TI 2024-049/III
Tinbergen Institute Discussion Paper

# Density Forecasting for Electricity Prices under Tail Heterogeneity with the t-Riesz Distribution

*Anne Opschoor*[1]
*Dewi Peerlings*[2]
*Luca Rossini*[3]
*Andre Lucas*[4]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

2 Vrije Universiteit Amsterdam and Tinbergen Institute

3 University of Milan and Fondazione Eni Enrico Mattei

4 Vrije Universiteit Amsterdam and Tinbergen Institute

# Density Forecasting for Electricity Prices under Tail Heterogeneity with the $t$-Riesz Distribution*

Anne Opschoor[a,b] Dewi Peerlings[a,b] Luca Rossini[c,d] Andre Lucas[a,b]

[a]Vrije Universiteit Amsterdam, The Netherlands [b]Tinbergen Institute, The Netherlands

[c]University of Milan, Italy [d]Fondazione Eni Enrico Mattei, Italy

July 16, 2024

**Abstract**

We introduce the vector-valued $t$-Riesz distribution for time series models of electricity prices. The $t$-Riesz distribution extends the well-known Multivariate Student's $t$ distribution by allowing for tail heterogeneity via a vector of degrees of freedom (DoF) parameters. The closed-form density expression allows for straightforward maximum likelihood estimation. A clustering approach for the DoF parameters is provided to reduce the number of parameters in higher dimensions. We apply the $t$-Riesz distribution to a 24-dimensional panel of Danish daily electricity prices over the period 2017-2024, considering each hour of the day as a separate coordinate. Results show that multivariate $t$-Riesz-based density forecasts improve significantly upon the standard Student's $t$ distribution and the $t$-copula. Further, the $t$-Riesz distribution produces superior implied univariate density forecasts during the afternoon for the distribution as a whole and during 8 a.m.- 8 p.m. in its left tail. Moreover, during crisis periods, this effect is even stronger and holds for almost every hour of the day. Finally, portfolio Value-at-Risk forecasts during the central hours of the day improve during crisis periods compared to the classical Student's $t$ distribution and the $t$-copula.

*Anne Opschoor and Dewi Peerlings thank the Dutch National Science Foundation (NWO) for financial support under grant VI.VIDI.201.079. Luca Rossini acknowledges financial support from the Italian Ministry MIUR under the PRIN project Modelling Non-standard Data and Extremes in Multivariate Environmental Time series (MNEMET) (grant 20223CEZSR).

**Key words:** multivariate distributions, (fat)-tail heterogeneity, (inverse) Riesz distribution, electricity prices

# 1 Introduction

Models for electricity prices have gained considerable attention over the past decade. This is due in part to recent crisis events such as the COVID-19 pandemic and the Russian invasion of Ukraine with its subsequent impact on energy markets worldwide. The volatility of electricity prices, however, is also partly rooted in developments related to the sustainable energy transition many countries are facing. A particularity of electricity as an important tradable good is that it cannot or can only partially be stored. It must typically be consumed when and where it is produced. In contrast to traditional fossil fuel energy plants, renewable energy sources (RES) such as solar or wind energy exhibit much more volatile electricity production patterns due to their dependence on partly unpredictable weather conditions. Such volatility may not only result in stress on the electricity transport infrastructure. It may also result in high volatility for electricity spot prices. High uncertainty in such prices may render further investments in renewable energy sources (RES) too risky or even unprofitable and in this way jeopardize the sustainable energy transition. It is therefore important to get a better grip on the dynamics of electricity price densities for forecasting and risk analysis.

The electricity price literature has thus far mainly focused on univariate models for electricity prices for individual hours of the day; see for instance Koopman et al. (2007), Gianfreda and Grossi (2012), Paraschiv et al. (2014) and Chan and Grant (2016). Maciejowska and Nowotarski (2016) and Ziel (2016) find that hours from the previous day influence the early morning hours of the following days. Only recently, Gianfreda et al. (2020, 2023) have shifted the attention to multivariate time series models for electricity prices, where the 24 hourly prices are analyzed jointly. They provide evidence that electricity prices are fat-tailed, particularly during the central hours of the day.

In this paper, we contribute to this line of literature by developing a model for electricity prices based on the recently proposed $t$-Riesz distribution (Ghorbel and Louati, 2019). We introduce a new dynamic version of the $t$-Riesz distribution to describe the dynamics of time-

varying volatilities in electricity prices and we present efficient ways to estimate the model's parameters in high dimensions using both targeting and clustering. The $t$-Riesz distribution generalizes the multivariate $t$-distribution by mixing a multivariate normal with an inverse Riesz. The inverse Riesz as opposed to an inverse chi-squared distribution is characterized by a vector of degrees of freedom (DoF) parameters, rather than by a single scalar DoF. In this way, it allows for tail heterogeneity. Such a generalization seems useful for empirical data, as found by the preliminary results of Gianfreda et al. (2020, 2023), who found that the typical Student's $t$ distribution with its single DoF parameter may be too restrictive for electricity prices.

Our dynamic extension of the $t$-Riesz distribution is observation-driven and therefore has an explicit expression for the likelihood function. Parameter estimation by maximum likelihood (ML) isstraightforward. However, when using the dynamic $t$-Riesz distribution in a higher dimensional setting, the number of parameters still increases for two reasons and may thus complicate the maximization of the likelihood. First, if the dimension $k$ of the data increases, the intercept term of the covariance matrix transition equation increases the number of parameters in the model quadratically by $k(k+1)/2$. We address this by using a targeting approach to estimate this intercept. Second, the number of DoF parameters increases with the dimension $k$. Here, for more parsimony, we introduce a clustering approach to reduce the number of free DoF parameters. This still allows for flexibility and tail heterogeneity but also reduces the computational burden during optimization.

We apply our new model to Danish Electricity prices over the period of January 2017 – May 2024. This period covers major events like the COVID-19 pandemic and the Russian invasion of Ukraine. In particular, February 2022 onwards was important for electricity prices due to the discussion and subsequent ban on Russian gas trading on European markets (see, Ravazzolo and Rossini, 2023). Our full sample results clearly show that the dynamic conditional $t$-Riesz distribution fits the 24 hour data significantly better than the multivariate Student's $t$ distribution. Out-of-sample, we find that the $t$-Riesz distribution produces superior multivariate density forecasts compared to the $t$ distribution, confirming our in-sample findings. Moreover, the $t$-Riesz distribution also outperforms the $t$-copula. Further, the $t$-Riesz distribution produces superior implied (full-domain) marginal density forecasts during the afternoon and better implied left-tail marginal density forecasts during

8 a.m. - 8 p.m. Focusing on the crisis period related to the Russian invasion of Ukraine, the result becomes even stronger: the $t$-Riesz distribution then outperforms the Student's $t$ distribution during almost all hours of the day. This underlines the importance of allowing for tail heterogeneity during crisis periods. Finally, during crisis periods, the $t$-Riesz distribution is also able to better predict upper risk quantiles of a portfolio of electricity prices during the central hours of the day better than the $t$-distribution and the $t$-copula. In particular, the number of Value-at-Risk violations is closer to the nominal coverage value.

The remainder of this paper is organised as follows. Section 2 introduces the new dynamic $t$-Riesz model. Section 3 discusses the empirical application. Section 4 concludes.

# 2 The model

## 2.1 The $t$-Riesz distribution

The $t$-Riesz distribution for a general vector $\boldsymbol{y} \in \mathbb{R}^{k\times 1}$ generalizes the multivariate Student's $t$ distribution by mixing the multivariate normal with an inverse Riesz distribution rather than with a scalar inverse chi-squared distribution. The inverse Riesz distribution is characterized by a vector-valued DoF parameter $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)^\top \in \mathbb{R}^{k\times 1}$ rather than by a single, scalar DoF parameter $\nu$. It therefore allows for tail heterogeneity in different directions for the vector $\boldsymbol{y}$. We refer to the appendix for more details on the inverse Riesz distribution. The pdf of the $t$-Riesz distribution is given by the following theorem (see Ghorbel and Louati, 2019).[1] For a proof, see the appendix.

**Theorem 1** (**the $t$-Riesz pdf**). *Let $\boldsymbol{y} \in \mathbb{R}^{k\times 1}$ with pdf*

$$p_{\mathcal{TR}}(\boldsymbol{y}; \boldsymbol{\Sigma}, \boldsymbol{\nu}) = \frac{\overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right)}{\overline{\Gamma}\left(\boldsymbol{\nu}/2\right)} \frac{{}_U|\boldsymbol{\Sigma}|_{\boldsymbol{\nu}/2}}{\pi^{k/2}} \cdot {}_U|\boldsymbol{\Sigma} + \boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\boldsymbol{\nu}+1)/2}\,, \tag{1}$$

*with the generalized multivariate Gamma function $\overline{\Gamma}(\boldsymbol{\nu}) = \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left(\nu_i + \frac{1-i}{2}\right)$ for $2\nu_i > i-1$ and $i = 1, \ldots, k$, and ${}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}} = \prod_{i=1}^{k} \boldsymbol{U}_{i,i}^{2\nu_i}$ the Upper Power Weighted Determinant (UPWD) for a positive definite matrix $\boldsymbol{Y}$ with (unique)* upper *triangular*

[1] We concentrate in this paper on the so-called type-I $t$-Riesz distribution, based on a type-I inverse Riesz distribution. There is also a type-II $t$-Riesz distribution, but in our empirical application, we found the differences between these two distributions to be only minor.

*Cholesky decomposition* $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{U}^\top$*, where* $\boldsymbol{U}$ *has positive diagonal elements.*
*Then* $\boldsymbol{y}$ *as characterized by the pdf in Eq. (1) has a t-Riesz distribution and the distribution of* $\boldsymbol{y}$ *can be obtained by mixing a multivariate normal distribution with an independent inverse Riesz distribution.*

Interestingly, except for the use of the non-standard generalized gamma functions rather than the regular gamma function, and of the power-weighted determinant rather than the regular determinant, the density expression of the $t$-Riesz distribution in Eq. (1) one-on-one mirrors and generalizes the pdf expression of the multivariate Student's $t$ distribution.

We note that the power weighted determinant is *not* a regular determinant and lacks well-known properties such as $|\boldsymbol{A} \cdot \boldsymbol{B}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$ for matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{k \times k}$. We refer to Lemma 1 in the appendix for more details. If, however, all elements of $\boldsymbol{\nu}$ are identical, i.e., $\boldsymbol{\nu} = (\nu, \ldots, \nu)^\top$, then $\overline{\Gamma}(\boldsymbol{\nu})$ collapses to the standard multivariate gamma function, and the Upper Power Weighted Determinant collapses to a simple power of the standard determinant ${}_U|\boldsymbol{Y}|_{\boldsymbol{\nu}} = |\boldsymbol{Y}|^\nu$. In that case, the $t$-Riesz distribution collapses to the multivariate Student's $t$ distribution. The following corollary establishes the precise link between the two.

**Corollary 1.** *Under the conditions of Theorem 1, if we assume* $\boldsymbol{\nu} = \nu \cdot \boldsymbol{\iota}_k$ *for a scalar* $\nu > k - 1$*, then* $\boldsymbol{y}$ *has a multivariate Student's t distribution with* $\nu - k + 1$ *degrees of freedom, i.e.,* $p_{\mathcal{TR}}(\boldsymbol{y}; \boldsymbol{\Sigma}, \nu \cdot \boldsymbol{\iota}_k) = p_T(\boldsymbol{y}; \boldsymbol{\Sigma}, \nu - k + 1)$.

The corollary makes clear that it is possible to test whether the $t$-Riesz distribution collapses to the multivariate Student's $t$ distribution by testing whether all elements in $\boldsymbol{\nu}$ are the same. It also makes clear that the $t$-Riesz can be quite fat-tailed even if some of the $\nu_i$s are quite large. For instance, if all $\nu_i$s are equal to 26 for $k = 24$ dimensions, then the corresponding multivariate Student's $t$ distribution only has $\nu - k + 1 = 3$ degrees of freedom. Empirically, we often find that for our $k = 24$ at least some of the entries $\nu_i$ are substantially smaller than 26, increasing the overall fat-tailedness of the distribution in some directions.

Figure 1 shows simulated draws with covariance matrix $I_2$ from a bivariate $t$ distribution with 7 degrees of freedom as well as draws from a $t$-Riesz distribution with $\boldsymbol{\nu} = (5, 10)^\top$. The scatter plot in the top panel of Figure 1 clearly shows the differences in the tails: in the corners, the $t$-Riesz distribution has more probability mass compared to the $t$ distribution. Also, the conditional densities differ substantially between the two distributions. The

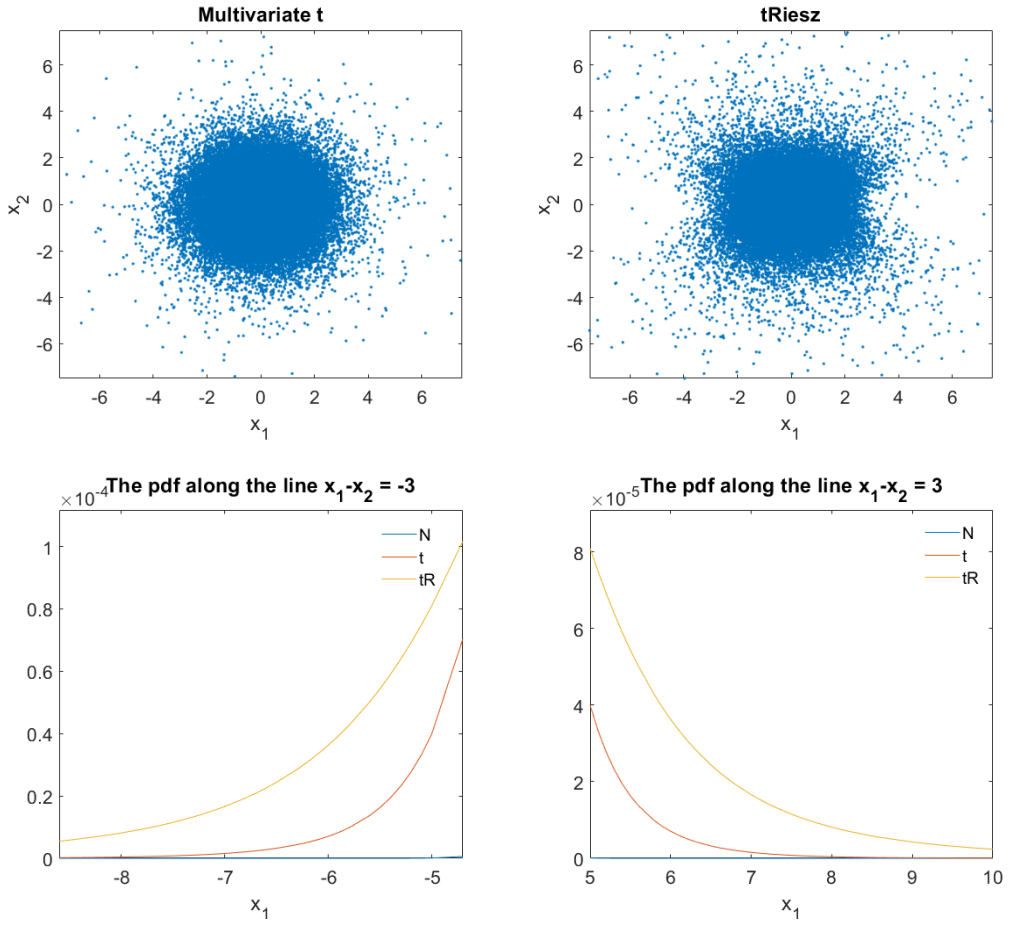


**Figure 1: The $t$-Riesz vs the multivariate Student's $t$ distribution**
This figure shows simulated draws and conditional distributions of two bivariate distributions: the $t$ distribution with 7 degrees of freedom and the $t$-Riesz distribution with $\boldsymbol{\nu} = (5, 10)^\top$ and covariance matrix $I_2$. The top panel shows a scatterplot of $x_2$ against $x_1$, while bottom panel show the pdf of $x_1$ given that $x_1 - x_2 = \pm 3$.

conditional density of $(X_1 \mid X_2 = 0)$ is fatter for the $t$ than the $t$-Riesz distribution, while the opposite holds for the distribution of $x_2$ given $x_1 = 0$. The plots in the bottom panels show a "joint identical crash or boom probability", i.e. the probability along the line $x_1 - x_2 = \pm 3$ according to the Gaussian, $t$ and $t$-Riesz distribution. The figure clearly shows that the latter has a much higher joint crash (and boom) probability than the $t$, and even more so compared to the Gaussian distribution. To sum up, the new distribution allows for much more flexibility in the tails than the standard multivariate Student's $t$ distribution.

## 2.2 Scale matrix dynamics, parameter estimation, and targeting

When applying the $t$-Riesz distribution in the multivariate time series context for modeling electricity prices, it is important to allow for a time-varying scaling matrix $\boldsymbol{\Sigma}$, i.e., for $\boldsymbol{\Sigma}_t$. We do this as follows. First, we define $\boldsymbol{V}_t = \text{Var}[\boldsymbol{y}_t \mid \mathcal{F}_{t-1}] = \boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{U}_{\boldsymbol{V}_t}^\top$ as the time-varying conditional covariance matrix of $\boldsymbol{y}_t$, where $\boldsymbol{M}(\boldsymbol{\nu})$ is the static diagonal covariance matrix of a standard ($\boldsymbol{\Sigma}_t = \mathbf{I}_k$) $t$-Riesz distribution, and $\boldsymbol{U}_{\boldsymbol{V}_t}$ is the unique upper triangular Choleski decomposition of $\boldsymbol{V}_t$.[2] Using the uniqueness of the Choleski decomposition, the diagonality of $\boldsymbol{M}(\boldsymbol{\nu})$, and the relation $\boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{U}_{\boldsymbol{V}_t}^\top = \boldsymbol{U}_{\boldsymbol{\Sigma}_t}\,\boldsymbol{M}(\boldsymbol{\nu})\,\boldsymbol{U}_{\boldsymbol{\Sigma}_t}^\top$, we obtain $\boldsymbol{U}_{\boldsymbol{\Sigma}_t} = \boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{M}(\boldsymbol{\nu})^{-1/2}$, and thus $\boldsymbol{\Sigma}_t = \boldsymbol{U}_{\boldsymbol{\Sigma}_t}\boldsymbol{U}_{\boldsymbol{\Sigma}_t}^\top = \boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{M}(\boldsymbol{\nu})^{-1}\boldsymbol{U}_{\boldsymbol{V}_t}^\top$. Using this expression for the time-varying scale matrix $\boldsymbol{\Sigma}_t$ in the $t$-Riesz distribution, we obtain the following time-varying parameter specification of the $t$-Riesz:

$$\boldsymbol{y}_t \sim \mathcal{TR}(\boldsymbol{\mu}_t\ ,\ \boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{M}(\boldsymbol{\nu})\boldsymbol{U}_{\boldsymbol{V}_t}^\top\ ,\ \boldsymbol{\nu}), \tag{2}$$

$$\boldsymbol{V}_{t+1} = (1 - A - B)\boldsymbol{\Omega} + A\,(\boldsymbol{y}_t - \boldsymbol{\mu}_t)(\boldsymbol{y}_t - \boldsymbol{\mu}_t)^\top + B\,\boldsymbol{V}_t, \tag{3}$$

with scalar parameters $A$ and $B$, $k \times 1$ DoF parameter $\boldsymbol{\nu}$, $k \times k$ matrix-valued intercept parameter $\boldsymbol{\Omega}$, and time-varying conditional mean $\boldsymbol{\mu}_t$ (to be specified later). Unlike a standard BEKK-MGARCH specification, the conditional covariance matrix $\boldsymbol{V}_t$ of $\boldsymbol{y}_t$ does not enter the conditional density in (2) directly, but only via its upper triangular Choleski decomposition $\boldsymbol{U}_{\boldsymbol{V}_t}$ interacted with the diagonal matrix $\boldsymbol{M}(\boldsymbol{\nu})$.

We can now easily compute the log-likelihood function of the dynamic $t$-Riesz model. Using a set of static parameters $A$, $B$, $\boldsymbol{\nu}$, $\boldsymbol{\Omega}$, and a starting value $\boldsymbol{V}_1$, we first obtain all the values of $\boldsymbol{V}_t$ for $t = 1, \ldots, T$, using the filter (3). Next, we compute the log-likelihood function using the $t$-Riesz pdf in (1) with $\boldsymbol{\Sigma}_t = \boldsymbol{U}_{\boldsymbol{V}_t}\boldsymbol{M}(\boldsymbol{\nu})\boldsymbol{U}_{\boldsymbol{V}_t}^\top$, i.e.,

$$\sum_{t=1}^{T} \log \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right) - \log \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) + \log {}_U|\boldsymbol{\Sigma}_t|_{\boldsymbol{\nu}/2} + \log {}_U|\boldsymbol{\Sigma}_t + \boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\boldsymbol{\nu}+1)/2}\,. \tag{4}$$

This log-likelihood function can subsequently be maximized with respect to the static parameters to obtain the maximum likelihood estimator (MLE).

[2] Note that the upper triangular decomposition $\boldsymbol{U}_{\boldsymbol{V}_t}$ can be obtained from the lower triangular Choleski decomposition of $\boldsymbol{V}_t$ as $\boldsymbol{U}_{\boldsymbol{V}_t} = (\boldsymbol{L}_{\boldsymbol{V}_t^{-1}}^{-1})^\top$.

In higher dimensions $k$, estimating the intercept $\boldsymbol{\Omega}$ becomes increasingly cumbersome. We therefore propose a targeting approach to estimate $\boldsymbol{\Omega}$. Assuming stationarity and the existence of an unconditional first moment of $\boldsymbol{V}_t$ (and thus a second unconditional moment of $\boldsymbol{y}_t$), we can take unconditional expectations of the left-hand and right-hand sides of (3) to obtain $\overline{\boldsymbol{V}} = \mathbb{E}[\boldsymbol{V}_t] = \boldsymbol{\Omega}$. Based on this result, we estimate $\boldsymbol{\Omega}$ by its sample equivalent $\hat{\boldsymbol{\Omega}} = T^{-1} \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{\mu}_t)(\boldsymbol{y}_t - \boldsymbol{\mu}_t)^\top$. Using $\hat{\boldsymbol{\Omega}}$, we can then estimate the remaining parameters by standard maximum likelihood.

An elaborate set of simulation results is provided in Appendix B. We first show the finite sample properties of the static $t$-Riesz distribution (with targeting). Second, we show that we can test the null hypothesis of a multivariate Student's $t$ distribution versus the alternative of the $t$-Riesz distribution and that this test has power in finite samples. Third, we show that our model in (**??**)–(3) performs well in finite samples: it estimates the correct parameters back correctly if the model is correctly specified.

## 2.3 Clustering of DoFs and ordering of variables

If the dimension $k$ of the data increases, the number of DoF parameters in $\boldsymbol{\nu}$ also increases. Estimating all of these DoF parameters separately may be empirically challenging. To introduce more parsimony in the model, we propose a clustering approach. Here, we group the elements of $\boldsymbol{\nu}$ into different clusters that each have the same value for the DoF parameter. A challenge for the clustering approach is that not only the value of the DoF parameter of the cluster matters but also the order of the coordinates in the random vector $\boldsymbol{y}_t$. The latter is due to the use of the inverse Riesz distribution in the construction of the $t$-Riesz distribution and is well-known and accepted in the Riesz literature; see also the Bartlett construction of the inverse Riesz in Eq. (A.1) in the appendix. We solve this challenge by combining the ordering and cluster assignment steps into one. Algorithm 1 summarizes the procedure.

In short, we make $G$ groups of consecutive coordinates, each of size $k_g - k_{g-1}$ for $g = 1, \ldots, G$, with $1 \leq k_1 < \ldots < k_G = k$ denoting the end-coordinates of each cluster. Initially, we randomly order the elements of $\boldsymbol{y}_t$ and estimate the static parameters of the model. We then consecutively take each coordinate $i$ in $\boldsymbol{y}_t$ and put it in each of the positions

**Algorithm 1** Clustering DoFs and Ordering the Coordinates

1: Fix the number of clusters $G$. Randomize the order of the elements of $\boldsymbol{y}_t$ and initialize $0 = k_0 < k_1 < \ldots < k_G = k$ to define the cluster end points $k_g$ for $g = 1, \ldots, G$.
2: Set $\nu_i = \tilde{\nu}_g$ for $i \in [k_{g-1} + 1, k_g]$, and maximize the log-likelihood with respect to $A$, $B$, and $\tilde{\boldsymbol{\nu}} = (\tilde{\nu}_1, \ldots, \tilde{\nu}_G)$ while putting $\hat{\Omega}$ at its targeting estimate from Section 2.2.
3: For $i = 1 : k$ do
   a: Shift coordinate $i$ from $\boldsymbol{y}_t$ to position $j$ for each $j = 1, \ldots, k$;
   b: Compute the value of the log-likelihood function at the current estimated parameters, using an equivariant transformation of the targeted $\hat{\Omega}$;
   c: if $j$ crosses a cluster boundary point, i.e., $j = k_g + 1$, then compute the log-likelihood value both for $\nu_j = \tilde{\nu}_g$ and $\nu_j = \tilde{\nu}_{g+1}$, and proceed with the new $\tilde{\nu}_{g+1}$ for the next positions $j$;
   d: retain the coordinate position $j^\star$ that results in the highest likelihood and put coordinate $i$ in the new position $j^\star$ with corresponding cluster assignment;
   e: if $j^* \neq i$, then re-estimate all static parameters of the model;
   f: move to the next coordinate $i$ from the original order of coordinates.
4: Possibly re-iterate steps 2-3 until cluster assignments stabilize; possibly also re-iterate steps 1-4 such for different random starting points and retain the best solution.

$j = 1, \ldots, k$, switching the value of its DoF parameter if it changes cluster membership. For each position and value of the DoF parameter, we calculated the value of the likelihood. Finally, we update the coordinate of $i$ and its corresponding DoF parameter to the position and cluster where it has the highest likelihood.

A final choice in Algorithm 1 is the choice of the number of clusters $G$. Different data-driven approaches can be used for this based on for instance within-cluster similarity and between-cluster dissimilarity (see Hastie et al., 2009, Chapter 14). Here, we apply Algorithm 1 for different values of $G$ and choose the value that minimizes the BIC; compare the approach of Oh and Patton (2023) for clustering in a copula context. Other model selection or cluster selection criteria could also be used.

# 3 Application: density forecasts of electricity prices

## 3.1 Data

In this paper, we focus on one-day-ahead electricity prices, where bids to buy and offers to sell are submitted voluntarily for each hour of the following day. More specifically, this market is opened several days in advance and closed at noon of the day preceding physical

delivery. The actual spot prices realized the next day may depart from these one-day-ahead prices. The electricity prices themselves are affected by factors such as the installed capacity, the international dynamics of fossil fuel prices, plant maintenance, outages, interconnections with foreign markets, weather conditions, and market microstructure noise (see Gianfreda et al., 2023, for a more complete description of electricity markets).

Our empirical analysis focuses on Denmark, and in particular the Danish zonal price area DK1, which is the Western Denmark area (see Pircalabu et al., 2017; Pircalabu and Benth, 2017). We focus on DK1 since it is part of the European continental electricity system and relies on higher Central European prices, while DK2 (Eastern Denmark) is part of the Nordic electricity system and thus refers to the cheaper northern prices. The dataset spans the period from January 1, 2017, to May 31, 2024, and comprises the COVID-19 pandemic and the Russian invasion of Ukraine with the subsequent price cap implementation.

Daily one-day-ahead auction prices for each hour of the next day are collected from Refinitiv-Eikon and are quoted in Euros per MegaWatt-hour (MWh). To account for daylight-saving changes, we remove the 25th hour in October and interpolate the missing 24th hour in March. Operators trading in these markets use forecasts at day $t$ of demand and of wind and solar activity to construct a set of 24 hourly price forecasts for delivery at day $t+1$. These prices are submitted before the closure of the market at around noon of day $t$.

Figure 2 provides an overview of the historical data for electricity prices. The top left panel shows the electricity prices for each hour over the day over the entire sample period. We can clearly see the strong increase in electricity prices after the Russian invasion of Ukraine when Danish electricity prices were strongly affected by the sanctions on Russian gas. The top right panel provides the sample distributions of electricity prices, with high variability across all the hours and with some negative spikes in the central hours. In particular, we can see some outliers varying between 800 Euros/MWh during the initial Russian invasion of Ukraine and $-440$ Euros/MWh during the second wave of the current war. Lastly, the bottom panel depicts the intraday dynamics across days of the week for the prices. The different markers provide evidence of different behaviors across the 24 hours and the different days of the week, with low values during the weekends and high values during the first two days of the week.

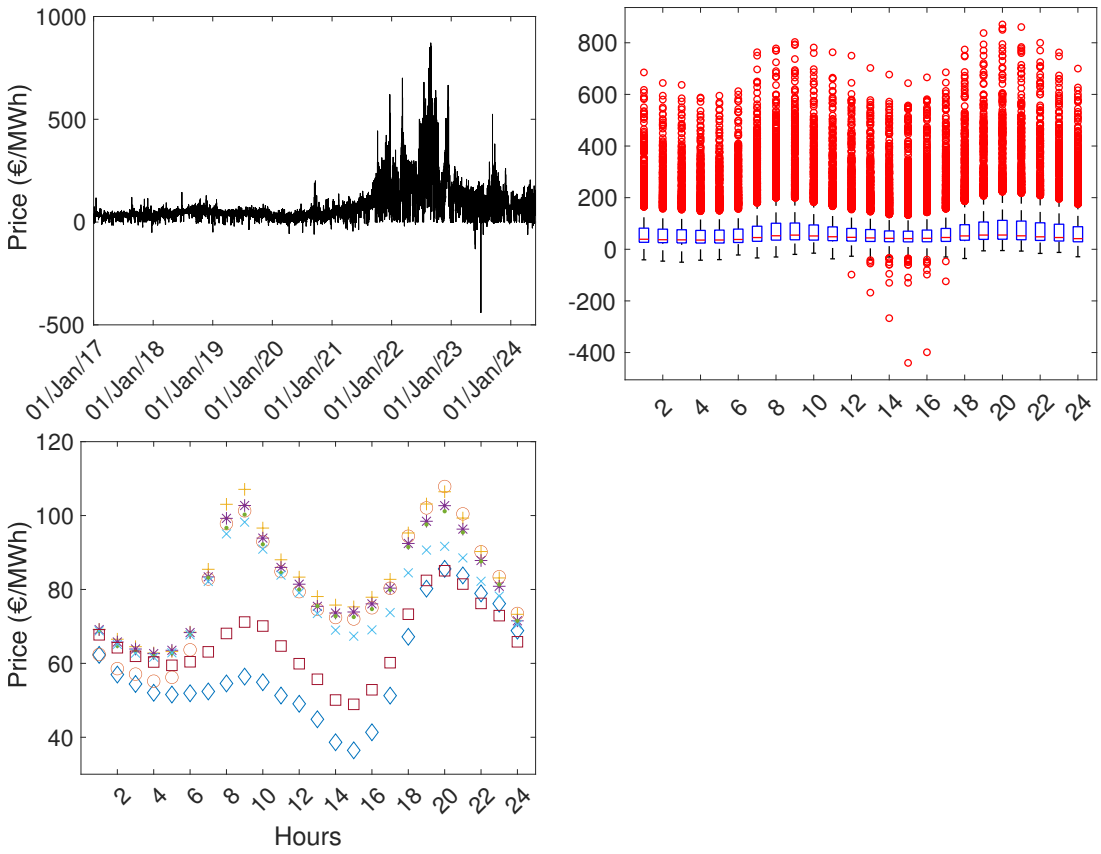


**Figure 2: Hourly Electricity Prices for Denmark.**
Hourly Danish (DK1) electricity one-day-ahead prices (top left), sample price distributions (top right), and Intraday profiles across days of the week (bottom left) from January 1, 2017 to May 31, 2024 ($T = 2,708$ days). [Monday ($\circ$), Tuesday ($+$), Wednesday ($\star$), Thursday ($\bullet$), Friday ($\times$), Saturday ($\square$), Sunday ($\diamond$)]

The electricity prices in the top-left panel of Figure 2 clearly also show outliers and even negative prices from time to time. This may call for a different distribution than the normal for density forecasting. This suspicion is confirmed in Figure 3, where we provide the kurtosis of the one-day-ahead electricity prices for each separate hour of the day jointly with the overall kurtosis of all pooled hours (in red). We see clear evidence of a strong kurtosis across all hours of the day, ranging from a low of 13 to a high of 17. The troughs appear during the first peak hours (8–10 a.m.) and early evening (6–8 p.m.), but even these still provide evidence of heavy tails.

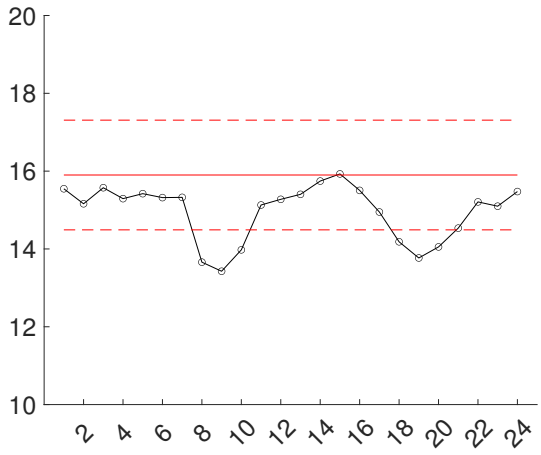


**Figure 3: Kurtosis of Hourly Electricity Prices**
This figure shows Kurtosis across hours (in black) and the overall kurtosis (in red) with 95% confidence bands (in dashed red) for the Danish (DK1) electricity one-day-ahead prices. The sample covers the period from January 1, 2017, to May 31, 2024 ($T = 2,708$ days).

## 3.2 Model set-up and evaluation criteria

Let $\boldsymbol{y}_t = (y_{1,t}, y_{2,t}, \ldots, y_{24,t}) \in \mathbb{R}^{24\times 1}$ be a vector of hourly one-day-ahead electricity prices at day $t$, such that $k = 24$. Following Gianfreda et al. (2023), we model the conditional mean of $y_{h,t}$ for hour $h$ at day $t$ as

$$y_{h,t} = \sum_{k=1}^{K} \psi_k d_{k,t} + \sum_{l=1}^{24} \phi_{1,l} y_{l,t-1} + \phi_2 y_{h,t-2} + \phi_7 y_{h,t-7} + \epsilon_{h,t}, \tag{5}$$

where we included twelve month dummies $(d_{1,t}, \ldots, d_{12,t})$, two weekend dummies (for Saturdays, $d_{13,t}$, and Sundays, $d_{14,t}$), and the lag of all 24 prices of the previous day ($\boldsymbol{y}_{t-1}$), and the 2nd and 7th lag of the hour-specific price.

We gather all hourly error terms $\epsilon_{h,t}$ into a vector $\boldsymbol{\epsilon}_t$, which we model by a $t$-Riesz distribution with time-varying covariance matrix $\boldsymbol{V}_t$ according to the model specified in Eqs. (2)–(3). We call this the BEKK $t$-Riesz model. Note that if $\boldsymbol{\nu} = \nu \boldsymbol{\iota}_k$, the model becomes a BEKK $t$ model with $\nu - k + 1$ degrees of freedom for $\nu > k - 1$. Parameter estimation is carried out in two steps. First, we estimate the parameters of the conditional mean equation by OLS. Next, we estimate the parameters of the BEKK $t$-Riesz model by Maximum Likelihood and targeting as described in Section 2.2. Hence, the BEKK $t$-Riesz likelihood only needs to be optimized with respect to the parameters $A$, $B$, and $\boldsymbol{\nu}$. Standard

errors are obtained using the standard sandwich covariance matrix estimator.

We investigate the one-step-ahead density forecast performance of the new dynamic $t$-Riesz distribution vis-a-vis the BEKK Student's $t$ model. We consider both the multivariate density forecasts for the full vector $\boldsymbol{y}_t$, as well as the implied marginal density forecasts of the electricity prices $y_{h,t}$ at each hour as obtained from the estimated multivariate density. To assess the differences in performance, we use the log scoring rule (see Mitchell and Hall, 2005; Amisano and Giacomini, 2007). Define the difference in log score between two density forecasts $M_1$ and $M_2$ corresponding to vector $\boldsymbol{y}_t$ as

$$d_{ls,t} = S_{ls,t}(\boldsymbol{y}_t, M_1) - S_{ls,t}(\boldsymbol{y}_t, M_2), \tag{6}$$

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

$$S_{ls,t}(\boldsymbol{y}_t, M_j) = \log p_{\boldsymbol{y}_t}(\boldsymbol{y}_t | \boldsymbol{V}_t, \mathcal{F}_{t-1}, M_j), \tag{7}$$

where $p_t(\cdot)$ is the pdf of the $t$-Riesz or the Student's $t$ BEKK model. The null hypothesis of equal predictive ability is given by $H_0 : \mathbb{E}[d_{ls}] = 0$ for all $P = T - R$ out-of-sample forecasts. This null can be tested through a Diebold and Mariano (1995) (DM) statistic

$$DM_{ls} = \frac{\bar{d}}{\sqrt{\hat{\sigma}^2/N}}, \tag{8}$$

with out-of-sample average $\bar{d}$ of the log score differences and HAC variance estimator $\hat{\sigma}^2$. Under the assumptions of the framework of Giacomini and White (2006), $DM_{ls}$ asymptotically follows a standard normal distribution. A significantly positive value means that model $M_1$ has a superior forecast performance over model $M_2$. We also consider the Model Confidence Set (MCS) of Hansen et al. (2011). The MCS automatically accounts for the dependence between model outcomes given that all models are based on the same data.

We follow Gianfreda et al. (2023) and also compute one-step-ahead marginal density forecasts of electricity prices for each hour of the day as implied by the multivariate density for $\boldsymbol{y}_t$, i.e., the pdf of $y_{h,t}$ for every $h = 1, \ldots, 24$. We evaluate these marginal density forecasts by using the quantile CRPS score (Gneiting and Ranjan, 2011), which is defined

as

$$S_{h,t}(y_{h,t+1}) = \int_0^1 QS_\alpha \left(F^{-1}(\alpha), y_{h,t+1}\right) \omega(\alpha) d\alpha, \tag{9}$$

with the cumulative distribution function (cdf) $F$ of the pdf $p_{\boldsymbol{y}_t}$, and the quantile score $QS_\alpha \left(F^{-1}(\alpha), y_{h,t+1}\right)$ given by

$$QS_\alpha \left(F^{-1}(\alpha), y_{h,t+1}\right) = 2 \left(I_{y_{h,t+1} \leq F^{-1}(\alpha)} - \alpha\right) \left(F^{-1}(\alpha) - y_{h,t+1}\right), \tag{10}$$

with the quantile forecast $F^{-1}(\alpha)$ and $0 \leq \alpha \leq 1$. Further, $\omega(\alpha)$ is a weight function, which is set equal to 1 (no weight), $\alpha^2$ (weight on the right tail), and $(1-\alpha^2)$ (weight on the left tail), respectively. Similar to the log scoring rule, we test for differences in the (weighted) Q-CRPS score between two distributions using the DM test. The lower the CRPS score, the better the one-step-ahead density forecasts are.

Finally, we investigate the implications of the different forecasting distributions from a risk management perspective. More specifically, we are interested in the probability that the (average) electricity price during the central hours of the day exceeds a certain number. This portfolio Value-at-Risk is formally defined as

$$Pr(y^p_{t+1} \geq pVaR^{1-q}_{t+1} | \mathcal{F}_{t-1}; \boldsymbol{\mu}_{t+1}, \boldsymbol{V}_{t+1}) = q, \tag{11}$$

with $y^p_{t+1} = \boldsymbol{y}^\top_{t+1} \boldsymbol{w}$ an equal weighted average of the electricity prices of tomorrow and $q$ a particular quantile. We choose the 8 central hours during the day (9 a.m. - 4 p.m.) for our analysis and set $q$ equal to 0.10, 0.05, 0.025, and 0.01 respectively. We evaluate our forecasted 1-step Value-at-Risk values by the classical Conditional Coverage test of Christoffersen (1998), and its two subtests: the Unconditional Coverage test and the test on the independence of the violations.

## 3.3 Full sample results

Table 1 presents the main results. We first discuss the results for the full sample of $T = 2,708$ observations. As the focus is on the density forecasts and the effect of the new tail heterogeneity model, we concentrate on the volatility and tail part of the results.

**Table 1: Full sample parameter estimates, likelihoods and information criteria**
This table reports maximum likelihood parameter estimates of the BEKK model of (2) - (3), assuming a $t$-Riesz or a multivariate Students's $t$ distribution, applied to a panel of 24 hourly Danish electricity prices. The Table lists the coefficients, as well as the standard errors, obtained by the sandwich covariance estimator. We report the log-likelihood $\mathcal{L}$, the BIC, and the $\chi^2$ statistic (plus critical value using a 1 % significance level) of the test on an identical DoF parameter across coordinates, $H_0 : \boldsymbol{\nu} = \nu\,\boldsymbol{\iota}_k$. The sample spans the period from January 1, 2017, until May 31, 2024 ($T = 2,708$ days).

| $t$-Riesz | | | | | | | | | | multivariate $t$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BEKK parameters | | | | | | | | | | | |
| $A$ | 0.005 | | | | | | | | | $A$ | 0.007 |
| | (0.001) | | | | | | | | | | (0.001) |
| $B$ | 0.993 | | | | | | | | | $B$ | 0.991 |
| | (0.001) | | | | | | | | | | (0.001) |
| Degrees of Freedom parameters | | | | | | | | | | | |
| $\nu_1$ | 4.205 | $\nu_6$ | 10.883 | $\nu_{11}$ | 18.745 | $\nu_{16}$ | 20.922 | $\nu_{21}$ | 25.872 | $\nu$ | 2.717 |
| | (0.393) | | (0.458) | | (1.893) | | (0.323) | | (0.339) | | (0.034) |
| $\nu_2$ | 5.500 | $\nu_7$ | 13.256 | $\nu_{12}$ | 19.685 | $\nu_{17}$ | 22.337 | $\nu_{22}$ | 25.948 | | |
| | (0.381) | | (1.307) | | (0.911) | | (0.409) | | (0.281) | | |
| $\nu_3$ | 6.605 | $\nu_8$ | 14.917 | $\nu_{13}$ | 22.553 | $\nu_{18}$ | 23.551 | $\nu_{23}$ | 26.891 | | |
| | (0.387) | | (1.093) | | (2.271) | | (0.358) | | (0.269) | | |
| $\nu_4$ | 7.432 | $\nu_9$ | 16.286 | $\nu_{14}$ | 20.503 | $\nu_{19}$ | 24.981 | $\nu_{24}$ | 25.617 | | |
| | (0.155) | | (1.149) | | (0.546) | | (0.461) | | (0.031) | | |
| $\nu_5$ | 8.872 | $\nu_{10}$ | 17.056 | $\nu_{15}$ | 20.597 | $\nu_{20}$ | 25.934 | | | | |
| | (0.198) | | (1.522) | | (0.465) | | (0.398) | | | | |
| $\mathcal{L}$ | -197,915 | | | | | | | | | | -200,039 |
| BIC | 396,036 | | | | | | | | | | 400,101 |
| LR test | 4,247 | | | | | | | | | | |
| $\chi^2_{0.99}(23)$ | 41.64 | | | | | | | | | | |

For a more thorough discussion on the conditional mean part, see for instance Gianfreda et al. (2020, 2023).

Looking at the log-likelihood values of the $t$-Riesz and multivariate $t$ BEKK specification in Table 1, the $t$-Riesz model is clearly superior with a log-likelihood improvement of about 2,100 points compared to the multivariate $t$. The corresponding Likelihood Ratio test statistic for $H_0 : \boldsymbol{\nu} = \nu\,\boldsymbol{\iota}_k$ is significant at the 1% level, hence rejecting the multivariate $t$ distribution. This is also confirmed by the BIC values: the BIC of the BEKK $t$-Riesz model is about 4,050 points lower than that of the BEKK $t$ model. Tail heterogeneity thus seems empirically important for electricity prices.

When looking at the estimated DoF parameters, we see they range from a low 4.205 to a high 25.934 for the $t$-Riesz, whereas the $\nu$ for the multivariate $t$ equals 2.717. Note that according to Corollary 1, a $t$-Riesz with DoF parameter $\boldsymbol{\nu} = (2.717 + 24 - 1)\boldsymbol{\iota}_k = 25.717\,\boldsymbol{\iota}_k$ would result in a multivariate $t$ distribution with 2.722 DoF. We see, however, in Table 1,

that many of the estimated DoF parameters $\nu_i$ for the $t$-Riesz distribution are substantially below 25.717. This implies that the $t$-Riesz is much fatter-tailed in certain directions than the multivariate Student's $t$ distribution.

We confirm this pattern by looking at the differences between the two distributions via the QQ plots in Figure 4. Using the two estimated models from Table 1, we simulate 100,000 draws from the multivariate $t$ and the $t$-Riesz distribution, respectively. We then plot the QQ plot for four different hours: 2 a.m., 9 a.m., 1 p.m., and 9 p.m.. If the $t$-Riesz and multivariate $t$ distribution were similar, the blue marks for the $t$-Riesz should be very close to the red dashed line. The figure clearly indicates that there is a substantial difference between the two distributions at 2 a.m., 9 a.m., while for 1 p.m. and 9 p.m., the marginal distributions are roughly equal. The differences do not only apply to the right tail (high electricity prices) but also to the left tail. Note that the estimated DoF parameters are already corrected for time-varying volatility via Eq. (3). The latter is important, as Table 1 also confirms the result of Gianfreda et al. (2023) that there is substantial time-varying volatility for electricity prices.

To sum up, hourly electricity prices do not only exhibit fat tails, but there is also substantial tail heterogeneity, even after correcting for time-varying volatility. The $t$-Riesz distribution statistically outperforms the multivariate $t$ distribution and seems a better description of the distribution of electricity prices. In particular, the multivariate $t$ distribution seems to underestimate the fat-tailedness and thus the risk of electricity prices for some hours of the day. In the next subsection, we investigate whether these results continue to hold out-of-sample and what the implications are for predicted portfolio Value-at-Risks.

## 3.4 Out-of-sample results

To assess the out-of-sample behavior of the new dynamic $t$-Riesz model, we perform two different analyses. In the first analysis, we compare density forecasts of the new $t$-Riesz distribution to those of the benchmark multivariate $t$ density using the joint vector $\boldsymbol{y}_t$ of all 24 hours of the day. As a second benchmark, we also consider a copula approach as a natural competitor of the $t$-Riesz distribution. Copulas are used in many applications in economics

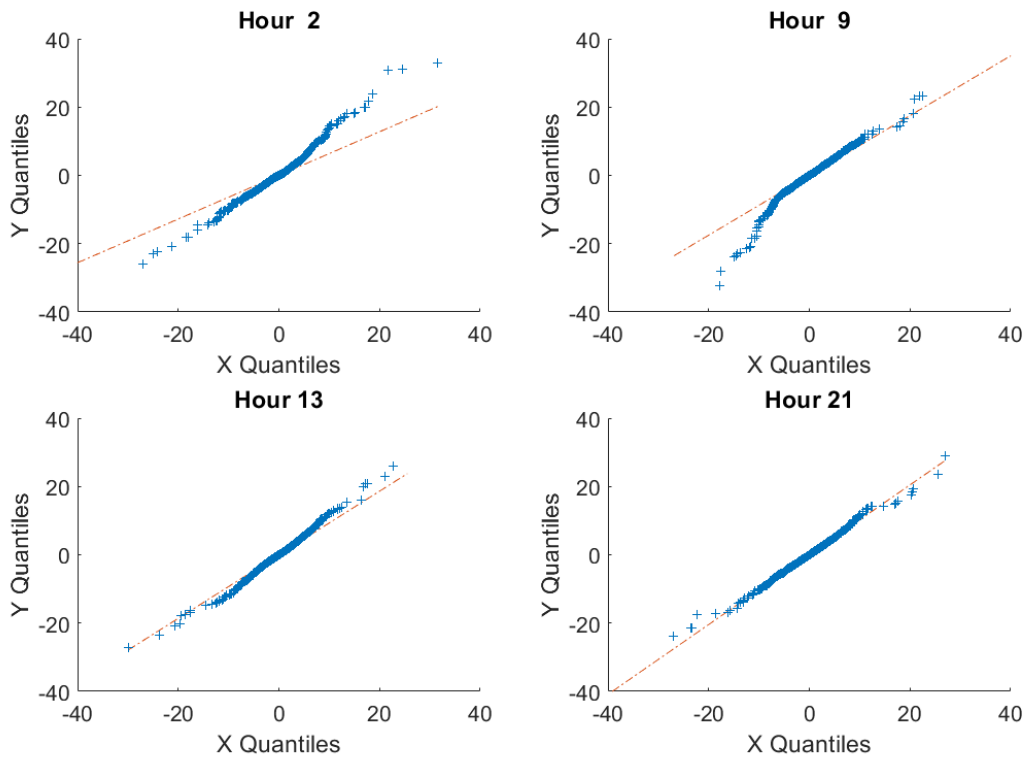


**Figure 4: QQ plots**
This figure shows QQ plots for the implied marginal distributions of electricity prices at 2 a.m., 9 a.m., 1 p.m., and 9 p.m. based on 100,000 simulated draws from the $t$ and $t$-Riesz distribution with covariance matrix $\mathbf{I}$ and full sample based DoF parameters of Table 1. The red dashed line corresponds to the $t$ distribution, while the blue scatters correspond to the $t$-Riesz distribution.

and finance (see e.g. Patton, 2009; Cherubini et al., 2011; McNeil et al., 2015). We follow Christoffersen et al. (2014) and consider the $t$-copula for the dependence structure with DCC type of correlations and model each marginal separately by a univariate Student's $t$ distribution with a mean equation as in (5), GARCH volatilities and $\nu_h$ degrees of freedom.

In addition, we compare the BEKK $t$ and BEKK $t$-Riesz performances in terms of the implied out-of-sample marginal density forecasts for each separate hour of the day. In our second analysis, we compare the out-of-sample risk implications of all models by computing the portfolio Value-at-Risk during 8a.m.-8p.m. from the one-step-ahead density forecasts. All results are based on a rolling window of 800 observations, where we re-estimate the static parameters of each model roughly every 2 months (66 days).

### 3.4.1 Density Forecasts

For the density forecasts, we adopt the following procedure. At each point in time, we evaluate the one-step-ahead density forecasts of the whole vector $\boldsymbol{y}_t$ using the log-scoring rule. To assess the quality of the one-step-ahead marginal density forecasts, we use the Q-CRPS score as given in (9). We approximate the required integral for the Q-CRPS by a numerical one over a grid of 1000 points. Note that the marginal density is unknown for the $t$-Riesz distribution. We approximate it by simulating 250,000 vectors $\boldsymbol{y}_t$ at each time $t$ to compute the daily CRPS for the $t$-Riesz for each hour separately.

Table 2 and Figure 5 summarize the results. We see in panel A of Table 2 that the *multivariate* out-of-sample density forecasts of the $t$-Riesz distribution are superior to those of the multivariate $t$. The DM test of 14.67 is clearly significant in favor of the $t$-Riesz. Even the difference in log score vis-a-vis the $t$-copula is positive and significant. This is partly surprising since the $t$-copula allows for more flexibility by modeling each marginal density separately with a time-varying mean, volatility, and DoF parameter. Its dependence structure, by contrast, is still very restrictive with one tail shape parameter rather than $k$ parameters as in the $t$-Riesz distribution. Also, the results for the model confidence sets point in the same direction. The $t$-Riesz is always in the model confidence set, whereas the multivariate $t$ and the $t$-copula never are. This implies that the heterogeneity in tail behavior for the full 24-hour multivariate distribution of electricity prices is not only important in-sample but also out-of-sample. Such tail heterogeneity cannot be captured by the multivariate Student's $t$ distribution.

Second, in panel B we see that the average daytime (8 a.m. – 8 p.m.) *marginal* density forecasts as implied by the multivariate $t$-Riesz distribution for $\boldsymbol{y}_t$ again outperform those of the Student's $t$ distribution, both for the full support and for the left tail. The DM statistics become even more negative when we focus on the second half of the out-of-sample period containing the crisis (i.e. the Russian invasion of Ukraine). Hence also during such times of turmoil, the $t$-Riesz distribution with its more flexible, heterogeneous tail behavior proves valuable compared to a plain-vanilla Student's $t$ distribution.

Third, Figure 5 shows the results for the marginal density forecasts per hour in more detail. The figure consists of 6 different sub-plots for all combinations of full versus crisis

**Table 2: Out-of-Sample one-step-ahead density forecasts results**

This table shows results of one-step-ahead density forecasts of the full and marginal densities of $\boldsymbol{y}_t$ (the vector with 24 electricity prices at day $t$) according to the BEKK model assuming a multivariate Student's $t$ or a $t$-Riesz I distribution and a $t$-copula with univariate GARCH-t marginals. We use a moving window of 800 observations and re-estimate our parameters after 66 days. Panel A shows the mean of the log-score as defined in (7) according to four different approaches: The BEKK $t$-Riesz full (estimating all 24 DoFs), BEKK $t$-Riesz cluster (BEKK Tr with a clustering approach using 3 clusters), the BEKK $t$ model and the $t$-copula model with GARCH-$t$ marginals. The panel also lists the pair-wise DM statistics on significant differences in the log-score against the BEKK Tr full model. Positive statistics indicate that the BEKK tRiesz full model has superior density forecasts. Finally, we list the MCS $p$-values based on a 5% significance level. The highest value of the predictive log-score across the models and the p-values of the models within the model confidence set are marked in bold. Panel B shows results on the average marginal density forecast during the hours 8 a.m. - 8 p.m., using the average Q-CRPS score of (9) $\omega(\alpha) = 1, \alpha^2$ and $(1-\alpha)^2$, respectively. The panel shows DM statistic of equal density forecasts between the multivariate Student's $t$ and the $t$-Riesz distribution is significant at a 5% level. Negative values indicate superior density forecasts of the $t$-Riesz distribution. The panel discriminates between the whole out-of-sample and the second half containing the crisis period (December 2021 - May 2024). The out-of-sample period goes from March 19, 2019, until May 31, 2024 ($T = 1,901$ days).

| Panel A: multivariate density forecasts | | | | |
|---|---|---|---|---|
| | $t$-Riesz (full) | $t$-Riesz cluster | mult $t$ | $t$-copula |
| $S_{ls}$ | **-79.581** | -79.635 | -80.254 | -80.120 |
| mcs p-value | **(1.00)** | 0.024 | (0.00) | (0.00) |
| $DM_{ls}$ | | 3.187 | 14.670 | 5.861 |

| Panel B: Average univariate density forecasts (8:00 - 20:00) | | |
|---|---|---|
| $DM_{Q-CRPS}$ stat ($t$ vs $t$-Riesz ) | | |
| | Full OOS | Crisis period |
| Full dist | -2.188 | -5.366 |
| Left tail | -5.024 | -7.238 |
| Right tail | 1.613 | -0.514 |

period, and the full support versus the left or right tail. Red bars indicate that the $t$-Riesz implied marginal distributions perform better than their Student's $t$ counterparts based on the Q-CRPS score and a DM statistic at a 5% significance level. For blue bars, the converse holds, while no bars are shown if the scores are not significantly different at a 5% level.

The first message from the figure is that the $t$-Riesz distribution with tail heterogeneity improves upon the $t$ distribution when considering the full support of the distribution and its left tail. This result holds regardless of whether we focus on the full out-of-sample (left panels) or on the second half of the out-of-sample period that includes the crisis period. The gain is statistically significant during 12a.m.-5p.m. and 8p.m. for the full distribution and holds during 7 a.m. - 8 p.m. when we focus solely on the left tail of the distribution. Only in the case of the right tail, the $t$ distribution sometimes does a better job during the full out-of-sample period, while the opposite holds during night time in crisis periods.

Second, comparing the left and right panels, the tail heterogeneous $t$-Riesz distribution is even more preferred during times of turmoil than when evaluated over the full out-of-sample period, irrespective of the focus of the distribution (full support, left tail, or right tail). Especially for the full support and the left tail, the $t$-Riesz distribution outperforms the $t$ distribution for almost every hour of the day.

To sum up, one-step ahead *multivariate* density forecasts of the BEKK $t$-Riesz model are superior to those of the $t$-distribution and the $t$-copula. In addition, the *univariate* marginal distributions implied by their multivariate counterparts show that the $t$-Riesz distribution outperforms the $t$-distribution for many hours of the day, both for the full marginal distribution and its left-tail. During crisis periods the result is even stronger and holds for almost all hours of the day. Again, this illustrates the empirical usefulness of the $t$-Riesz distribution in allowing for tail heterogeneity in electricity prices over the course of the day.

Finally, Table 2 shows the performance of the clustering-based dynamic $t$-Riesz model using the approach described in Section 2.3. This reduces the number of DoF parameters of the $t$-Riesz distribution and checks whether the $t$-Riesz is not overfitting the data with its $k$ different DoF parameters in the vector $\boldsymbol{\nu}$. To estimate the number of clusters and the order of the variables for the $t$-Riesz, we iterate the clustering algorithm twice over the in-sample period of the first 800 observations, resulting in 3 clusters with cluster-adjusted DoF values $\nu_i = \tilde{\nu}_g + i + 1$ for $\tilde{\nu}_1 = 2.29$, $\tilde{\nu}_2 = 5.57$, and $\tilde{\nu}_3 = 3.33$, respectively; see Algorithm 1 in Section 2.3. Table 2 shows that the multivariate density forecasts of this more parsimonious dynamic $t$-Riesz distribution also outperform those of the multivariate Student's $t$ and the $t$-copula. A pairwise Diebold-Mariano test of the clustered dynamic $t$-Riesz distribution against the $t$ (-copula) distributions results in t-statistics of 12.545 and 5.158 respectively. Hence the differences are statistically significant. The clustering approach performs worse than the full $t$-Riesz model, but the difference is relatively small compared to improvements vis-á-vis the other competitors. Hence, clustering the DoF parameters in our current application and reducing the number of parameters can be implemented without a substantial loss in density forecast performance.

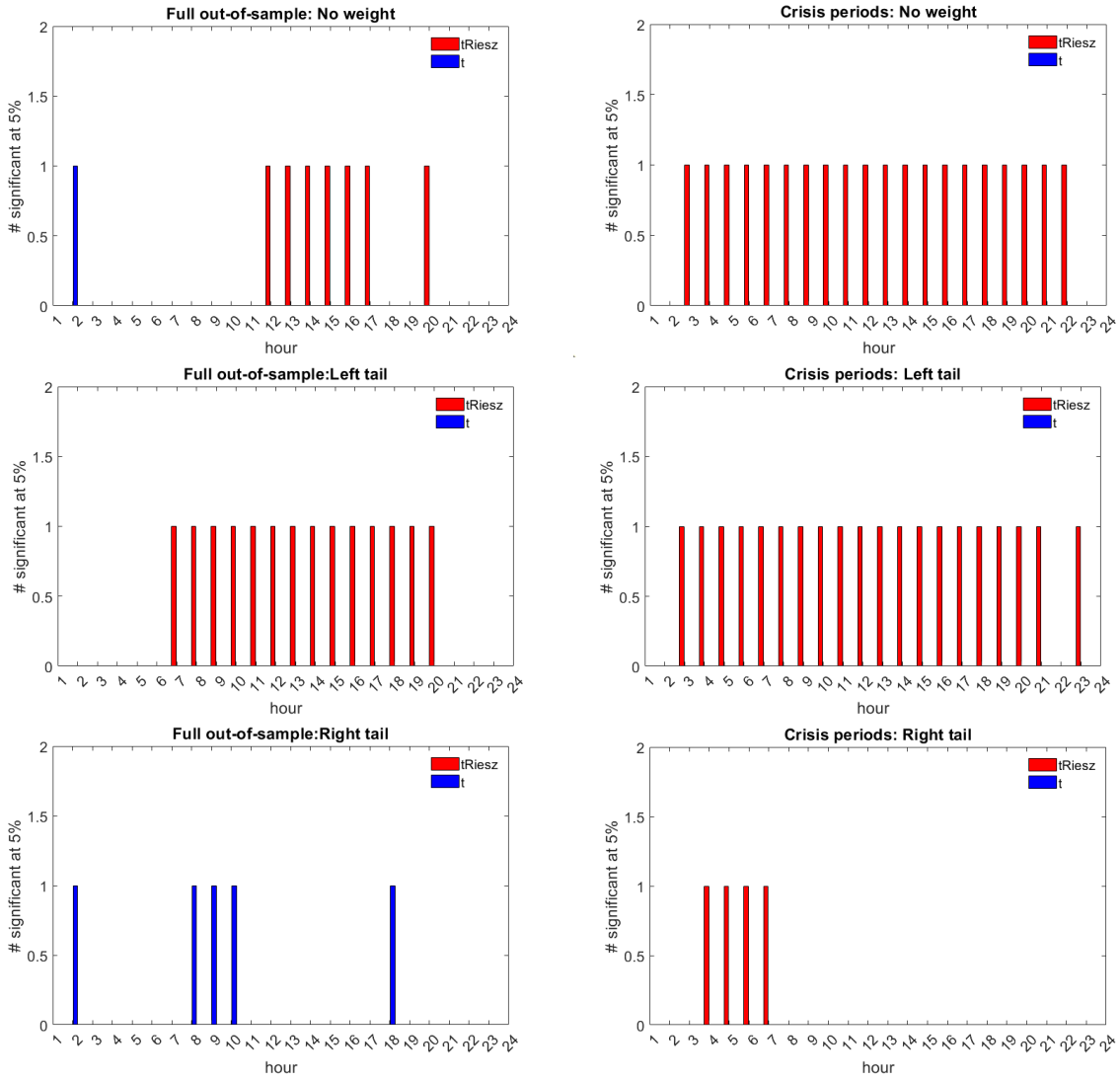


**Figure 5: Testing univariate density forecasts using the Q-CRPS score**
This figure shows results on the DM test of one-step-ahead density forecasts of hourly electricity prices between the BEKK $t$-Riesz and BEKK $t$ model using the Q-CRPS score of (9) based on three different weighting schemes. The top panel shows results when $\omega(\alpha) = 1$ (no weight), the middle panel depicts the case when $\omega(\alpha) = \alpha^2$ (right tail), and the bottom panel corresponds to the case $\omega(\alpha) = (1-\alpha)^2$ (left tail). The left panel shows results based on the full out-of-sample, while the right panel focuses on the crisis periods (December 2021 - May 2024). Each sub-figure shows for each hour (1 a.m. - 12 a.m.) whether the difference in the Q-CRPS score is statistically significant at a 5% level. The out-of-sample period goes from March 19, 2019, until May 31, 2024 ($T = 1,901$ days).

### 3.4.2 Forecasting Risk Measures

We end the empirical section by evaluating the 1-step ahead portfolio Value-at-Risk forecasts of the average electricity price during 9a.m.-4p.m. Table 3 lists the results of these VaR predictions during the second half of the out-of-sample, which includes the crisis

period. We list the number and percentage of violations, as well as the $p$-values for the (un)conditional coverage and independence test. A bold number indicates that the null of correct (un)conditional coverage or independence cannot be rejected.

The table shows that the portfolio VaR forecasts of the $t$-Riesz distribution are unconditionally correct and closest to their nominal values. In contrast to this, the $t$-copula with its flexible margins is too conservative and overpredicts the portfolio volatility as all violation rates are substantially below their nominal value. For the 90% and 95% VaR the difference is even so large that the null hypothesis of correct unconditional coverage is rejected for the $t$ copula model. Finally, $t$-distribution and $t$-Riesz distribution behave similarly from a VaR perspective, with a slight edge for the $t$-Riesz model as the independence test rejects rejects for the $t$ distribution at the 97.5% VaR. Also, the unconditional coverage percentages for the $t$-Riesz distribution are (slightly) closer to their nominal values than for the $t$ distribution.

To sum up, during crisis periods, the $t$-Riesz distribution outperforms the $t$ distribution and the $t$-copula with respect to the number of violations and the independence of violations.

# 4 Conclusions

In this paper, we introduced the conditional dynamic $t$-Riesz distribution into time series econometrics with a focus on electricity prices. The $t$-Riesz distribution was obtained by mixing the inverse Riesz distribution (Tounsi and Zine, 2012) with the multivariate normal distribution, allowing for heterogeneity in tail behavior compared to the well-known multivariate Student's $t$ distribution. While the multivariate $t$ only depends on a single, scalar degrees of freedom parameter, the $t$-Riesz is characterized by a vector of degrees of freedom parameters. The model could easily be estimated by straightforward maximum likelihood methods. The different degrees of freedom parameters could be grouped into clusters using the clustering algorithm provided in this paper, thus introducing more parsimony into the model.

We applied the new dynamic conditional distribution to a panel of 24 Danish daily electricity prices during the period 2017-2024, including the start of the Russian invasion of Ukraine. The full sample analysis revealed that the $t$-Riesz distribution fits electricity prices

**Table 3: Out-of-Sample portfolio Value-at-Risk forecasts during Crises periods**
This table shows results of one-step-ahead portfolio Value-at-Risk forecasts of the average electricity price between 9 a.m. and 4 p.m. according to the BEKK model assuming a multivariate Student's $t$ or a $t$-Riesz I distribution and a $t$-copula with univariate GARCH-t marginals. We use a moving window of 800 observations and re-estimate our parameters after 66 days. The Table shows for four different quantiles (90, 95, 97.5, and 99) the number (percentage) of violations, and p-values associated with the (Un)conditional Coverage and Independence tests of Christoffersen (1998). Bold numbers show that the null hypothesis can not be rejected (UC/IND/CC tests) The table shows results on the second half of the out-of-sample (December 2021 - May 2024), including the Crisis period and contains 901 days.

| | Risk quantile (1-q) | | | |
|---|---|---|---|---|
| | 0.9 | 0.95 | 0.975 | 0.99 |
| BEKK $t$-Riesz | | | | |
| # violations | 94 | 47 | 22 | 9 |
| perc violations | 10.43% | 5.22% | 2.44% | 1.00% |
| p-value UC | **0.667** | **0.767** | **0.910** | **0.997** |
| p-value IND | 0.000 | 0.000 | **0.111** | **0.074** |
| p-value CC | 0.000 | 0.000 | **0.280** | **0.203** |
| BEKK $t$ | | | | |
| # violations | 98 | 54 | 24 | 9 |
| perc violations | 10.88% | 5.99% | 2.66% | 1.00% |
| p-value UC | **0.386** | **0.184** | **0.755** | **0.997** |
| p-value IND | 0.000 | 0.000 | 0.025 | **0.074** |
| p-value CC | 0.000 | 0.000 | **0.076** | **0.203** |
| $t$-Copula | | | | |
| # violations | 69 | 32 | 20 | 6 |
| perc violations | 7.66% | 3.55% | 2.22% | 0.67% |
| p-value UC | 0.015 | 0.036 | **0.583** | **0.283** |
| p-value IND | 0.005 | 0.027 | **0.461** | **0.777** |
| p-value CC | 0.001 | 0.010 | **0.656** | **0.540** |

significantly better than the multivariate $t$-distribution. The degrees of freedom parameters of the $t$-Riesz varied significantly across the different coordinates. In addition, multivariate density forecasts improved significantly when using the $t$-Riesz distribution compared to the $t$ distribution and the $t$-copula.

Especially during the second half of the out-of-sample period (December 2021 - May 2024), the $t$-Riesz distribution also produces superior implied univariate density forecasts during most hours of the day, as well as better density forecasts in the left tail. Finally, the $t$-Riesz also produced better risk quantiles than the $t$ distribution and the $t$-copula since the start of the Russian invasion of Ukraine. Overall, the results support the empirical relevance of tail heterogeneity for electricity prices, and the proposed dynamic $t$-Riesz distribution can

be a useful device to capture this feature of the data in an empirically congruent way.

# References


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

Anderson, T. W. (1962). An Introduction to Multivariate Statistical Analysis. Technical report, Wiley New York.

Chan, J. C. C. and A. L. Grant (2016). Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics 54*, 182–189.

Cherubini, U., S. Mulinacci, F. Gobbi, and S. Romagnoli (2011). *Dynamic Copula Methods in Finance*. John Wiley and Sons.

Christoffersen, P., V. Errunza, K. Jacobs, and X. Jin (2014). Correlation dynamics and international diversification benefits. *International Journal of Forecasting 30*(3), 807–824.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*(4), 841–862.

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

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

Ghorbel, E. and M. Louati (2019). The multiparameter $t$ distribution. *Filomat 33*(13), 4137–4150.

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

Gianfreda, A. and L. Grossi (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics 34*(6), 2228–2239.

Gianfreda, A., F. Ravazzolo, and L. Rossini (2020). Comparing the forecasting performance of linear models for electricity prices with high RES penetration. *International Journal of Forecasting 36*(3), 974–986.

Gianfreda, A., F. Ravazzolo, and L. Rossini (2023). Large Time-Varying Volatility Models for Hourly Electricity Prices. *Oxford Bulletin of Economics and Statistics 85*(3), 545–573.

Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics 29*(3), 411–422.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.

Koopman, S. J., M. Ooms, and M. A. Carnero (2007). Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association 102*(477), 16–27.

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

Maciejowska, K. and J. Nowotarski (2016). A hybrid model for GEFCom2014 probabilistic electricity price forecasting. *International Journal of Forecasting 32*(3), 1051 – 1056.

McNeil, A., R. Frey, and P. Embrechts (2015). *Quantitative risk management: Concepts, techniques and tools*. Princeton University Press.

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

Oh, D. H. and A. J. Patton (2023). Dynamic factor copula models with estimated cluster assignments. *Journal of Econometrics 237*, 105374.

Opschoor, A., A. Lucas, and L. Rossini (2023). Tail heterogeneity for dynamic covariance matrices: the FRiesz distribution. *Tinbergen Institute Discussion Paper 2021-010/III*.

Paraschiv, F., D. Erni, and R. Pietsch (2014). The impact of renewable energies on EEX day-ahead electricity prices. *Energy Policy 73*, 196 – 210.

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

Pircalabu, A. and F. E. Benth (2017). A regime-switching copula approach to modeling day-ahead prices in coupled electricity markets. *Energy Economics 68*, 283–302.

Pircalabu, A., T. Hvolby, J. Jung, and E. Høg (2017). Joint price and volumetric risk in wind power trading: A copula approach. *Energy Economics 62*, 139–154.

Ravazzolo, F. and L. Rossini (2023). Is the price cap for gas useful? Evidence from European countries. *FEEM Working Papers 023.2023*.

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

Ziel, F. (2016). Forecasting electricity spot prices using Lasso: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems 31*(6), 4977–4987.

# A Lemmas and Proofs

**Lemma 1** (**see Appendix B of Opschoor et al., 2023**). *Given a scalar $\nu$, a vector $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^\top \in \mathbb{R}^{k\times 1}$, a vector of ones $\boldsymbol{\iota}_k \in \mathbb{R}^{k\times 1}$, and $\boldsymbol{Y} \in \mathbb{R}^{k\times k}$ a positive definite matrix, then we have the following identities.*

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

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

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

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

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

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 (DoF) parameters $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)^\top$, with $\nu_i > i-1$ for $i = 1, \dots, k$. Let $\boldsymbol{G} \in \mathbb{R}^{k\times k}$ be the random matrix

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

where all elements of $\boldsymbol{G}$ are independent random variables. Then $\boldsymbol{Y} = \boldsymbol{L}\boldsymbol{G}\boldsymbol{G}^\top\boldsymbol{L}^\top$ has a type-I Riesz distribution, $\boldsymbol{Y} \sim \mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$, where type-I relates to the fact that we have taken a lower triangular Cholesky decomposition in the Bartlett decomposition. A type-II Riesz distribution uses the upper-triangular decomposition instead but is not used in the current paper. If $\nu_i \equiv \nu$ for $i = 1, \dots, k$, (A.1) collapses to the well-known Bartlett decomposition of a standard Wishart distribution (Anderson, 1962) and the Riesz collapses to the Wishart distribution. In contrast to the Wishart distribution, the Riesz distribution thus allows for heterogeneous tail behavior in the cross-section.

Tounsi and Zine (2012) introduce tail fatness into the Riesz distribution by developing the inverse Riesz distribution. If $\boldsymbol{Y} \sim \mathcal{R}^I(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\nu})$, then $\boldsymbol{X} = \boldsymbol{Y}^{-1}$ has a (type-I) inverse Riesz distribution, which we denote as $\boldsymbol{X} \sim i\mathcal{R}^I(\boldsymbol{\Sigma}, \boldsymbol{\nu})$. Let $\boldsymbol{\Sigma}^{-1} = \boldsymbol{L}_{\boldsymbol{\Sigma}^{-1}}\boldsymbol{L}_{\boldsymbol{\Sigma}^{-1}}^\top$, then $\mathbb{E}[\boldsymbol{X}] = (\boldsymbol{L}_{\boldsymbol{\Sigma}^{-1}}^\top)^{-1}\boldsymbol{M}(\boldsymbol{\nu})\boldsymbol{L}_{\boldsymbol{\Sigma}^{-1}}^{-1}$ with $\boldsymbol{M}(\boldsymbol{\nu})$ the first moment of a type-I inverse Riesz distribution with $\boldsymbol{\Sigma} = \mathbf{I}$; see Louati and Masmoudi (2015) and Theorem 2 below.

**Proof of Theorem 1:** The pdf of the inverse Riesz is given by

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

We therefore obtain the joint density of $\boldsymbol{y}$ and $\boldsymbol{X}$ as

$$\begin{aligned}
p(\boldsymbol{y},\boldsymbol{X}) &= \frac{\exp\left(-\frac{1}{2}\boldsymbol{y}^\top\boldsymbol{X}^{-1}\boldsymbol{y}\right)}{|2\pi\boldsymbol{X}|^{1/2}} \cdot \frac{|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)} \cdot \operatorname{etr}\left(-\frac{1}{2}\boldsymbol{\Sigma}\boldsymbol{X}^{-1}\right)}{|\boldsymbol{\Sigma}^{-1}|_{0.5\boldsymbol{\nu}} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \\
&= \frac{|\boldsymbol{X}^{-1}|_{\boldsymbol{\iota}_k/2}\ |\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+k+1)}\ \operatorname{etr}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)\ \boldsymbol{X}^{-1}\right)}{(2\pi)^{k/2}\ \ |\boldsymbol{\Sigma}^{-1}|_{\boldsymbol{\nu}/2} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \\
&= \frac{|\boldsymbol{X}^{-1}|_{0.5(\boldsymbol{\nu}+1+k+1)}\ \operatorname{etr}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)\ \boldsymbol{X}^{-1}\right)}{|\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)^{-1}|_{(\boldsymbol{\nu}+1)/2} \cdot \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right) \cdot 2^{(\boldsymbol{\nu}+1)^\top \boldsymbol{\iota}_k/2}} \quad \times \\
&\qquad \frac{|\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)^{-1}|_{(\boldsymbol{\nu}+1)/2} \cdot \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right) \cdot 2^{(\boldsymbol{\nu}+1)^\top \boldsymbol{\iota}_k/2}}{(2\pi)^{k/2}\ \ |\boldsymbol{\Sigma}^{-1}|_{\boldsymbol{\nu}/2} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}}.
\end{aligned}$$

The first factor can be recognized as the pdf of the inverse Riesz of type-I. Therefore, integrating out $\boldsymbol{X}$, we only retain the last factor, which we can rewrite as

$$\begin{aligned}
p(\boldsymbol{y}) &= \int p(\boldsymbol{y},\boldsymbol{X})\,\mathrm{d}\boldsymbol{X} \\
&= \frac{|\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)^{-1}|_{(\boldsymbol{\nu}+1)/2} \cdot \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right) \cdot 2^{(\boldsymbol{\nu}+1)^\top \boldsymbol{\iota}_k/2}}{(2\pi)^{k/2}\ \ |\boldsymbol{\Sigma}^{-1}|_{\boldsymbol{\nu}/2} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right) \cdot 2^{\boldsymbol{\nu}^\top \boldsymbol{\iota}_k/2}} \\
&= \frac{|\left(\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top\right)^{-1}|_{(\boldsymbol{\nu}+1)/2} \cdot \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right)}{(\pi)^{k/2}\ \ |\boldsymbol{\Sigma}^{-1}|_{\boldsymbol{\nu}/2} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right)} \\
&\overset{\text{From Lemma (1iv)}}{=} \frac{{}_U|\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\boldsymbol{\nu}+1)/2} \cdot \overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right)}{{}_U|\boldsymbol{\Sigma}|_{-\boldsymbol{\nu}/2} \cdot \overline{\Gamma}\left(\boldsymbol{\nu}/2\right)(\pi)^{k/2}} \\
&\overset{\text{From Lemma (1iii)}}{=} \frac{\overline{\Gamma}\left((\boldsymbol{\nu}+1)/2\right)\ {}_U|\boldsymbol{\Sigma}|_{\boldsymbol{\nu}/2}}{\overline{\Gamma}\left(\boldsymbol{\nu}/2\right)(\pi)^{k/2}} \cdot {}_U|\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\boldsymbol{\nu}+1)/2}.
\end{aligned}$$

□

**Proof of Corollary 1:** We use the following matrix calculus result for a matrix $\boldsymbol{\Sigma}$: For a $k \times 1$ column vector $\boldsymbol{y}$ and row vector $\boldsymbol{y}^\top$, it holds that:

$$|\boldsymbol{\Sigma}+\boldsymbol{y}\,\boldsymbol{y}^\top| = |\boldsymbol{\Sigma}||1+\boldsymbol{y}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}|.$$

The pdf of the $t$-Riesz distribution with $\boldsymbol{\nu} = \nu\boldsymbol{\iota}_k$ is given by

$$p_{\mathcal{TR}}(\boldsymbol{y};\boldsymbol{\Sigma}^*,\nu\cdot\boldsymbol{\iota}_k) = \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k+1)/2\right)\ {}_U|\boldsymbol{\Sigma}^*|_{\nu\boldsymbol{\iota}_k/2}}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} \cdot {}_U|\boldsymbol{\Sigma}^*+\boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\nu\boldsymbol{\iota}_k+1)/2}.$$

Since $\boldsymbol{\nu} = \nu\boldsymbol{\iota}_k$, $\operatorname{Var}(\boldsymbol{y}) = \boldsymbol{\Sigma} = \frac{\boldsymbol{\Sigma}^*}{\nu-k-1} = \frac{\boldsymbol{\Sigma}^*}{\nu_T-2}$ as $\nu_T = \nu-k+1$. Hence $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\nu_T-2)$. Using Lemma (1i),

the equalities $\nu_T = \nu - k + 1$ and $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}(\nu_T - 2)$ and the matrix calculus result of above we obtain

$$
\begin{aligned}
p_{\mathcal{TR}}(\boldsymbol{y}; \boldsymbol{\Sigma}^*, \nu \cdot \boldsymbol{\iota}_k) &= \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} {}_U|\boldsymbol{\Sigma}^*|_{\nu\boldsymbol{\iota}_k/2} {}_U|\boldsymbol{\Sigma}^* + \boldsymbol{y}\,\boldsymbol{y}^\top|_{-(\boldsymbol{\nu}\cdot\boldsymbol{\iota}_k+1)/2} \\
&\overset{\text{From Lemma (1i)}}{=} \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} |\boldsymbol{\Sigma}^*|^{\nu/2} |\boldsymbol{\Sigma}^* + \boldsymbol{y}\,\boldsymbol{y}^\top|^{-(\nu+1)/2} \\
&\overset{\text{Matr Calc result}}{=} \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} |\boldsymbol{\Sigma}^*|^{\nu/2} |\boldsymbol{\Sigma}^*|^{-(\nu+1)/2} \left(1 + \boldsymbol{y}\,(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{y}^\top\right)^{-(\nu+1)/2} \\
&= \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} |(\nu_T - 2)\boldsymbol{\Sigma}|^{\nu_T + k - 1/2} |(\nu_T - 2)\boldsymbol{\Sigma}|^{-(\nu_T+k)/2} \left(1 + \frac{\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^\top}{\nu_T - 2}\right)^{-(\nu_T+k)/2} \\
&= \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)(\pi)^{k/2}} |(\nu_T - 2)\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^\top}{\nu_T - 2}\right)^{-(\nu_T+k)/2} \\
&= \frac{\overline{\Gamma}\left((\nu\boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu\boldsymbol{\iota}_k/2\right)((\nu_T - 2)\pi)^{k/2} |\boldsymbol{\Sigma}|^{1/2}} \left(1 + \frac{\boldsymbol{y}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}^\top}{\nu_T - 2}\right)^{-(\nu_T+k)/2} \\
&= \frac{\Gamma\left((\nu_T + k)/2\right)}{\Gamma\left(\nu_T/2\right)((\nu_T - 2)\pi)^{k/2} |\boldsymbol{\Sigma}|^{1/2}} \left(1 + \frac{\boldsymbol{y}^\top\,\boldsymbol{\Sigma}^{-1}\boldsymbol{y}}{\nu_T - 2}\right)^{-(\nu_T+k)/2} \\
&= p_T(\boldsymbol{y}; \boldsymbol{\Sigma}, \nu_T),
\end{aligned}
$$

where the second last step comes from the fact that if $\nu = \nu_T + k - 1$, then

$$
\begin{aligned}
\frac{\overline{\Gamma}\left((\nu \cdot \boldsymbol{\iota}_k + 1)/2\right)}{\overline{\Gamma}\left(\nu \cdot \boldsymbol{\iota}_k/2\right)} &= \frac{\pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left((\nu_T + k - 1 + 1)/2 + \frac{1-i}{2}\right)}{\pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\left((\nu_T + k - 1)/2 + \frac{1-i}{2}\right)} \\
&= \frac{\Gamma\left(\frac{\nu_T+k+1-1}{2}\right)\Gamma\left(\frac{\nu_T+k+1-2}{2}\right)\cdots\Gamma\left(\frac{\nu_T+k+1-(k-1)}{2}\right)\Gamma\left(\frac{\nu_T+k+1-k}{2}\right)}{\Gamma\left(\frac{\nu_T+k-1+1-1}{2}\right)\Gamma\left(\frac{\nu_T+k-1+1-2}{2}\right)\cdots\Gamma\left(\frac{\nu_+k-1+1-(k-1)}{2}\right)\Gamma\left(\frac{\nu_T+k-1+1-k}{2}\right)} \\
&= \frac{\Gamma\left(\frac{\nu_T+k}{2}\right)}{\Gamma\left(\frac{\nu_T}{2}\right)}.
\end{aligned}
$$

□

To make the paper self-contained, we also state the first moment of the inverse Riesz distribution. We refer directly to Díaz-García (2013) and Louati and Masmoudi (2015) for the proofs.

**Theorem 2** (**Expectation of the inverse Riesz distribution**).

*(i) Let* $\boldsymbol{Y} \sim i\mathcal{R}^I(\mathbf{I}, \boldsymbol{\nu})$, *then* $\mathbb{E}[\boldsymbol{Y}] = \sum_{i=1}^{k} c_i a_i$ *with* $c_i = \operatorname{diag}(e_i)$ *and*

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

*(ii) Let* $\boldsymbol{Y} \sim i\mathcal{R}^{II}(\mathbf{I}, \boldsymbol{\nu})$*, then* $\mathbb{E}[\boldsymbol{Y}] = \sum_{i=1}^{k} c_i a_i$ *with* $c_i = \text{diag}(e_i)$ *and*

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

# B Simulation results

To investigate the properties of the new model, we perform four simulation experiments. All simulation experiments are based on samples of 1000 observations. We then use Maximum Likelihood to estimate the parameters of interest. In addition, we estimate their standard errors by computing the inverse of the (negative) Hessian at the optimum. We replicate each experiment 1000 times.

The first experiment aims to assess the small sample properties of the MLE for the DoF parameters and the elements of $\boldsymbol{V}$. We simulate vectors $\boldsymbol{y}_t$ of dimension $k = 2$ from the $t$-Riesz distribution. We set $\boldsymbol{V} = \boldsymbol{L}\boldsymbol{L}^\top$ with $\boldsymbol{L}$ the lower Cholesky matrix with elements $\boldsymbol{L}_{11} = 2.752$, $\boldsymbol{L}_{21} = 2.125$, $\boldsymbol{L}_{22} = 3.006$ and $\boldsymbol{\nu} = (5, 10)$. To save space, we only consider the type-I distributions. Results on the type-II distributions are qualitatively similar and are available upon request.

The second experiment focuses on the estimation of the DoF parameters in a 5-dimensional case, where now the elements of $\boldsymbol{V}$ are estimated using a targeting approach as explained in Section 2.2. We set $\boldsymbol{\nu} = (4, 6, 8, 10, 12)$ for the $t$-Riesz distribution.

Panel A of Table B.1 presents the results of all the first two simulation experiments. We find that all parameters are estimated near their true values. Comparing the Monte-Carlo standard error of the estimates (std column) with the mean of the estimated standard error over all replications (mean(s.e.) column), we find that our computed standard errors fairly reflect estimation uncertainty. We see an increase in the variability of the $\boldsymbol{\nu}$ parameters in both panels A.1 and A.2. It appears that the sensitivity of the DoF parameters decreases with respect to the log-likelihood (and variance) of a standard $t$-Riesz distribution.

The third simulation experiment investigates the statistical gain of the $t$-Riesz distribution with various DoF parameters over the standard multivariate Student's $t$ distribution with just one DoF parameter. Guided by the empirical application, we focus on a 5-variate $t$-Riesz I distribution with DoF vector $\boldsymbol{\nu} = (3.97, 5.71, 6.90, 12.4, 9.24)^\top$. We define $\bar{\nu} = 10$ and $\boldsymbol{\nu}_{range} = \boldsymbol{\nu} - \bar{\nu}\boldsymbol{\iota}_k$. The simulation experiment now consists of the following steps. First, we simulate 1000 vectors $\boldsymbol{y}_t$ from a $\mathcal{TR}(0, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\nu}})$ with $\tilde{\boldsymbol{\nu}} = \bar{\nu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\nu}_{range}$ for $\lambda = (0, 0.15, \ldots, 0.60, 0.90)$. Note that if $\lambda = 0$, the $t$-Riesz distribution collapses to a multivariate Student's $t$ distribution with DoF parameter $\bar{\nu}$. On the other hand, if $\lambda = 1$ we have a $t$-Riesz distribution with DoF vector $\boldsymbol{\nu}$. Second, we estimate $\boldsymbol{\Sigma}$ (using the targeting approach) and the DoF parameters assuming a $t$-Riesz distribution or a multivariate Student's $t$ distribution. For each $\lambda$ we test the null-hypotheses $\boldsymbol{\nu} = \bar{\nu}\boldsymbol{\iota}_k$. This boils down to the Likelihood-Ratio test with $k - 1$ degrees of freedom. We repeat this exercise 1000 times.

The results of the third experiment are shown in Panel B of Table B.1. If we simulate from a multivariate Student's $t$ distribution (i.e. $\lambda = 0$), the likelihood ratio test has been rejected in 10 % of all cases. Further, when we deviate slightly from the multivariate Student's $t$ distribution, we increasingly reject the null hypothesis of a scalar $\nu$. For example, if $\lambda = 0.6$, corresponding to $\boldsymbol{\nu} = (6.38, 7.42, 8.13, 11.44, 9.54)^\top$, we reject in 83% of all cases.

The final simulation experiment investigates the small sample properties of the dynamic BEKK $t$-Riesz model of (2)–(3). Guided by the empirical results, we simulate time-varying covariance matrices and return vectors with $k = 10$. We set $\boldsymbol{\nu} = (5, 7.5, 10, 12, 12, 15, 14, 14, 16, 16)^\top$, $A = 0.01$ and $B = 0.98$. Similar to the second and third experiments, we use a targeting approach to estimate $\boldsymbol{\Omega}$ and use maximum likelihood to estimate the $A$, $B$, and the DoF parameters.

Panel C of Table Table B.1 lists the small sample properties of the parameters of the dynamic BEKK $t$-Riesz model. We see that $A$ and $B$ and most of the DoF parameters are estimated near their true values. There seems to be a small upward bias in the DoF parameters. Note again that our computed standard errors fairly reflect estimation uncertainty.

**Table B.1: Parameter estimations of the $t$-Riesz distributions**

This table shows Monte Carlo averages and standard deviations (in parentheses) of parameter estimates of simulated vectors from the $t$-Riesz distribution (type-I) of dimensions two and five. Panel A.1 corresponds to the bivariate case, where both the (Cholesky elements $\boldsymbol{L}_{11}$ ($\boldsymbol{L}_{22}$), $\boldsymbol{L}_{21}$ of $\boldsymbol{V}$ as well as the degrees of freedom (DoF) parameters $\boldsymbol{\nu}$ are estimated. Panel A.2 shows the results of the five-variate case, where the elements of $\boldsymbol{V}$ are estimated by targeting in the first step, and the DoF parameters are estimated in the second step by maximum likelihood. Panel B lists Monte Carlo results on the difference between the multivariate Student's $t$ distribution and the $t$-Riesz distribution. We simulate 1000 vectors from a $\mathcal{TR}(0, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\nu}})$ with $\tilde{\boldsymbol{\nu}} = \bar{\nu}\boldsymbol{\iota}_k + \lambda\boldsymbol{\nu}_{range}$ for $\lambda = (0, 0.15, \ldots, 0.60, 0.90)$ with $\bar{\nu} = 10$ and $\boldsymbol{\nu}_{range} = \boldsymbol{\nu} - \bar{\nu}\boldsymbol{\iota}_k$ with $\boldsymbol{\nu} = (3.97, 5.71, 6.90, 12.4, 9.24)^\top$. We estimate the parameters assuming a $t(0, \boldsymbol{\Sigma}, \tilde{\nu})$ or a $\mathcal{TR}(0, \boldsymbol{\Sigma}, \tilde{\boldsymbol{\nu}})$ distribution. For each value of $\lambda$ we perform a likelihood ratio test on the null hypothesis $\boldsymbol{\nu} = \bar{\nu}\boldsymbol{\iota}_k$. The table lists the percentage rejections of this hypothesis for different values of $\lambda$. Panel C reports results on the 10-dimensional BEKK $t$-Riesz model of (2)–(3) with $\boldsymbol{\nu} = (5, 7.5, 10, 12, 12, 15, 14, 14, 16, 16)^\top$ with $T = 1000$. The table reports the true values, the mean and standard deviation of the estimated coefficients, as well as the mean of the computed standard error. Results are based on 1000 Monte Carlo replications.

Panel A: Small sample properties of MLE

| Panel A.1: dimension 2 | | | | | Panel A.2: dimension 5 (targeting) | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Coef. | True | mean | std | mean(s.e.) | Coef. | True | mean | std | mean(s.e.) |
| $L_{11}$ | 2.752 | 2.752 | 0.088 | 0.090 | $\nu_1$ | 4 | 4.07 | 0.46 | 0.40 |
| $L_{21}$ | 2.125 | 2.127 | 0.123 | 0.125 | $\nu_2$ | 6 | 6.07 | 0.58 | 0.55 |
| $L_{22}$ | 3.006 | 3.003 | 0.085 | 0.085 | $\nu_3$ | 8 | 8.13 | 0.81 | 0.77 |
| | | | | | $\nu_4$ | 10 | 10.03 | 0.85 | 0.92 |
| $\nu_1$ | 5 | 5.12 | 0.77 | 0.79 | $\nu_5$ | 12 | 12.16 | 1.09 | 1.15 |
| $\nu_2$ | 10 | 10.26 | 1.81 | 1.84 | | | | | |

Panel B: multivariate Student's $t$ vs $t$-Riesz

| $\lambda$ | 0 | 0.15 | 0.3 | 0.45 | 0.6 | 0.75 | 0.90 |
|---|---|---|---|---|---|---|---|
| perc rejections | 0.10 | 0.16 | 0.29 | 0.55 | 0.83 | 0.97 | 1.00 |

Panel C: BEKK $t$-Riesz model parameters

| Coef. | True | mean | std | mean(s.e.) |
|---|---|---|---|---|
| A | 0.010 | 0.010 | 0.001 | 0.001 |
| B | 0.980 | 0.975 | 0.003 | 0.003 |
| $\nu_1$ | 5 | 5.28 | 0.764 | 0.755 |
| $\nu_2$ | 7.5 | 7.82 | 0.926 | 1.042 |
| $\nu_3$ | 10 | 10.37 | 1.242 | 1.452 |
| $\nu_4$ | 12 | 12.33 | 1.271 | 1.630 |
| $\nu_5$ | 12 | 12.33 | 1.151 | 1.250 |
| $\nu_6$ | 15 | 15.13 | 1.299 | 1.737 |
| $\nu_7$ | 14 | 14.24 | 0.992 | 1.094 |
| $\nu_8$ | 14 | 14.25 | 0.806 | 0.787 |
| $\nu_9$ | 16 | 16.16 | 0.922 | 1.029 |
| $\nu_{10}$ | 16 | 16.24 | 0.769 | 0.767 |