz-García (2016); see Supplementary Appendix A. The correction relates to the part of the pdf involving the scaling matrix and is therefore crucial in our context, where we consider time-varying scaling matrices.

distributional flexibility, which appears relevant in our application within our class of observation-driven models when comparing the behavior of the *F*-Riesz model with the (Inverse) Riesz.[2]

We apply the dynamic *F*-Riesz distribution to a sample of daily realized covariance matrices of dimensions 5 and 15 using U.S. stock data over the period 2001–2019. The results show that both tail heterogeneity and fat tailedness as captured by the *F*-Riesz distribution are empirically relevant compared to tail heterogeneity only (Riesz) or fat-tailedness only (matrix-*F*). The data strongly reject both the dynamic Riesz and the dynamic matrix-*F* specifications for the dynamics of realized covariance matrices compared to the *F*-Riesz alternative, despite both the Riesz and matrix-*F* already performing significantly better in terms of likelihood fit than the dynamic Wishart and inverse Wishart models. In addition, weekly and biweekly point forecasts of the predicted covariance matrices from the *F*-Riesz distribution are at par or better than those of the alternative model specifications, while its density forecasts measured by the log-scoring rule are clearly better. The one-step-ahead predicted covariance matrices from the *F*-Riesz distribution also perform well in a global minimum variance portfolio strategy: they yield statistically lower ex-post portfolio standard deviations and higher economic gains after taking into account transaction costs compared to competing models and several standard benchmarks such as the HAR-DRD model of Oh and Patton (2016). We conclude that the dynamic *F*-Riesz distribution can prove useful for the statistical analysis of covariance-matrix-valued time-series.

The rest of this article is set up as follows. In Section 1, we introduce the model. Section 2 considers filter invertibility and the consistency properties of the two-step targeted MLE and also studies the new model's performance in a simulated setting. Section 3 presents the empirical results. Section 4 concludes. The Supplementary Appendices gather further technical and empirical results. As a notational convention in this article, we write scalars in normal type face, vectors in bold, and matrices in capitalized bold font.

1 The Conditional Autoregressive *F*-Riesz Model

1.1 Dynamic Model Specification

Let $\boldsymbol{X}_t \in \mathbb{R}^{k\times k}$ for $t = 1,\dots,T$ denote a time-series of realized covariance matrices, and define the filtration $\mathcal{F}_{t-1} = \{\boldsymbol{X}_1,\dots,\boldsymbol{X}_{t-1}\}$ containing all lagged observations of $\boldsymbol{X}_t$. We describe the dynamics of $\boldsymbol{X}_t$ by a conditional, matrix-valued distribution with an autoregressive, time-varying mean,

$$\boldsymbol{X}_t = \boldsymbol{L}_{\boldsymbol{V}_t}\boldsymbol{E}_t\boldsymbol{L}_{\boldsymbol{V}_t}^{\top}, \qquad \boldsymbol{V}_{t+1} = (1-A-B)\boldsymbol{\Omega} + A\boldsymbol{X}_t + B\boldsymbol{V}_t, \tag{1}$$

where $\boldsymbol{L}_{\boldsymbol{V}_t}$ is the lower-triangular Cholesky decomposition of $\boldsymbol{V}_t$, such that $\boldsymbol{V}_t = \mathbb{E}[\boldsymbol{X}_t|\mathcal{F}_{t-1}] = \boldsymbol{L}_{\boldsymbol{V}_t}\boldsymbol{L}_{\boldsymbol{V}_t}^{\top}$, and where $\boldsymbol{E}_t$ is a matrix-valued innovation with conditional expectation equal to the identity matrix, $\mathbb{E}[\boldsymbol{E}_t|\mathcal{F}_{t-1}] = \mathbf{I}_k$. We can view Equation (1) as the matrix version of the scalar MEM model of Engle and Gallo (2006). We explicate the conditional pdf of $\boldsymbol{X}_t$ in Equation (4) further below. The parameter matrix $\boldsymbol{\Omega}$ in Equation (1) is symmetric and positive definite. For simplicity, we take A and B as scalar parameters like in the original DCC model of Engle (2002), but generalizations of this can easily be accommodated. Note that for positive definite $\boldsymbol{\Omega}$ and $\boldsymbol{V}_1$ and for $A > 0$, $B > 0$, and $A + B < 1$, $\boldsymbol{V}_t$ is automatically positive definite for all t. We initialize $\boldsymbol{V}_1$ by the unconditional mean of $\boldsymbol{X}_t$. Also note that more complex dynamic structures can easily be allowed for in Equation (1).

[2]As one of the referees pointed out correctly, a state-space model with a Riesz measurement equation as in Gribisch and Hartkopf (2023) can also result in conditional fat-tailedness of the realized covariance matrix measurements via the mixing with the state innovation. The state innovations in Gribisch and Hartkopf (2023) do not give rise to the closed-form Riesz distribution, making the two models less comparable.

For instance, in the empirical application in Section 3 we incorporate HAR-type dynamics as in Corsi (2009) to better capture the possible long-memory behavior of realized covariance matrices.

If we assume the conditional distribution of $\boldsymbol{X}_t|\mathcal{F}_{t-1}$ to be Wishart, the dynamics in Equation (1) resemble the Conditional Autoregressive Wishart (CAW) model (see Golosnoy, Gribisch, and Liesenfeld 2012). For the Wishart case, the model also collapses to one of the two core equations of the Multivariate HEAVY model of Noureldin, Shephard, and Sheppard (2012). As argued in the introduction, however, assuming an (Inverse) Wishart or even a matrix-*F* distribution for $\boldsymbol{E}_t$ can be too restrictive in terms of the tail behavior it implies. This is why we introduce the *F*-Riesz distribution in Section 1.2 to allow for tail heterogeneity.

Note that model (1) is observation-driven. Therefore, it allows for easy parameter estimation via standard maximum likelihood methods employing a prediction error decomposition using the expression for the *F*-Riesz pdf in Equation (4). To reduce the dimensionality of the optimization, we can use a targeting approach to pre-estimate $\boldsymbol{\Omega}$. This can be done in the following way. Assuming stationarity and the existence of an unconditional first moment of $\boldsymbol{X}_t$, one can take unconditional expectations on both sides of the equations in Equation (1) to obtain $\overline{\boldsymbol{V}} = \mathbb{E}[\boldsymbol{V}_t] = \mathbb{E}[\boldsymbol{X}_t] = \boldsymbol{\Omega}$. This result can be used to estimate $\boldsymbol{\Omega}$ by the sample average of the realized covariance matrices, $\hat{\boldsymbol{\Omega}} = T^{-1}\sum_{t=1}^{T}\boldsymbol{X}_t$. Plugging this expression into the log-likelihood function, the resulting function then only depends on the remaining parameters A and B, plus any parameters describing the conditional distribution of $\boldsymbol{E}_t$.

1.2 The *F*-Riesz Distribution

The family tree of the *F*-Riesz distribution considered in this article is provided in Figure 1. The Wishart and to a lesser extent the matrix-*F* distributions are assumed to be sufficiently well known. The Riesz distribution, however, may be less familiar. Therefore, we first briefly recapitulate the basics of the Riesz distribution before presenting the *F*-Riesz distribution. A more extensive introduction to the different distributions and some more technical results can be found in the Supplementary Appendices A and B.

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 $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top$, with $\nu_i > i-1$ for $i = 1, \ldots, k$. Arguably the easiest way to introduce the Riesz distribution is via its so-called Bartlett decomposition, which is

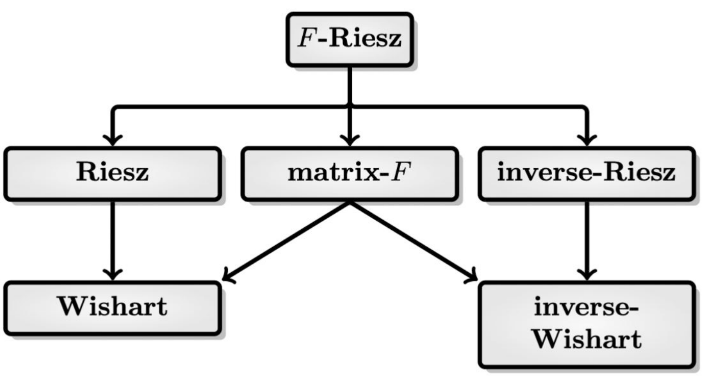

Figure 1. Family of matrix distributions.

Notes: This figure shows a family tree of the *F*-Riesz distributions. Connected lines mean that distributions are related by generalization.

a familiar simulation device for the standard Wishart distribution; see Anderson (1962). Consider a random matrix $\boldsymbol{G}\in\mathbb{R}^{k\times k}$, defined as

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

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, which we write as $\boldsymbol{Y}\sim\mathcal{R}^I(\boldsymbol{\Sigma},\boldsymbol{\nu})$.[3]

Given the use of the Cholesky decomposition, it is clear that the ordering of the variables matters. This is a well-known and accepted feature in the Riesz literature, as the order of the variables can also 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 small dimensions, this can be done by full enumeration. For moderate dimensions, we provide a heuristic algorithm in Section 1.3 that works well in our simulated and empirical setting. Empirically, we find that the step of allowing for tail heterogeneity by generalizing the Wishart to the Riesz or the matrix-*F* to the *F*-Riesz distribution increases the likelihood much more than subsequent likelihood increases due to determining the precise ordering of the coordinates. The latter, in our setting, are typically of second-order importance.

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 $\boldsymbol{G}$ in (2) all have the same value ν. In contrast to the Wishart, the Riesz distribution thus allows for heterogeneous tail behavior in the cross section. Tail fatness, however, is left unaffected and is still exponential (thin) in all directions.

To introduce fatter tails for the Riesz distribution, we draw the analogy between the Wishart and matrix-*F* distribution (Konno 1991). If $\boldsymbol{Y}$ has a Wishart distribution, $\boldsymbol{Y}\sim\mathcal{W}(\mathbf{I}_k,\nu)$, and $\boldsymbol{X}$ given $\boldsymbol{Y}$ also has a Wishart distribution, $\boldsymbol{X}|\boldsymbol{Y}\sim\mathcal{W}(\boldsymbol{Y}^{-1},\mu)$, then the unconditional distribution of $\boldsymbol{X}$ is a matrix-*F*: $\boldsymbol{X}\sim\mathcal{F}(\mathbf{I}_k,\mu,\nu)$. Replacing the Wishart distribution by its generalization, the Riesz, we might expect that a similar result can be obtained which allows for both fat and heterogeneous tails. This is confirmed in Theorem 2.3, which we prove in Supplementary Appendix A.

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

Definition 2.1 (**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 $\overline{\Gamma}(\boldsymbol{\nu})=\pi^{k(k-1)/4}\prod_{i=1}^k\Gamma(\nu_i+\frac{1-i}{2})$, with $2\nu_i>i-1$ for $i=1,\ldots,k$. Similarly, the *upper* generalized multivariate gamma function is defined as $\overline{\Gamma}_U(\boldsymbol{\nu})=\pi^{k(k-1)/4}\prod_{i=1}^k\Gamma(\nu_i+\frac{i-k}{2})=\overline{\Gamma}(\boldsymbol{\nu}+\tilde{\boldsymbol{\gamma}})$, for $2\nu_i>k-i$ for $i=1,\ldots,k$, and $\tilde{\boldsymbol{\gamma}}=\frac{1}{2}\Big(k+1)-(k,k-1,\ldots,1)^\top=(-\frac{1}{2}(k-1),\ldots,\frac{1}{2}(k-1))^\top$.

[3]The type I relates to the fact that we have taken a lower triangular Cholesky decomposition in the Bartlett decomposition. Type II versions of the distribution based on an upper triangular Cholesky decomposition also exist, and we refer to Supplementary Appendix B for details. As the performance of both types of *F*-Riesz distributions was similar in our context, we only discuss the type I *F*-Riesz distributions in the main text.

The upper and lower generalized multivariate gamma functions enter the integrating constant of the Riesz and F-Riesz distributions. If $\boldsymbol{\nu} = (\nu, \ldots, \nu)^\top$, then $\overline{\Gamma}(\boldsymbol{\nu}) = \Gamma_k(\nu)$, where $\Gamma_k(\nu)$ is the standard multivariate gamma function.

Next, we introduce the Lower Power Weighted Determinant (LPWD). The power weighted determinant takes a similar role in the expressions for the density of the F-Riesz distribution as standard determinants do for the Wishart and the matrix-F distribution.

Definition 2.2 (**LPWD**). Consider the vector $\boldsymbol{\nu} \in \mathbb{R}^{k\times 1}$ and a positive definite matrix $\boldsymbol{Y}$. Let $\boldsymbol{L}$ be the (unique) *lower* triangular Cholesky decomposition of $\boldsymbol{Y}$, that is, $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{L}^\top$. Then the LPWD $|\boldsymbol{Y}|_{\boldsymbol{\nu}}$ is given by $|\boldsymbol{Y}|_{\boldsymbol{\nu}} = \prod_{i=1}^{k} \boldsymbol{L}_{i,i}^{2\nu_i}$.

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 article, we instead use the notation of power weighted determinants as it is closer to the econometric literature and stresses the simplification of the Riesz to the Wishart and of the F-Riesz to the matrix-F distribution if $\boldsymbol{\nu} = \nu\boldsymbol{\iota}_k$. Note that the power weighted determinant is *not* as simple as a regular determinant. Properties like $|\boldsymbol{A}\cdot\boldsymbol{B}| = |\boldsymbol{A}|\cdot|\boldsymbol{B}|$ for matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{k\times k}$ either do not hold or hold in modified form for power weighted determinants. Supplementary Appendix Lemma B.1 provides manipulation rules for power weighted determinants.

We now obtain the following formulation of the pdf of the F-Riesz distribution.

Theorem 2.3 (***F*-Riesz distribution**).

i) *If $\boldsymbol{X}|\boldsymbol{Y} \sim \mathcal{R}^I(\boldsymbol{Y}, \boldsymbol{\mu})$ 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 $\mathcal{FR}^I(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$ 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) *Let $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$ for a lower triangular matrix $\boldsymbol{L}$. If $\boldsymbol{X} \sim \mathcal{FR}^I(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \boldsymbol{\nu})$, then $\boldsymbol{L}^{-1}\boldsymbol{X}(\boldsymbol{L}^\top)^{-1} \sim \mathcal{FR}^I(\mathbf{I}_k, \boldsymbol{\mu}, \boldsymbol{\nu})$.*

Theorem 2.3 corrects a result from Díaz-García (2016) on the pdf of the generalized Beta II distribution. The result in that article is only valid for $\boldsymbol{\Sigma} = \mathbf{I}_k$, and incorrect otherwise. The correction is therefore crucial for our application, where we allow for time-varying scale matrices $\boldsymbol{\Sigma}_t$.

Interestingly, except for the use of the non-standard generalized gamma functions and power weighted determinants, the density expression of the F-Riesz one-on-one mirrors and generalizes that of the matrix-F distribution. The following corollary establishes this link and shows that the matrix-F distribution of Konno (1991) as used by Opschoor et al. (2018) is a special case of the F-Riesz distribution.

Corollary 2.4. *Under the conditions of Theorem 2.3 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*

$$p_{\mathcal{FR}^I}(\boldsymbol{X};\boldsymbol{\Sigma},\mu\cdot\boldsymbol{\iota}_k,\nu\cdot\boldsymbol{\iota}_k)=p_{\mathcal{F}}(\boldsymbol{X};\boldsymbol{\Sigma},\mu,\nu)=\frac{\Gamma_k\left(\frac{\mu+\nu}{2}\right)\cdot|\boldsymbol{\Sigma}|^{0.5\nu}}{\Gamma_k\left(\frac{\nu}{2}\right)\Gamma_k\left(\frac{\mu}{2}\right)}\cdot\frac{|\boldsymbol{X}|^{0.5(\mu-k-1)}}{|\boldsymbol{\Sigma}+\boldsymbol{X}|^{0.5(\mu+\nu)}},\tag{3}$$

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 $\boldsymbol{\nu}$.

Similar to the scalar or matrix-*F* distributions, the expectation of the *F*-Riesz does not always exist. The following theorem derives expressions for the expectation and the conditions for its existence. An expression for the first moment of the *F*-Riesz distribution is important for our dynamic model in Equation (1), as $\boldsymbol{E}_t$ is normalized to have the unit matrix as its conditional expectation.

Theorem 2.5 (Expectation of the *F*-Riesz distribution). *Let* $\boldsymbol{Y}\sim\mathcal{FR}^I(\mathbf{I},\boldsymbol{\mu},\boldsymbol{\nu})$, *then* $\mathbb{E}[\boldsymbol{Y}]=\boldsymbol{M}(\boldsymbol{\mu},\boldsymbol{\nu})$, *where* $\boldsymbol{M}(\boldsymbol{\mu},\boldsymbol{\nu})$ *is a diagonal matrix with ith diagonal element* $\boldsymbol{M}_{i,i}(\boldsymbol{\mu},\boldsymbol{\nu})$ *equal to*

$$\boldsymbol{M}_{i,i}(\boldsymbol{\mu},\boldsymbol{\nu})=\begin{cases}\dfrac{\mu_1}{\nu_1-k-1}, & \text{for } i=1,\\[2ex] \dfrac{1}{\nu_i-k+i-2}\left(\mu_i+\sum_{i=1}^{i-1}a_i\right), & \text{for } i=2,\ldots,k,\end{cases}$$

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

Combining this result with Theorem 2.3, we immediately obtain that the disturbance term $\boldsymbol{E}_t$ in our dynamic *F*-Riesz model in Equation (1) should have scale matrix $\boldsymbol{M}_{i,i}(\boldsymbol{\mu},\boldsymbol{\nu})^{-1}$ in order to have conditional expectation I_k. As a result, the conditional pdf of $\boldsymbol{X}_t=\boldsymbol{L}_{V_t}\boldsymbol{E}_t\boldsymbol{L}_{V_t}^\top$ is given by

$$p_{\mathcal{FR}^I}(\boldsymbol{X}_t\,;\,\boldsymbol{L}_{V_t}\boldsymbol{M}(\boldsymbol{\mu},\boldsymbol{\nu})^{-1}\boldsymbol{L}_{V_t}^\top\,,\,\boldsymbol{\mu}\,,\,\boldsymbol{\nu}),\tag{4}$$

with $p_{\mathcal{FR}^I}(\cdot)$ as defined in Equation (1), and where $\boldsymbol{V}_t$ follows the recursion in Equation (1) with lower triangular Cholesky decomposition $\boldsymbol{L}_{V_t}$. The static parameters to be estimated by maximum likelihood now comprise $\boldsymbol{\Omega}$, A, B, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$, where $\boldsymbol{\Omega}$ can be targeted as explained in Section 1.1. For a given set of static parameters, $\boldsymbol{V}_t$ can be obtained from Equation (1) for every t. From these, we can compute the Cholesky decompositions $\boldsymbol{L}_{V_t}$ for every t, which can finally be inserted into the expression for the pdf in Equation (4) to obtain the value of the likelihood.

1.3 Ordering of Variables

The order of the coordinates in $\boldsymbol{X}_t$ matters for the specification of Riesz and *F*-Riesz distributions. As mentioned before, this is well accepted in the Riesz literature due to the use of the Cholesky decomposition in the construction of the Riesz distribution. The order of the coordinates can be regarded as another variable that can be optimized over. For sufficiently small dimensions, enumeration of all possible orders is possible. However, such an approach quickly becomes unwieldy: for $k=10$, we already would have to estimate and compare more than 3.6M models. To approximate the optimal ordering, we therefore propose the following heuristic algorithm.The algorithm above ensures that the maximized

Algorithm 2.6 (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 $\boldsymbol{X}_t$. 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 $\boldsymbol{\theta}$ and obtain $\hat{\boldsymbol{\theta}}^{(j)}$.*
Step 3: *Loop over asset $i, i = 1, \ldots, k$:*

Step 3a: *Find $i^\star$ such that $i = o_{i^\star}^{(j)}$, that is, 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 $o^{(j)}$, that is, 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 $(o_1^{(j)}, \ldots, o_{i^\star-1}^{(j)}, o_{i^\star+1}^{(j)}, \ldots, o_k^{(j)}, o_{i^\star}^{(j)})$. For each re-ordering, re-estimate $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}^{(j,i)}$ and retain the ordering plus estimated $\hat{\boldsymbol{\theta}}^{(j,i)}$ that yields the highest log-likelihood value and store it as $o^{(j+1)}$ and $\hat{\boldsymbol{\theta}}^{(j+1)}$.*

Step 3c: *Increase j to $j + 1$.*

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

Step 4 (optional): *Repeat steps 1–3 p_1 times (or until convergence), possibly repeating the entire process for p_2 different random initial orderings. Retain the final order and estimate that yields the highest log-likelihood. Call this final order $o^{(opt)}$ with corresponding parameter estimate $\hat{\boldsymbol{\theta}}^{(opt)}$.*

likelihood never decreases when searching over different orderings. Moreover, the algorithm is relatively efficient for moderate dimensions since it limits the number of times we re-estimate $\boldsymbol{\theta}$. The latter is costly due to the required non-linear optimization. The algorithm re-estimates $\boldsymbol{\theta}$ about $p_1 p_2 k^2$ times, which is substantially smaller than the full $k!$ enumerated possibilities, thus providing a substantial computational gain. Though no guarantee is given that we arrive at the true optimum using this heuristic algorithm, the simulation evidence in the next section shows that even without using Step 4 non-negligible likelihood increases can be obtained. Also, the algorithm typically lands close to the correct ordering of the variables in terms of rank correlations.

2 Theory and Simulation Evidence

In Section 2.1, we establish the invertibility of the conditional autoregressive F-Riesz (CAFr) filter as defined in Equations (1) and (4) for the dynamic parameters $\boldsymbol{V}_t$. In addition, we prove the consistency of the MLE of the static parameters $\boldsymbol{\Omega}$, A, B, $\boldsymbol{\mu}$, and $\boldsymbol{\nu}$. In Section 2.2, we then study the performance of the MLE and of the heuristic algorithm introduced in Section 1 in a simulated setting.

2.1 Filter Invertibility and MLE Consistency

To establish the consistency of the MLE for the unknown static parameter $\boldsymbol{\theta}$ of the CAFr model,[4] we follow the usual two-step targeting approach that is typically found in

[4]For the i.i.d. case, we provide a separate set of conditions and results in Supplementary Appendix C. We also confine ourselves to proving consistency. Though we expect asymptotic normality of the MLE to hold as well, the expressions for first and second-order derivatives of the F-Riesz pdf are complex in $\boldsymbol{\theta}$ and $\boldsymbol{V}_t$ such that

empirical work and described in Section 1.1. We first estimate $\mathbf{\Omega}$ using a simple sample mean of $\mathbf{X}_t$. Next, fixing this estimate of $\mathbf{\Omega}$, we estimate the remaining parameters by non-linear maximum likelihood optimization.

We make the following assumptions for consistent estimation of $\mathbf{\Omega}$.

Assumption 3.1. *The sequence* $\{\mathbf{X}_t\}_{t=1,\ldots,T}$ *is generated by (4) and (1) for some* $(\mathbf{\Omega}_0, A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0)$ *for every* $t = 1,\ldots,T$.

Assumption 3.2. $\mathbf{\Omega}_0$ *is positive definite,* $\mu_{0,j} > K+1\,\forall j$, $\nu_{0,i} > i+1\,\forall i$, $A_0 > 0$, $B_0 \geq 0$, *and* $|A_0 + B_0| < 1$.

Assumption 3.1 is a standard assumption on correct specification. Assumption 3.2 then allows us to establish the stationarity and ergodicity of the model as a data generating process. The strong consistency of the sample average $\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t$ to $\mathbf{\Omega}_0$ then follows by an application of the ergodic theorem.

Proposition 3.3. *Let Assumptions 3.1–3.2 hold. Then* $\hat{\mathbf{\Omega}} = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t \xrightarrow{a.s.} \mathbf{\Omega}_0$ *as* $T \to \infty$.

The consistent estimate of $\mathbf{\Omega}$ can be used as a plug-in for a targeted estimation approach for the remaining static parameters of the model.

We now turn to the invertibility of the filtering Equation (1). To do so, we first introduce some new notation. Let $\hat{\mathbf{V}}_t(\boldsymbol{\theta})$ denote the filtered sequence from (1), initialized at some point $\hat{\mathbf{V}}_1$, and evaluated at some parameter vector $\boldsymbol{\theta} \in \mathbf{\Theta}$. Following the literature (e.g., Straumann and Mikosch 2006; Wintenberger 2013), invertibility ensures that the filter "forgets" the possibly incorrect initialization; that is, the filtered sequence $\{\hat{\mathbf{V}}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ converges path-wise and exponentially fast to a unique stationary and ergodic limit sequence $\{\mathbf{V}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$. This means that for every $\boldsymbol{\theta}$ in the parameter space $\mathbf{\Theta}$ there is a $c > 1$ such that $c^t||\hat{\mathbf{V}}_t(\boldsymbol{\theta}) - \mathbf{V}_t(\boldsymbol{\theta})|| \xrightarrow{a.s.} 0$ as $t \to \infty$, regardless of the initialization $\hat{\mathbf{V}}_1$. We also write $\hat{\mathbf{V}}_t = \hat{\mathbf{V}}_t(\boldsymbol{\theta}_0)$ and $\mathbf{V}_t = \mathbf{V}_t(\boldsymbol{\theta}_0)$, such that the filter asymptotically recovers the true $\mathbf{V}_t$ series from the data generating process if the filter is evaluated at the true static parameter $\boldsymbol{\theta}_0$.

In the current setting, filter invertibility can be obtained by ensuring that the following conditions hold: (i) stationarity of the data $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$; (ii) a *logarithmic bounded moment* for $\mathbf{X}_t\,\forall t$; (iii) a *contraction condition* for the filtering equation. The stationarity of the data in (i), and the logarithmic moment in (ii) follow directly from Assumptions 3.1 and 3.2. The contraction condition for the filtering equation, however, requires additional restrictions on the parameter space $\mathbf{\Theta}$. Assumption 3.4 ensures that the filtered $\hat{\mathbf{V}}_t(\boldsymbol{\theta})$ matrices are positive definite and that the stochastic filtering equation is contracting in the appropriate sense. It also ensures identification of the ordering of the variables in the system by requiring the degrees of freedom parameters to be different across coordinates i and j.

Assumption 3.4. *The parameter space* $\mathbf{\Theta}$ *is compact and satisfies* $A \geq 0$, $B \geq 0$, $\sup_B |B| < 1$, *and* $\min_j \inf_{\mu_j}(\mu_j - K - 1) > 0$, $\min_i \inf_{\nu_i}(\nu_i - i - 1) > 0$, $\mu_i \neq \mu_j$ *for* $i \neq j$, *and* $(\mu_i + \nu_i) \neq (\mu + j + \nu_j)$ *for* $i \neq j$.

establishing the asymptotic normality in a tractable way is hard without high-level assumptions. We leave this part for further research. In the empirical section, we use the standard sandwich estimator for the standard errors.

Proposition 3.5 now establishes the invertibility of the initialized filter $\hat{V}_t(\boldsymbol{\theta})$ for its stationary and ergodic limit $V_t(\boldsymbol{\theta})$ and opens the door to the consistency of the MLE.

Proposition 3.5. *Let Assumptions 3.1-3.4 hold. Then the filter $\{\hat{V}_t(\boldsymbol{\theta})\}_{t\in\mathbb{N}}$ is invertible.*

We are now ready to formulate our consistency result of the MLE $(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T)$. This MLE takes the form of a targeted two-step estimator as it depends on the first-step estimator for $\boldsymbol{\Omega}_0$. As is common for filtering models, the log-likelihood depends directly on the properties of the filter $\hat{V}_t(\boldsymbol{\theta})$, 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. To make this clearer in the notation, we write explicitly $\hat{V}_t(\hat{\boldsymbol{\Omega}}_T, A, B)$ rather than $\hat{V}_t(\boldsymbol{\theta})$. In addition, we write $\hat{o}_T$ as the ordering of coordinates that maximizes the log-likelihood. Putting all elements together, we define the MLE as the maximizer of the plug-in log-likelihood $\log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu})$ for a specific ordering o,

$$(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T, \hat{o}_T) = \arg \max_{(A,B,\boldsymbol{\mu},\boldsymbol{\nu},o)} \sum_{t=2}^{T} \log \tilde{p}_{\mathcal{FR}^I}(\boldsymbol{X}_t; \hat{\boldsymbol{V}}_t(\hat{\boldsymbol{\Omega}}_T, A, B), \boldsymbol{\mu}, \boldsymbol{\nu}). \quad (5)$$

Theorem 3.6. *Let Assumptions 3.1–3.4 hold. Then for $T \to \infty$ the targeted MLE $(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T, \hat{o}_T)$ defined in (5) satisfies*

$$(\hat{A}_T, \hat{B}_T, \hat{\boldsymbol{\mu}}_T, \hat{\boldsymbol{\nu}}_T, \hat{o}_T) \quad \overset{a.s.}{\to} \quad (A_0, B_0, \boldsymbol{\mu}_0, \boldsymbol{\nu}_0, o_0).$$

Theorem 3.6 provides the consistency of the MLE. Though the simulation results in Table 1 suggest that the ML estimator for the static parameters can be well approximated by a normal distribution in finite samples, we instead choose to focus on the predictive performance of the model using Diebold–Mariano (DM) tests (Diebold and Mariano 1995) rather than on the behavior of the static parameters. The latter is typically deemed of less

Table 1. Parameter estimations of CAFr model

Coef.	True	Mean	Std	Mean (S.E.)
A	0.1600	0.1596	0.0048	0.0044
B	0.8300	0.8296	0.0049	0.0049
μ_1	16.64	16.69	1.125	1.128
μ_2	27.15	27.07	1.567	1.544
μ_3	41.61	41.52	2.284	2.229
μ_4	58.18	58.03	3.073	3.030
μ_5	84.67	84.12	4.402	4.381
ν_1	20.05	20.45	1.584	1.510
ν_2	18.72	18.94	0.957	0.880
ν_3	19.36	19.57	0.780	0.722
ν_4	20.59	20.85	0.782	0.734
ν_5	14.61	14.77	0.512	0.442

Notes: This table shows Monte Carlo averages and standard deviations of parameter estimates of simulated covariance matrices from the five-dimensional CAFr model of (1). Guided by empirical results, we set $\boldsymbol{\mu} = (16.64, 27.15, 41.61, 58.18, 86.67)$ and $\boldsymbol{\nu} = (20.05, 18.72, 19.36, 20.59, 14.61)$. We estimate Ω by targeting in a first step, while A, B, 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 using the inverse of the Hessian. Results are based on 1000 Monte Carlo replications.

interest in dynamic parameter models such as the CAFr, where the focus is mostly on the filtered paths of $\hat{V}_t$ and the model's predictive performance. Note that the consistency of the filtered paths $\hat{V}_t$ follows directly from the consistency of the MLE and the filter invertibility established earlier.

We use the DM test based on two loss functions. First, we use the log-scoring rule $d_t = \ell_t^1 - \ell_t^2$, where ℓ_t^1 and ℓ_t^2 are the log-likelihood contributions of two different model specifications. The test requires d_t to be a finite variance martingale difference series under the null of equal model performance. The existence of a second moment of the log-likelihood for the *F*-Riesz distribution is easily obtained using similar arguments as for the consistency proof, where a bounded first moment of the log-likelihood was established. Second, we use a more economic perspective to compare the different models by constructing global minimum variance portfolios and comparing their ex-post portfolio variance performance using the DM test. See Section 3.4 for further details.

2.2 Simulation Experiment

This section presents the results of a Monte Carlo study for the statistical properties of the MLE of the conditional autoregressive *F*-Riesz (CAFr) model. We simulate from a $k = 5$ dimensional version of the CAFr model with empirically relevant values for the static parameters. We do so 1000 times and for each simulated series estimate the static parameters of the model by MLE, as well as their standard errors.

Table 1 presents the results. We see that all parameters are estimated near their true values. This holds both for the dynamic parameters A and B, as well as for the degrees of freedom parameters $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, underlining the consistency result from Section 2. We also note that the Monte Carlo standard deviation of the MLE across simulations (in the *std* column) is close to the average of the estimated standard errors using the inverse Hessian (in the *mean(s.e.)* column).

The second 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 five-dimensional *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 $\tilde{\boldsymbol{\nu}} = \bar{\nu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\nu}_{range}$ 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 the Likelihood-Ratio test with $2 \times k - 2$ degrees of freedom. We repeat this exercise 1000 times.

Table 2 shows the results. In Panel A, we see that if we simulate from a matrix-*F* distribution (ie $\lambda = 0$), 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 correspond to the simulated values of 69.2 and 23.25.

In Supplementary Appendix E, we present further simulation results showing that (i) the targeting approach and full estimation approach for $\boldsymbol{\Omega}$ perform similarly well; (ii) all distributions (from Wishart to *F*-Riesz) exhibit a similar quality of the MLE in finite samples if that distribution is correctly specified; (iii) the heuristic approach for ordering the variables results in further likelihood increases; (iv) even running the heuristic once already closes much of the gap with the optimal ordering in small dimensional settings; (v) the rank

Table 2. The matrix F versus the F-Riesz distributions

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				
	Std	5.72	0.63				
F-Riesz		$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$	$\boldsymbol{\mu}_3$	$\boldsymbol{\mu}_4$	$\boldsymbol{\mu}_5$	
	True	69.20	69.20	69.20	69.20	69.20	
	Mean	69.60	69.47	69.54	69.44	69.42	
	Std	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	
	Std	0.99	0.91	0.99	1.17	1.52	

Notes: 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 the standard deviation of the estimated coefficients. All results are based on 1000 Monte Carlo replications.

correlations between the true ordering and the ordering found by the heuristic algorithm is 99% on average, and 90% of the top 5 ranks are correctly identified in the $k = 15$ dimensional case. We thus feel confident to proceed with our empirical application.

3 Empirical Application

3.1 Data and Setup

In this section, we apply the F-Riesz distribution to an empirical data set of 45 U.S. equities from the S&P 500 index over the period January 2, 2001, until December 6, 2019, a total of 4,696 trading days. We extract transaction prices from the Trade and Quote 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. We refer to Supplementary Appendix F for more information about the data (Tickers) and the cleaning procedure.

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. We also allow our dynamic F-Riesz model from Equations (1) and (4) to include HAR-type dynamics by considering an extension of Equation (1) to

$$\boldsymbol{V}_{t+1} = (1 - A_1 - A_2 - A_3 - B)\boldsymbol{\Omega} + A_1\boldsymbol{X}_t + A_2\boldsymbol{X}_t^w + A_3\boldsymbol{X}_t^m + B\boldsymbol{V}_t, \tag{6}$$

with $\boldsymbol{X}_t^w = (1/5)\sum_{i=0}^4 \boldsymbol{X}_{t-i}$ and $\boldsymbol{X}_t^m = (1/22)\sum_{i=0}^{21} \boldsymbol{X}_{t-i}$, respectively. We use the two-step targeting approach from Section 1.1 to estimate $\boldsymbol{\Omega}$, and the algorithm from 2.3 with $p_1 = p_2 =$

Table 3. Parameter estimates, likelihoods, and information criteria

Distribution	A	B	μ_{min}	μ_{max}	ν_{min}	ν_{max}	$\mathcal{L}$	AIC	No. para
Panel A: XOM/PG/WMT/PFE/MCD									
Wishart	0.281	0.691			16.09		−5178	10,362	3
	(0.004)	(0.004)			(0.075)				
Riesz	0.260	0.713			8.53	19.57	−3942	7898	7
	(0.004)	(0.004)			(0.170)	(0.157)			
i-Wishart	0.195	0.791			18.63		1162	−2319	3
	(0.003)	(0.004)			(0.067)				
i-Riesz	0.190	0.796			13.09	20.45	1493	−2971	7
	(0.003)	(0.004)			(0.223)	(0.155)			
F	0.215	0.770	56.90		26.31		2229	−4450	4
	(0.003)	(0.004)	(1.350)		(0.277)				
F-Riesz	0.166	0.823	15.49	82.74	13.63	22.85	7220	−14,416	12
	(0.003)	(0.003)	(0.460)	(1.681)	(0.174)	(0.485)			
Panel B: JPM/GE/HON/BA/IBM/XOM/CAT/HD/PG/KO/AA/WMT/AXP/MCD/PFE									
Wishart	0.176	0.810			27.83		96,115	−192,224	3
	(0.001)	(0.001)			(0.041)				
Riesz	0.155	0.830			6.27	36.34	113,988	−227,942	17
	(0.001)	(0.001)			(0.123)	(0.149)			
i-Wishart	0.100	0.895			32.10		143,001	−285,995	3
	(0.001)	(0.001)			0.03				
iRiesz	0.095	0.899			10.57	35.34	151,316	−302,599	17
	(0.001)	(0.001)			(0.170)	(0.140)			
F	0.116	0.879	78.25		46.26		159,642	−319,276	4
	(0.001)	(0.001)	(0.473)		0.14				
F-Riesz	0.087	0.908	12.19	105.2	14.09	46.60	185,170	−370,275	32
	(0.001)	(0.001)	(0.260)	(0.702)	(0.271)	(0.539)			

Notes: This table reports maximum likelihood parameter estimates of the conditional autoregressive models (4)–(1), assuming a Wishart, Riesz, Inverse Wishart, Inverse Riesz, matrix-*F*, or *F*-Riesz distribution in (4). Data consist of realized covariance matrices with the optimal ordering based on the algorithm from 2.3 with $p = 1$ on the CAFr model. Panels A and B list results for a randomly chosen subset of 5 and 15 different assets, respectively. Standard errors are provided in parentheses and based on the (sandwich) robust covariance matrix estimator. We report the likelihood $\mathcal{L}$, the AIC, and the number of estimated parameters. The sample goes from January 2, 2001, until December 12, 2019 ($T = 4696$ trading days).

1 to determine the ordering of the variables in the CAFr model. We use the same ordering for the other models.

3.2 Full Sample Results

Table 3 and Figure 2 report the results for the full sample. For each dimension ($k = 5, 15$), we randomly choose stocks (without replacement) from our pool of 45 assets. We only present parameter estimates for models with dynamics as in Equation (1) and compare their fit graphically afterward to models with the HAR specification from Equation (6).

The results provide four main takeaways. First, the maximized log-likelihood values show that the model with the *F*-Riesz distribution performs better than all the other specifications, including the Riesz distribution.[5] This is most clearly seen in Figure 2, which shows the AIC improvements of all models compared to the Wishart specification for both the original (1) and the extended HAR specification (6). 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

[5]Note that our results and those of Gribisch and Hartkopf (2023) cannot be compared directly, as they use a parameter-driven model specification, whereas ours is an observation-driven approach.

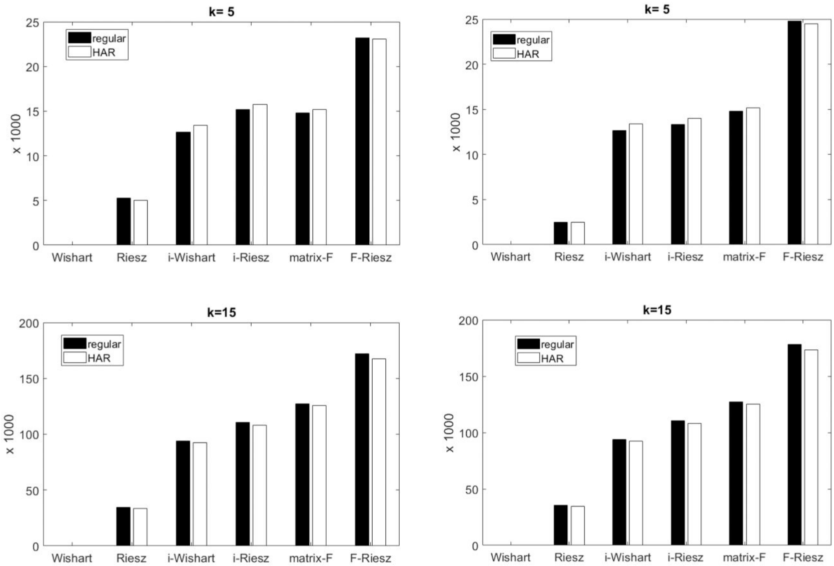

Figure 2. AIC improvements.

Notes: The figure shows the difference between the AIC of the Wishart and that of the other distributions for the models in Table 3 (black bars) and their HAR extension of (6) (white bars). The left panel of the graph depicts results of a random initial ordering of the constructed realized covariance matrices. The right panel is based on the optimized ordering using the algorithm from Section 1.3.

and the matrix-*F* distribution equals 5000 and 25,000 log-likelihood points for dimensions 5 and 15, respectively. This increase suggests substantial heterogeneity *and* fatness of the tails. The AIC values underline that the large likelihood differences outweigh the increased number of 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, irrespective of the ordering of the assets; see Figure 2. This underlines the importance of tail heterogeneity. The same holds when relaxing the inverse Wishart to the Inverse Riesz. Tail fatness is also clearly important: the AIC improvement for the matrix-*F* is large compared to the Wishart. With only two parameters, the matrix-*F* succeeds in having a similar or higher AIC as the Inverse Riesz, which needs $k-2$ additional parameters compared to the matrix-*F*. 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. Including tail heterogeneity in the matrix-*F* by using the *F*-Riesz distribution provides a further substantial gain in likelihood and AIC. Tail heterogeneity thus appears important for both the thin and fat-tailed distributional specifications.

Third, the importance of allowing for tail heterogeneity is confirmed by looking at the estimates of the degrees of freedom parameters. To save space, 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 55, while the elements of $\boldsymbol{\mu}$ of the *F*-Riesz distribution vary from around 16–83. 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 Table 3, rising the dimension of the system from 5 to 15 increases 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 and B. 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/matrix-*F* can only accommodate this by using some sort of average degrees of freedom across all assets due to their one or two parameter set-up, the degrees of freedom for these two distributions increases empirically when increasing the number of assets. By contrast, the *F*-Riesz (and also the (Inverse) Riesz) distributions do not show this behavior.

Fourth, we see that the heterogeneity biases discussed above 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 distribution. Again, this stems from 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 Inverse Riesz. 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 fail to describe the heavy-tailed behavior in some of the realized volatilities. Empirically, this typically leads to a lower estimated persistence due to the more frequent unexpected occurrence of incidental large observations. As a result, the *F*-Riesz and Riesz have a higher persistence *B* compared to the matrix-*F* and Wishart, respectively.

3.3 Out-of-sample Setup and Metrics

We also apply our new model in an out-of-sample exercise. First, we calculate daily, weekly, and biweekly point forecasts for $\boldsymbol{V}_t$. Second, we forecast the (joint) density of the realized covariance matrices. Third, we conduct an economic application by considering Global Minimum Variance Portfolios (GMVP) as in for example Engle, Ledoit, and Wolf (2019). All these exercises directly depend on one-step and multi-step-ahead forecasts of $\boldsymbol{V}_t$.

We use a moving-window approach in the forecasting exercise with an in-sample period of 1000 observations. This corresponds roughly to four calendar years. To avoid that the results are driven by a particular selection of stocks, we choose three disjoint sets of stocks for each of the settings $k = 5, 15$.

The out-of-sample period contains $P = 3,696$ observations including the Great Financial Crisis and the European Sovereign Debt crisis. The period, therefore, provides an important test for the robustness of the model. We re-estimate the models after every 250 observations, which roughly corresponds to updating the parameters annually. As we show later on, though optimizing the ordering of the variables empirically leads to some gains in likelihood, there are no substantial gains in *point prediction* quality compared to a single random ordering. We therefore consider only one random initial ordering for the main results.

Based on Table 3, we consider a subset of five distributions: the classical Wishart distribution as a benchmark, the Riesz, the Inverse Riesz, the matrix-F, and the *F*-Riesz distribution. Moreover, we include the EWMA filter with $\lambda = 0.96$, the DCC-GARCH model (Engle 2002), and the HAR-DRD model of Oh and Patton (2016) as further benchmark models.

Given a predicted $\boldsymbol{V}_{t+1}$ and the true realized covariance matrix $\boldsymbol{X}_{t+1}$, we evaluate the point forecasts by the Frobenius norm and the QLIK loss function:

$$FR_{t+1} = \sqrt{\text{trace}\Big((\boldsymbol{X}_{t+1} - \boldsymbol{V}_{t+1})^\top (\boldsymbol{X}_{t+1} - \boldsymbol{V}_{t+1})\Big)}, \tag{7}$$

$$QLIK_{t+1} = \log|\boldsymbol{V}_{t+1}| + \text{trace}(\boldsymbol{V}_{t+1}^{-1}\boldsymbol{X}_{t+1}). \tag{8}$$

We use the model confidence set (MCS) of Hansen, Lunde, and Nason (2011) at a significance level of 5% to select the best models. The MCS automatically accounts for the dependence between model outcomes, given that all models are based on the same data.

Turning toward the density forecasts, we use the log scoring rule (see Mitchell and Hall 2005; Amisano and Giacomini 2007) to differentiate between the density forecasts of the different models. Define the difference in log score between the two density forecasts M_1 and M_2 corresponding to the realized covariance matrix $\boldsymbol{X}_{t+1}$ as

$$d_{ls,t+1} = S_{ls,t+1}(\boldsymbol{X}_{t+1}, M_1) - S_{ls,t+1}(\boldsymbol{X}_{t+1}, M_2), \tag{9}$$

for $t = R+1, \ldots T$, with $R = 1000$ the length of the rolling estimation window and $S_{ls,t+1}(\boldsymbol{X}_{t+1}, M_j)$ $(j = 1,\, 2)$ the log score of the density forecast corresponding to model M_j at time $t+1$,

$$S_{ls,t+1}(\boldsymbol{X}_{t+1}, M_j) = \log p_{t+1}(\boldsymbol{X}_{t+1}|\boldsymbol{V}_{t+1}, \mathcal{F}_t, M_j), \tag{10}$$

where $p_{t+1}(\cdot)$ is one of the densities discussed. For multi-step ahead density forecasts, we consider the joint density $\prod_{h=1}^{H} p_{t+h}(\boldsymbol{X}_{t+h}|\boldsymbol{V}_{t+h}, \mathcal{F}_{t+h-1}, M_j)$ such that the log score boils down to $S_{ls,t+1:t+H}(\boldsymbol{X}_{t+1} : \boldsymbol{X}_{t+H}, M_j) = \sum_{h=1}^{H} \log p_{t+h}(\boldsymbol{X}_{t+h}|\boldsymbol{V}_{t+h}, \mathcal{F}_{t+h-1}, M_j)$. The null hypothesis of equal predictive ability is given by $H_0 : \mathbb{E}[d_{ls}] = 0$ for all T—R out-of-sample forecasts. Similar to the evaluation of the point forecasts, we again use the MCS with a 5% significance level to test equal predictive ability.

The GMVP application is motivated by the mean-variance optimization setting of Markowitz (1952). Assuming that the investor aims at minimizing the 1-step-ahead portfolio variance at time t subject to a fully invested portfolio, we have the quadratic optimization problem

$$\begin{cases} \min \boldsymbol{w}_{t+1|t}^\top \boldsymbol{V}_{t+1} \boldsymbol{w}_{t+1|t}, \\ \text{s.t.} \quad \boldsymbol{w}_{\text{t}+1|\text{t}}^\top \boldsymbol{\iota} = 1, \end{cases} \quad \Rightarrow \quad w_{t+1|t} = \frac{\boldsymbol{V}_{t+1|t}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}^\top \boldsymbol{V}_{t+1|t}^{-1}\boldsymbol{\iota}}. \tag{11}$$

We assess the predictive ability of the different models by comparing the results to the ex-post portfolio volatility $\sigma_{p,t+1} = (\boldsymbol{w}_{t+1|t}' \boldsymbol{X}_{t+1} \boldsymbol{w}_{t+1|t})^{1/2}$ using the MCS.

Alongside the GMVP's volatility, we also calculate several other relevant quantities, such as portfolio turnover (TO_t), concentration (CO_t), and the total short position (SP_t) for each competing model at time t. A model that produces more stable covariance matrix forecasts implies in general less turnover and hence less transaction costs. This effect would lead to a gain in trading strategies. We follow Bollerslev, Patton, and Quaedvlieg (2018) and assume that there is a fixed transaction cost c. The total turnover at time t is defined as

$$TO_t = \sum_{i=1}^{k} \left| w_{i,t+1|t} - w_{i,t|t-1} \frac{1 + r_{i,t}}{1 + \boldsymbol{w}_{t|t-1}^\top r_t} \right|, \tag{12}$$

where $w_{i,t|t-1}$ is the i-th element of the weight vector $\boldsymbol{w}_{t|t-1}$ and $r_{i,t}$ the return of asset i at time t. It measures the value of the change in portfolio holdings when rebalancing the portfolio to its new optimal position from time t to $t+1$. For given proportional transaction costs c, the portfolio return net of transaction costs then equals

$$r_{p,t} = \boldsymbol{w}_{t|t-1}^{\top} r_t - c\,TO_t, \tag{13}$$

with c equal to 0, 1, or 2 percent, respectively.

Portfolio concentration and total portfolio short position both measure the amount of extreme portfolio allocations. Again, more stable forecasts of $\boldsymbol{V}_{t+1}$ should result in less extreme portfolio weights. The portfolio concentration (CO_t) and short position (SP_t) are defined as

$$CO_t = \left(\sum_{i=1}^{k} w_{i,t|t-1}^2\right)^{1/2}, \qquad SP_t = \sum_{i=1}^{k} w_{i,t|t-1} \cdot I[w_{i,t|t-1} < 0], \tag{14}$$

with $I[\cdot]$ an indicator function that takes the value 1 if the i-the element of the weight vector is negative.

Finally, we follow Bollerslev, Patton, and Quaedvlieg (2018) and also evaluate the economic significance of different forecasting models by considering the utility-based framework of Fleming, Kirby, and Ostdiek (2001, 2003). This framework is based on the assumption that an investor has a quadratic utility with a risk aversion parameter γ. Then the realized utility of the portfolio return based on the forecasted covariances from model j reads

$$U(r_{p,t}^{j}, \gamma) = (1 + r_{p,t}^{j}) - \frac{\gamma}{2(1+\gamma)}(1 + r_{p,t}^{j})^2. \tag{15}$$

Given two different models j and l, Δ_γ denotes the return an investor with risk aversion parameter γ is willing to forfeit for model l to make her indifferent between models j to l. It can be obtained by solving

$$\sum_{t=1}^{P} U(r_{p,t}^{j}, \gamma) = \sum_{t=1}^{P} U(r_{p,t}^{l} - \Delta_\gamma, \gamma). \tag{16}$$

We test the null hypothesis $\Delta_\gamma = 0$ using the Reality Check of White (2000), based on the stationary bootstrap of Politis and Romano (1994) for 999 bootstrap samples with an average block length of 22 days.

Before presenting the main out-of-sample results, we first briefly investigate the effect of the ordering of the variables on point forecast performance. We do so for two sets of dimension 5. Table 4 shows the results for three settings: an arbitrary ordering, the optimal ordering over all 120 possibilities, and a time-varying ordering scheme based on Algorithm 2.6 implemented every 250 observations in line with the re-estimation of the model's static parameters. The short summary of the table is that all approaches behave similarly in terms of the Frobenius norm: none of the models falls outside the MCS based on the simulated p-values. The same can be said in terms of the QLIK if we consider the left-panel of the table. Only in case of the assets AA, AXP, BA, CAT, and GE, the original and time-varying ordering scheme do not belong to the confidence set. We conclude that the optimal ordering does not automatically lead to better point predictions. In addition, varying the ordering over time also does not seem to lead to significant improvements in point forecasts. In the remainder of the analysis, we, therefore, stick to one specific order for the model

Table 4. Ordering and out-of-sample point forecasts

	MCD/PFE/PG/WMT/XOM			AA/AXP/BA/CAT/GE		
	Original	Optimal	TV ord	Original	Optimal	TV ord
Frob	5.656	**5.648**	5.658	2.280	**2.278**	2.283
mcs p-value	**(0.36)**	**(1.00)**	**(0.31)**	**(0.37)**	**(1.00)**	**(0.12)**
QLIK	**7.477**	7.480	7.478	4.364	**4.361**	4.366
mcs p-value	**(1.00)**	**(0.06)**	**(0.44)**	(0.00)	**(1.00)**	(0.00)

Notes: This table shows the mean of the Frobenius norm and QLIK values based on daily predictions of the (5×5) covariance matrix of two sets of five stocks according to the Conditional Autoregressive model assuming a F-Riesz (FR) distribution using three different approaches. The first approach takes simply the original ordering created by the construction of the realized covariance matrix. The second ordering corresponds with the lowest Frobenius Norm. Finally, the third approach uses a time-varying ordering scheme (TV ord), by applying Heuristic 2.6 iteratively after 250 observations. Parameters are estimated with a moving window of 1,000 observations and re-estimated after 250 observations. The lowest value of the Frobenius norm and QLIK across the models are marked in bold. In addition, we report the MCS p-values based on a 5% significance level. The p-values of the models within the model confidence set are marked in bold. The out-of-sample period goes from January 2005 until December 2019 and contains 3696 observations.

comparisons (as indicated in the tables), keeping in mind that the results might be improved somewhat further in favor of the F-Riesz and Riesz models by re-ordering the variables.

3.4 Out-of-sample Results

We now turn to the discussion of our point and density forecasts, followed by the results of the GMV forecasts. Given that we have three data sets of dimensions 5 and 15, we have many results. For reasons of space, we only present results for one selection of 5 and 15 assets. The remaining results can be found in Supplementary Appendix G and do not change the main conclusions.

We first consider the point forecasts. For point forecasts, it is not clear a priori why a good model for tail behavior would improve the point forecasts. Tables 5 and 6 show the mean of the Frobenius norm and QLIK values with the associated MCS p-values for daily, weekly, and biweekly covariance forecasts for the first set of 5 and 15 assets. Supplementary Appendix Tables G.1–G.4 show the results for the other two sets of 5 and 15 assets. All tables show a similar pattern. For 1-step-ahead forecasts, the HAR-DRD model is superior. For the Frobenius norm, it is often accompanied in the MCS for $k = 5$ by the conditional autoregressive models with one of the matrix distributions and HAR dynamics, illustrating that the point forecasts for the different models are comparable 1-day-ahead. This is not surprising as the parameters of the HAR-DRD are estimated by minimizing the Frobenius norm. The HAR-DRD model loses its superiority for point forecasts, however, for weekly and biweekly point forecasts. Here, the F-Riesz distribution always belongs to the MCS, either with the classical dynamic specification in Equation (1) or with the HAR specification from Equation (4). This does not hold for other matrix distributions, such as the Riesz distribution. Finally, we see that the forecasts of the EWMA and DCC-GARCH forecast are relatively worse than the other models as they are rarely within the MCS.

Next, we consider the models' density forecast performances. In contrast to milder differences in the point forecasts, the results for the density forecasts exhibit much stronger differences between the different models. Table 7 (as well as Supplementary Appendix Table G.5) reports results on the daily, weekly, and biweekly density forecasts. The tables list the average log-score values together with the MCS p-values. The message from the table is clear: the F-Riesz distribution is superior in density forecasts against all other matrix

Table 5. Out-of-sample point forecasts (dimension 5, set I)

	CA specifications					EWMA	DCC	CA-HAR specifications					HDRD
	W	R	iR	F	FR			W	R	iR	F	FR	
Panel A: daily forecasts													
Frob	2.274	2.276	2.275	2.274	2.280	2.627	2.620	2.250	2.252	2.248	**2.246**	2.253	2.231
mcs p-value	(0.02)	(0.01)	(0.01)	**(0.05)**	(0.01)	(0.01)	(0.00)	**(0.43)**	**(0.37)**	**(0.43)**	**(0.66)**	**(0.34)**	**(1.00)**
QLIK	4.354	4.353	4.358	4.356	4.364	4.525	4.911	4.341	4.341	4.345	4.343	4.351	**4.325**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(0.07)**	**(0.07)**	**(0.02)**	**(0.06)**	(0.00)	**(1.00)**
Panel B: weekly forecasts													
Frob	9.369	9.314	9.271	9.303	9.295	10.799	11.230	9.259	9.240	**9.229**	9.239	9.288	9.507
mcs p-value	**(0.25)**	**(0.61)**	**(0.87)**	**(0.61)**	**(0.64)**	**(0.11)**	(0.02)	**(0.79)**	**(0.87)**	**(1.00)**	**(0.87)**	**(0.64)**	**(0.61)**
QLIK	12.578	12.573	12.572	12.573	12.574	12.674	13.151	12.567	**12.564**	12.568	12.569	12.574	12.575
mcs p-value	**(0.11)**	**(0.25)**	**(0.33)**	**(0.21)**	**(0.21)**	(0.01)	(0.00)	**(0.42)**	**(1.00)**	**(0.42)**	**(0.42)**	**(0.21)**	**(0.42)**
Panel C: biweekly forecasts													
Frob	18.937	18.723	**18.547**	18.676	18.557	20.814	21.906	19.273	19.252	18.738	18.830	18.782	18.953
mcs p-value	(0.05)	**(0.39)**	**(1.00)**	**(0.39)**	**(0.92)**	(0.01)	(0.01)	(0.01)	(0.04)	**(0.67)**	**(0.34)**	**(0.39)**	**(0.39)**
QLIK	16.175	16.165	16.160	16.165	16.162	16.218	16.718	16.162	16.158	**16.150**	16.153	16.155	16.173
mcs p-value	**(0.10)**	**(0.30)**	**(0.32)**	**(0.27)**	**(0.32)**	(0.06)	(0.01)	**(0.32)**	**(0.32)**	**(1.00)**	**(0.32)**	**(0.32)**	**(0.32)**

Notes: This table shows the mean of the Frobenius norm and QLIK values based on daily, weekly, and biweekly predictions of the (5×5) covariance matrix of 5 stocks (MCD/PFE/PG/WMT/XOM) according to the Conditional Autoregressive model with and without HAR dynamics, assuming a Wishart (W), Riesz (R), Inverse Riesz (iR), matrix-F (F), or a F-Riesz (FR) distribution. In addition, we include the EWMA, DCC-GARCH, and the HAR-DRD (HDRD) models as benchmarks. Parameters are estimated with a moving window of 1,000 observations and re-estimated after 250 observations. Panel A shows results of 1-step ahead predictions, and Panels B and C list results of cumulative forecasts for a week (five-step ahead) and two weeks (10 steps ahead). The lowest value of the Frobenius norm and QLIK across the models are marked in bold. In addition, we report the MCS p-values based on a 5% significance level. The p-values of the models within the model confidence set are marked in bold. The out-of-sample period goes from January 2005 until December 2019 and contains 3,696 observations.

Table 6. Out-of-sample point forecasts (dimension 15, set I)

	CA specifications					EWMA	DCC	CA-HAR specifications					HDRD
	W	R	iR	F	FR			W	R	iR	F	FR	
Panel A: daily forecasts													
Frob	10.411	10.456	10.536	10.486	10.603	11.801	12.373	10.311	10.359	10.431	10.372	10.504	**10.150**
mcs p-value	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(0.18)**	(0.01)	(0.00)	(0.01)	(0.00)	**(1.00)**
QLIK	15.278	15.293	15.310	15.286	15.320	15.561	16.793	15.231	15.248	15.264	15.234	15.281	**15.206**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(0.32)**	(0.01)	(0.00)	**(0.32)**	(0.00)	**(1.00)**
Panel B: weekly forecasts													
Frob	42.170	42.488	42.197	**41.885**	42.372	47.955	53.859	42.124	42.503	42.536	42.008	42.696	42.770
mcs p-value	**(0.64)**	**(0.19)**	**(0.64)**	**(1.00)**	**(0.43)**	(0.05)	(0.00)	**(0.64)**	**(0.14)**	**(0.14)**	**(0.67)**	**(0.11)**	**(0.64)**
QLIK	39.919	39.933	39.886	39.870	39.876	39.995	41.405	39.877	39.901	39.893	**39.857**	39.891	39.922
mcs p-value	(0.01)	(0.00)	**(0.06)**	**(0.26)**	**(0.21)**	(0.00)	(0.00)	**(0.06)**	(0.01)	(0.03)	**(1.00)**	**(0.05)**	(0.03)
Panel C: biweekly forecasts													
Frob	84.342	85.190	83.281	**82.517**	83.181	91.937	106.858	84.862	85.590	84.761	83.394	84.418	85.784
mcs p-value	**(0.11)**	(0.02)	**(0.42)**	**(1.00)**	**(0.42)**	(0.01)	(0.00)	(0.04)	(0.01)	(0.03)	**(0.42)**	**(0.10)**	**(0.11)**
QLIK	50.676	50.702	50.615	50.590	50.582	50.607	52.059	50.601	50.641	50.623	**50.573**	50.609	50.646
mcs p-value	(0.00)	(0.00)	**(0.11)**	**(0.67)**	**(0.67)**	**(0.67)**	(0.00)	**(0.45)**	(0.01)	(0.03)	**(1.00)**	**(0.18)**	**(0.11)**

Notes: This table shows the mean of the Frobenius norm and QLIK values based on daily, weekly, and biweekly predictions of the (15×15) covariance matrix of 15 stocks (AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM) according to the Conditional Autoregressive model with and without HAR dynamics, assuming a Wishart (W), Riesz (R), Inverse Riesz (iR), matrix-F (F), or a F-Riesz (FR) distribution. In addition, we include the EWMA, DCC-GARCH, and the HAR-DRD (HDRD) models as benchmarks. Parameters are estimated with a moving window of 1000 observations and re-estimated after 250 observations. Panel A shows results of 1-step ahead predictions, and Panels B and C list results of cumulative forecasts for a week (five-step ahead) and two weeks (10 steps ahead). The lowest value of the Frobenius norm and QLIK across the models are marked in bold. In addition, we report the MCS p-values based on a 5% significance level. The p-values of the models within the model confidence set are marked in bold. The out-of-sample period goes from January 2005 until December 2019 and contains 3696 observations.

Table 7. Out-of-sample log-scores

	CA specifications					CA-HAR specifications				
	W	**R**	**iR**	**F**	**FR**	**W**	**R**	**iR**	**F**	**FR**
Panel A: MCD/PFE/PG/WMT/XOM										
$S^{ls}(1)$	0.725	1.192	2.510	2.515	3.348	0.827	1.276	2.693	2.670	**3.455**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
$S^{ls}(5)$	3.481	5.821	12.39	12.43	16.60	3.991	6.240	13.31	13.20	**17.13**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
$S^{ls}(10)$	6.966	11.65	24.82	24.88	33.22	7.984	12.49	26.65	26.43	**34.29**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM										
$S^{ls}(1)$	34.62	38.50	48.68	49.99	54.86	35.24	38.96	49.11	50.74	**55.00**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
$S^{ls}(1)$	171.9	191.3	242.3	249.0	273.2	175.0	193.6	244.4	252.6	**273.9**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
$S^{ls}(10)$	343.7	382.6	484.7	498.0	546.5	350.0	387.2	489.0	505.3	**547.9**
mcs p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**

Notes: This table shows the mean of log scores, defined in (10), based on daily, weekly, and biweekly ahead predictions of the covariance matrix, according to the Conditional Autoregressive model with and without HAR dynamics, assuming a Wishart (W), Riesz (R), Inverse Riesz (iR), matrix-F (F), or a F-Riesz I (FR) distribution. More specifically, we show the average value of $S_t^{ls}(H) = \sum_{h=1}^{H} \log p_{t+h}(\mathbf{X}_{t+h}|\mathbf{V}_{t+h}, \mathcal{F}_{t+h-1})$ with $H = 1$, 5, and 10, respectively. Parameters are estimated with a moving window of 1,000 observations and re-estimated after 250 observations. Panel A shows results of dimension 5, and Panel B lists results of dimension 15. The highest value of the log score across the models is marked in bold. In addition, we report the MCS p-values based on a 5% significance level. The p-values of the models within the model confidence set are marked in bold. The out-of-sample period goes from January 2005 until December 2019 and contains 3696 observations.

distributions, irrespective of the forecast horizon. The gains are large, even with respect to the matrix-F distribution. Second, the HAR specification is most of the time better than the regular conditional autoregressive specification from Equation (1), except for the third data set of dimension 15 (see Supplementary Appendix Table G.5). We conclude that though the differences between the point forecasts of the different models are modest, the differences in density fits are clearly in favor of the F-Riesz model with HAR dynamics for $\boldsymbol{V}_t$.[6]

Finally, we evaluate the different models in terms of their GMV portfolio performance. To save space, we only present Tables 8 and 9 in the main text, reporting the results for 1-day and 10-days (2 weeks) ahead forecasts for the first set of 5 and 15 assets. The 5-day-ahead forecasts for these assets, as well as the 1-, 5-, and 10-day-ahead forecasts for the other sets of assets, can be found in Supplementary Appendix Tables G.6–G.12. We only show results for the HAR specifications as these were most often in the MCSs in Table 7. Each table first reports the annualized mean and ex-post (realized) standard deviation of the portfolio returns, with MCS p-values on the lowest ex-post realized portfolio volatility. Next, we report portfolio statistics such as the average turnover, concentration, and total short position. The second part of each table sheds light on the economic significance of using the CAFr HAR model against all other models using the utility-based framework. Bold positive values of Δ_γ indicate that an investor is willing to sacrifice an annual return of Δ_γ basis points to switch from a particular model to the CAFr HAR model.

Table 8 shows three important results. First, the CAFr HAR model has significantly lower ex-post realized portfolio volatility in the five-dimensional case. For the case of 15 assets, the CAFr model belongs to the MCS and behaves at par with the matrix-F and Inverse Riesz. This is the most important signal in the table, as the GMV criterion function only steers toward minimizing the portfolio variance, not taking into account its return, short positions, concentration, or any other performance measure of the portfolio returns. Second, apart from the EWMA filter, the F-Riesz distribution also has the lowest turnover amongst all models. Third, taking into account transaction costs, there are economic gains in switching from any other matrix distribution and the HAR-DRD model to the CAFr HAR model (FR). These gains are more pronounced for the 15-dimensional case, ranging from 14 to 136 basis points per year. Only the DCC model appears to perform better, mainly due to the higher mean return during the sample period. This performance of the DCC model, however, is not robust: for other sets of stocks, the performance of the DCC is easily found to be much worse among the different models, such as in Supplementary Appendix Tables G.7–G.9. We note again that the GMV criterion only takes the variance of the portfolio returns into account, and not its for instance its (notoriously hard to estimate) expected return. Altering the objective to correct for this might result in other portfolios and possibly different rankings.

Most of the results are robust for the 10-day-ahead (bi-weekly) forecasting horizon in Table 9 (as well as the additional tables and settings in Supplementary Appendix G). In particular, the F-Riesz distribution either has the lowest (realized) ex-post portfolio volatility ($k = 5$) or is in the MCS for models with the lowest variance ($k = 15$). Particularly at this longer horizon, the lower variance typically comes at the cost of a lower average return, resulting in a less favorable performance on the return-based performance criteria. Again, the GMV criterion does not take this dimension explicitly into account. Still, it is comforting to see that the economic gains at the bi-weekly horizon are either insignificant

[6]Note that the other benchmarks like DCC, HAR-DRD, and EWMA are not included in this comparison, because they either (i) do not produce a density forecast for lack of a density assumption, or (ii) take the return *vector* rather than the realized covariance *matrix* as the random variable of interest, thus inhibiting a density comparison with the matrix-valued F-Riesz distribution.

Table 8. The GMV portfolio (daily)

		EWMA	DCC	HDRD	W	R	iR	F	FR
Panel A: MCD/PFE/PG/WMT/XOM									
Mean ret		4.573	5.887	5.760	5.333	5.289	5.172	5.210	5.228
$\hat{\sigma}_p$		10.201	10.593	10.163	10.161	10.159	10.152	10.153	**10.149**
mcs p-value		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(1.00)**
StDev ret		11.699	12.104	11.693	11.644	11.637	11.628	11.634	**11.622**
TO		0.025	0.156	0.240	0.194	0.183	0.156	0.166	0.140
CO		0.500	0.550	0.512	0.510	0.510	0.507	0.508	0.506
SP		−0.005	−0.011	−0.005	−0.007	−0.007	−0.007	−0.007	−0.006
$c = 0\%$	Sharpe	0.391	0.486	0.493	0.458	0.454	0.445	0.448	0.450
	Δ_1	66	−60	−52	−10	−6	6	2	
	Δ_{10}	74	−9	−45	−8	−4	6	3	
$c = 1\%$	Sharpe	0.386	0.454	0.441	0.416	0.415	0.411	0.412	0.419
	Δ_1	37	−56	−27	3	5	**10**	8	
	Δ_{10}	45	−5	−20	6	7	**10**	10	
$c = 2\%$	Sharpe	0.380	0.421	0.389	0.374	0.375	0.377	0.376	0.389
	Δ_1	8	−52	−2	17	16	**14**	**15**	
	Δ_{10}	16	0	5	**19**	18	**14**	**16**	
Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM									
Mean ret		7.054	9.253	8.124	7.705	7.734	7.584	7.616	7.613
$\hat{\sigma}_p$		9.248	10.051	9.274	9.224	9.223	9.212	**9.212**	9.212
mcs p-value		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	**(0.66)**	**(1.00)**	**(0.66)**
StDev ret		10.685	11.344	10.754	10.587	10.582	10.569	10.575	**10.568**
TO		0.055	0.293	0.502	0.273	0.247	0.193	0.214	0.170
CO		0.415	0.504	0.439	0.431	0.430	0.426	0.428	0.425
SP		−0.104	−0.195	−0.127	−0.119	−0.117	−0.112	−0.115	−0.110

(continued)

Table 8. (continued)

Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM									
$c = 0\%$	Sharpe	0.660	0.816	0.755	0.728	0.731	0.718	0.720	0.720
	Δ_1	57	−156	−49	−9	−12	3	0	
	Δ_{10}	68	−79	−31	−7	−11	3	0	
$c = 1\%$	Sharpe	0.647	0.750	0.638	0.663	0.672	0.672	0.669	0.680
	Δ_1	28	−124	35	17	7	9	11	
	Δ_{10}	39	−48	52	19	9	9	12	
$c = 2\%$	Sharpe	0.634	0.685	0.520	0.598	0.613	0.626	0.618	0.639
	Δ_1	−1	−93	**118**	**43**	27	**14**	**22**	
	Δ_{10}	10	−16	**136**	**45**	**28**	**14**	**23**	

Notes: This table shows portfolio statistics of the Global Minimum Variance portfolio, based on one-step-ahead predictions of the covariance matrix, according to the Conditional Autoregressive model with HAR dynamics, assuming a Wishart (W), Riesz (R), Inverse Riesz (iR), matrix-*F* (F), or a *F*-Riesz (FR) distribution. In addition, we include the EWMA, the DCC-GARCH, and HAR-DRD models as benchmarks. We report the average annualized return and standard deviation (both using the true realized covariance matrix and daily returns), as well as the turnover (TO), portfolio concentration (CO), and short positions (SP). Lowest portfolio volatilities are marked in bold. We also list the *p*-value that belongs to the model confidence set approach (MCS) of lowest ex-post daily volatility using the true realized covariance matrix based on a 5% significance level. Bold *p*-values correspond to models that belong to the model confidence set. Finally, the table reports the economic gains of switching from each model listed in the column to the CAFr HAR model in annual basis points, Δ_γ, for various transaction cost levels c and risk aversion coefficients γ. A bold Δ_γ means significantly different from zero at the 5% level. Panel A shows results of dimension 5, and Panel B lists results of dimension 15. The out-of-sample period goes from January 2005 until December 2019 and contains 3696 observations.

Table 9. The GMV portfolio (biweekly)

		EWMA	DCC	HDRD	W	R	iR	F	FR
Panel A: MCD/PFE/PG/WMT/XOM									
Mean ret		4.414	5.961	5.539	5.432	5.385	5.259	5.297	5.319
$\hat{\sigma}_p$		10.583	10.892	10.605	10.568	10.569	10.564	10.564	**10.563**
mcs p-value		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	**(0.40)**	**(0.40)**	**(1.00)**
StDev ret		10.383	10.600	10.307	10.341	10.348	10.327	10.326	**10.317**
TO		0.031	0.110	0.152	0.108	0.106	0.089	0.092	0.083
CO		0.500	0.537	0.506	0.501	0.501	0.500	0.500	0.500
SP		−0.005	−0.009	−0.003	−0.005	−0.005	−0.005	−0.005	−0.005
$c = 0\%$	Sharpe	0.425	0.562	0.537	0.525	0.520	0.509	0.513	0.516
	Δ_1	**182**	−122	−47	−22	−12	**12**	5	
	Δ_{10}	**195**	−68	−49	−17	−7	**14**	6	
$c = 1\%$	Sharpe	0.424	0.560	0.534	0.523	0.518	0.507	0.511	0.514
	Δ_1	**180**	−121	−43	−21	−11	**13**	5	
	Δ_{10}	**192**	−66	−45	−16	−5	**15**	7	
$c = 2\%$	Sharpe	0.424	0.557	0.530	0.520	0.515	0.505	0.509	0.512
	Δ_1	**177**	−120	−40	−20	−10	**13**	5	
	Δ_{10}	**190**	−65	−42	−15	−4	**15**	7	
Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM									
Mean ret		6.755	8.943	7.583	7.167	7.179	7.038	7.061	7.055
$\hat{\sigma}_p$		**9.651**	10.343	9.737	9.669	9.672	9.661	9.659	9.660
mcs p-value		**(1.00)**	(0.00)	(0.00)	(0.00)	(0.00)	**(0.17)**	**(0.52)**	**(0.52)**
StDev ret		10.149	10.681	10.317	10.144	10.137	10.113	10.128	**10.108**
TO		0.063	0.213	0.331	0.166	0.154	0.132	0.145	0.123
CO		0.416	0.492	0.427	0.426	0.425	0.424	0.424	0.424
SP		−0.104	−0.181	−0.110	−0.109	−0.108	−0.106	−0.107	−0.106

(continued)

Table 9. (continued)

Panel B: AA/AXP/BA/CAT/GE/HD/HON/IBM/JPM/KO/MCD/PFE/PG/WMT/XOM									
$c=0\%$	Sharpe	0.666	0.837	0.735	0.707	0.708	0.696	0.697	0.698
	Δ_1	61	−366	−103	−22	−24	3	−1	
	Δ_{10}	68	−254	−63	−15	−19	4	3	
$c=1\%$	Sharpe	0.664	0.832	0.727	0.702	0.704	0.693	0.694	0.695
	Δ_1	58	−361	−93	−20	−23	4	0	
	Δ_{10}	65	−250	−53	−13	−17	5	4	
$c=2\%$	Sharpe	0.663	0.827	0.719	0.698	0.701	0.689	0.690	0.692
	Δ_1	55	−356	−82	−17	−21	4	1	
	Δ_{10}	62	−245	−42	−11	−16	5	5	

Note: This table shows portfolio statistics of the Global Minimum Variance portfolio, based on biweekly predictions of the covariance matrix, according to the Conditional Autoregressive model with HAR dynamics, assuming a Wishart (W), Riesz (R), Inverse Riesz (iR), matrix-*F* (F), or a *F*-Riesz (FR) distribution. In addition, we include the EWMA, the DCC-GARCH, and HAR-DRD models as benchmarks. We report the average annualized return and standard deviation (both using the true realized covariance matrix and daily returns), as well as the turnover (TO), portfolio concentration (CO), and short positions (SP). Lowest portfolio volatilities are marked in bold. We also list the p-value that belongs to the model confidence set approach (MCS) of lowest ex-post daily volatility using the true realized covariance matrix based on a 5% significance level. Bold p-values correspond to models that belong to the model confidence set. Finally, the table reports the economic gains of switching from each model listed in the column to the CAFr HAR model in annual basis points, Δ_γ, for various transaction cost levels c and risk aversion coefficients γ. A bold Δ_γ means significantly different from zero at the 5% level. Panel A shows results of dimension 5, and Panel B lists results of dimension 15. The out-of-sample period goes from January 2005 until December 2019 and contains 3696 observations.

(non-bold) or positive in favor of the *F*-Riesz model. We also note that the *F*-Riesz again has the lowest turnover amongst all models, safe the EWMA filter.

To sum up our out-of-sample results, the *F*-Riesz distribution does very well out-of-sample in terms of density forecasts, indicating that the *F*-Riesz distribution captures the distributional shape of realized variances and covariances better than any of the other models considered here at different forecasting horizons. In terms of point forecasts, the *F*-Riesz model is also in the MCS for different forecasting horizons and different numbers of assets. Finally, in terms of GMV performance, the new model is again always in the MCS for yielding the lowest ex-post portfolio variance. It also results in stable allocations, as the model has the lowest turnover across all models considered, safe the EWMA filter. The lower ex-post variance may come at the cost of a lower ex-post average return. Still, the economic gains are for many of the models either in favor of the *F*-Riesz model or insignificantly different.

4 Conclusions

In this article, we introduced the new conditional autoregressive *F*-Riesz model for capturing the dynamics of matrix-valued random variables. The *F*-Riesz distribution was obtained by mixing the Riesz distribution (Hassairi and Lajmi 2001) with an Inverse Riesz distribution (Tounsi and Zine 2012), thus allowing for much more heterogeneity in tail behavior compared to the well-known matrix distributions like the thin-tailed Wishart, the inverse Wishart, or the fat-tailed matrix-*F* distribution. While the latter distributions depend on one or two degrees of freedom parameters, the new distribution allows vector-valued degrees of freedom parameters. These can easily be estimated by a two-step targeted maximum likelihood approach. In higher dimensions, the elements of the vector-valued degrees-of-freedom parameters might even be clustered to impose further parsimony using the likelihood fit as a guiding mechanism; compare the clustering approach of Oh and Patton (2023) in a copula context.

An empirical application to realized covariance matrices of dimensions 5 and 15 and different samples of U.S. stocks over 19 years of daily data showed a remarkably high increase in the likelihood of the *F*-Riesz distribution compared to the (Inverse) Wishart, (Inverse) Riesz, and matrix-*F* distributions. The margin of outperformance in terms of density forecasts was significant, both in-sample and out-of-sample. Also, the degrees of freedom parameters varied significantly over the different coordinates. The model was also always in the MCS for the lowest ex-post portfolio variance for a GMV analysis, as well as for most of the simple point forecast comparisons. Overall these results show that there is strong heterogeneity of tail behavior of realized covariance matrices, as well as fat-tailedness, and that the *F*-Riesz distribution can be a helpful vehicle to obtain better empirical models.[7]

Funding

Funding support for this article was provided by the Dutch National Science Foundation (NWO) (VI.VIDI.201.079).

References

Amisano, G., and R. Giacomini. 2007. Comparing Density Forecasts via Weighted Likelihood Ratio Tests. *Journal of Business & Economic Statistics* 25: 177–190.

[7]As one of the referees remarked, the *F*-Riesz distribution may also be an interesting alternative in modeling covariance matrices in macroeconomics. For instance, Arias, Rubio-Ramírez, and Shin (2023) found that a Wishart-based "approach may be too restrictive, as a single scalar parameter (ν) controls the overall tightness of the distribution." The *F*-Riesz distribution would be less susceptible to this drawback.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71: 579–625.

Anderson, T. 1962. *An Introduction to Multivariate Statistical Analysis*, Technical report. New York: Wiley.

Andersson, S, and T. Klein. 2010. On Riesz and Wishart Distributions Associated with Decomposable Undirected Graphs. *Journal of Multivariate Analysis* 101: 789–810.

Arias, J. E., J. F. Rubio-Ramírez, and M. Shin. 2023. Macroeconomic Forecasting and Variable Ordering in Multivariate Stochastic Volatility Models. *Journal of Econometrics* 235: 1054–1086.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *Econometrics Journal* 12: 1–32.

Barndorff-Nielsen, O, and N. Shephard. 2004. Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics. *Econometrica* 72: 885–925.

Bollerslev, T., J. Li, A. Patton, and R. Quaedvlieg. 2020. Realized Semicovariances. *Econometrica* 88: 1515–1551.

Bollerslev, T., A. Patton, and R. Quaedvlieg. 2018. Modeling and Forecasting (un)Reliable Realized Covariances for More Reliable Financial Decisions. *Journal of Econometrics* 207: 71–91.

Brownlees, C, and G. Gallo. 2006. Financial Econometric Analysis at Ultra-High Frequency: Data Handling Concerns. *Computational Statistics & Data Analysis* 51: 2232–2245.

Callot, L., A. Kock, and M. Medeiros. 2017. Modeling and Forecasting Large Realized Covariance Matrices and Portfolio Choice. *Journal of Applied Econometrics* 32: 140–158.

Chen, R., D. Yang, and C.-H. Zhang. 2022. Factor Models for High-Dimensional Tensor Time Series. *Journal of the American Statistical Association* 117: 94–116.

Chiriac, R, and V. Voev. 2011. Modelling and Forecasting Multivariate Realized Volatility. *Journal of Applied Econometrics* 26: 922–947.

Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–196.

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

Díaz-García, J. 2016. Riesz and Beta-Riesz Distributions. *Austrian Journal of Statistics* 45: 35–51.

Diebold, F., and R. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13: 253–263.

Engle, R. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–350.

Engle, R., and G. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intra-Daily Data. *Journal of Econometrics* 131: 3–27.

Engle, R., O. Ledoit, and M. Wolf. 2019. Large Dynamic Covariance Matrices. *Journal of Business & Economic Statistics* 37: 363–375.

Fleming, J., C. Kirby, and B. Ostdiek. 2001. The Economic Value of Volatility Timing. *The Journal of Finance* 56: 329–352.

Fleming, J., C. Kirby, and B. Ostdiek. 2003. The Economic Value of Volatility Timing Using "Realized" Volatility. *Journal of Financial Economics* 67: 473–509.

Golosnoy, V., B. Gribisch, and R. Liesenfeld. 2012. The Conditional Autoregressive Wishart Model for Multivariate Stock Market Volatility. *Journal of Econometrics* 167: 211–223.

Gribisch, B, and J. Hartkopf. 2023. Modeling Realized Covariance Measures with Heterogeneous Liquidity: A Generalized Matrix-Variate Wishart State-Space Model. *Journal of Econometrics* 235: 43–64.

Gross, K, and D. Richards. 1987. Special Functions of Matrix Argument. I. Algebraic Induction, Zonal Polynomials, and Hypergeometric Functions. *Transactions of the American Mathematical Society* 301: 781–811.

Hansen, P., A. Lunde, and J. Nason. 2011. The Model Confidence Set. *Econometrica* 79: 453–497.

Hassairi, A., and S. Lajmi. 2001. Riesz Exponential Families on Symmetric Cones. *Journal of Theoretical Probability* 14: 927–948.

Jin, X., and J. Maheu. 2013. Modeling Realized Covariances and Returns. *Journal of Financial Econometrics* 11: 335–369.

Jin, X., and J. Maheu. 2016. Bayesian Semiparametric Modeling of Realized Covariance Matrices. *Journal of Econometrics* 192: 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: 504–518.
Markowitz, H. 1952. Portfolio Selection. *The Journal of Finance* 7: 77–91.
Markowitz, H. 1991. Foundations of Portfolio Theory. *The Journal of Finance* 46: 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: 995–1033.
Noureldin, D., N. Shephard, and K. Sheppard. 2012. Multivariate High-Frequency-Based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 27: 907–933.
Oh, D., and A. Patton. 2016. High-Dimensional Copula-Based Distributions with Mixed Frequency Data. *Journal of Econometrics* 193: 349–366.
Oh, D., and A. Patton. 2017. Modeling Dependence in High Dimensions with Factor Copulas. *Journal of Business & Economic Statistics* 35: 139–154.
Oh, D., and A. Patton. 2018. Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads. *Journal of Business & Economic Statistics* 36: 181–195.
Oh, D., and A. Patton. 2023. Dynamic Factor Copula Models with Estimated Cluster Assignments. *Journal of Econometrics* 237: 105374.
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: 643–657.
Opschoor, A., A. Lucas, I. Barra, and D. van Dijk. 2021. Closed-Form Multi-Factor Copula Models with Observation-Driven Dynamic Factor Loadings. *Journal of Business & Economic Statistics* 39: 1066–1079.
Patton, A. 2009. Copula-Based Models for Financial Time Series. In: Mikosch, T., Kreiß, JP., Davis, R., Andersen, T. (eds), *Handbook of Financial Time Series*. Berlin, Heidelberg: Springer, pp. 767–785.
Politis, D., and J. Romano. 1994. The Stationary Bootstrap. *Journal of the American Statistical Association* 89: 1303–1313.
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: 2449–2495.
Tounsi, M., and R. Zine. 2012. The Inverse Riesz Probability Distribution on Symmetric Matrices. *Journal of Multivariate Analysis* 111: 174–182.
Wang, D., X. Liu, and R. Chen. 2019. Factor Models for Matrix-Valued High-Dimensional Time Series. *Journal of Econometrics* 208: 231–248.
White, H. 2000. A Reality Check for Data Snooping. *Econometrica* 68: 1097–1126.
Wintenberger, O. 2013. Continuous Invertibility and Stable QML Estimation of the EGARCH(1, 1) Model. *Scandinavian Journal of Statistics* 40: 846–867.

Journal of Financial Econometrics, 2024, 00, 1–29
https://doi.org/10.1093/jjfinec/nbae023
Article