A DYNAMIC CONDITIONAL SCORE MODEL FOR THE LOG CORRELATION MATRIX

Christian M. Hafner, Linqi Wang

REPRINT • 2022 / 06

UCLouvain

LFIN | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

A dynamic conditional score model for the log correlation matrix

Christian M. Hafner* Linqi Wang†

February 24, 2022

Abstract

This paper proposes a new model for the dynamics of correlation matrices, where the dynamics are driven by the likelihood score with respect to the matrix logarithm of the correlation matrix. In analogy to the exponential GARCH model for volatility, this transformation ensures that the correlation matrices remain positive definite, even in high dimensions. For the conditional distribution of returns, we assume a student-t copula to explain the dependence structure and univariate student-t for the marginals with potentially different degrees of freedom. The separation into volatility and correlation parts allows for a two-step estimation, which facilitates estimation in high dimensions. We derive estimation theory for one-step and two-step estimation. In an application to a set of six asset indices including financial and alternative assets we show that the model performs well in terms of diagnostics, specification tests, and out-of-sample forecasting.

Keywords: score, correlation, matrix logarithm, identification

JEL classification: **C14, C43, Z11**

*Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM), and ISBA, UCLouvain, christian.hafner@uclouvain.be

†Louvain Institute of Data Analysis and Modelling in economics and statistics (LIDAM), and LFIN, UCLouvain, linqi.wang@uclouvain.be.

1 Introduction

Correlation matrices are pervasive in many applied sciences such as finance and econometrics. Reliable models for the dynamics of correlations between different assets or underlying risk factors are essential for asset pricing, portfolio selection and risk management. Nonetheless, correlation matrices are notoriously difficult to model in high dimensions, for at least two reasons: the positivity constraint and the balance between flexibility and computational feasibility. These challenges are well documented and addressed in the framework of multivariate GARCH models, see e.g. Bauwens et al. (2006), Bauwens et al. (2012) and Francq and Zakoian (2019).

This paper proposes a new model for the joint modelling of volatilities and correlations, which both are driven by the respective conditional scores of the likelihood, and hence fits into the general modelling framework of *dynamic conditional score* or, in an alternative terminology, *generalized autoregressive score* models, as introduced by the seminal works of Creal et al. (2011), Creal et al. (2013) and Harvey (2013). Score driven correlation models have been proposed recently in the context of modelling realized variances and covariances. For example, Gorgi et al. (2018) use the Wishart distribution, and Opschoor et al. (2018) and Vassallo et al. (2021) the more flexible matrix-F distribution.

Our aim is to allow for fat tails in the conditional distributions, and in particular, to generalize the popular Beta-t-EGARCH model proposed in Harvey (2013) (Section 4.1), which uses the student-t distribution, to the multivariate case. The conditional covariance matrix can be decomposed into volatility and correlation components, which facilitates estimation since it can be split into two parts as in the DCC model of Engle (2002) and in Vassallo et al. (2021).

For the correlation part, we specify the dynamics for the matrix logarithm of the correlation matrix to guarantee that positive definiteness is preserved in arbitrary dimensions. This approach is analogous to the Beta-t-EGARCH model in which the log volatility process is modelled to ensure positivity. In the bivariate case, the matrix log reduces to the Fisher transformation of correlations which has been adopted e.g. by Harvey (2013) (Sec-

tion 7.3.1) and Hansen et al. (2014). The idea of modelling the matrix log of a correlation or covariance matrix is not new. For example, Kawakatsu (2006) proposed a multivariate version of the EGARCH model of Nelson (1991), but with *ad hoc* rather than score-driven dynamics. In a static framework, Hafner et al. (2020) impose a multiplicative structure for the correlation matrix, which becomes additive under the matrix log transformation, and hence easier to estimate using minimum distance type estimators.

Modelling the log correlation matrix circumvents the need to impose parameter restrictions to ensure positive definiteness of the resulting correlation matrix and thus allows the implementation of an unconstrained model for the vectorized log correlation parameter. Moreover, block structures are preserved under the matrix log transformation, as shown by Archakov and Hansen (2020), so that factor structures may be imposed for the log correlation parameter to allow for estimation in high dimensions. This idea is used by Archakov et al. (2020) to estimate a multivariate realized GARCH model.

The parameter characterizing the dynamics of the log correlation matrix will be driven by the score of the likelihood with respect to this parameter. The distribution characterizing the dependence part of the model is assumed to be a student-t copula. If all conditional marginal distributions happen to be univariate student-t with the same degrees-of-freedom, equal to the parameter of the t-copula, then we have a multivariate student-t distribution. However, we do not impose this restriction, so we can have what is commonly called a Meta-t distribution, see Demarta and McNeil (2005), i.e. a student-t copula combined with arbitrary student-t marginals.

There is of course an identification issue because for an $(N \times N)$ correlation matrix, only $N(N-1)/2$ elements are freely determinable. Consequently, for the matrix log of a correlation matrix, which is symmetric by definition, the diagonal elements are implied by the off-diagonal elements, although no analytical form is available in higher dimensions. Archakov and Hansen (2021) propose an algorithm that reconstructs the original correlation matrix based on the lower-triangular part, excluding the diagonal, of the log correlation matrix. We follow the same approach here by endowing this $N(N-1)/2$-dimensional parameter with score-driven dynamics.

We propose two approaches for estimation, the first is a one-step estimator where all parts of the model are estimated simultaneously, which ensures statistical efficiency but may not be computationally feasible in high dimensions. In this case, we assume equality of the conditional marginal distributions, so that the conditional distribution is indeed a multivariate student-t with only one degree-of-freedom parameter. The second is a two-step estimator where the volatilities are estimated in a first step, and the correlations in a second step. The marginal student-t distributions are allowed to have different degrees-of-freedom parameters. This estimator, although statistically inefficient, is computationally simple and feasible even in higher dimensions. We derive the asymptotic estimation theory for both estimators.

In an empirical application to a set of six financial, commodity and cryptocurrency indices, we show that the proposed model fits well the data using two-step estimation and a variety of diagnostics and specification tests. We conclude that parameter restrictions across the components of the log correlation parameter may be employed to reduce the dimension of the parameter space and are supported by classical information criteria. Furthermore, we conduct a performance comparison of the proposed log correlation model with alternative parameterizations for both in-sample estimation and out-of-sample forecasting. We show that our model provides the best in-sample model fit and has similar out-of-sample performance compared with other score-driven models.

The remainder of the paper is organized as follows. The following section recalls the definition and the structure of the matrix log of a correlation matrix. Section 3 introduces the model and proposes a one-step and two-step maximum likelihood estimator, including their corresponding asymptotic estimation theories. Section 4 presents a detailed empirical application of the model, and Section 5 concludes. The appendix is split into two parts, Appendices A and B are at the end of this paper version, while Appendices C to G are delegated to a separate file that is available online. Theoretical material including proofs of the theorems is collected in Appendix A, and some tables and figures of the empirical example are shown in Appendix B. Appendix C shows that our model reduces to the one of Harvey (2013), Section 7.3.1, in the bivariate Gaussian case. Appendix D

includes additional derivations on the Lyapounov exponent of the stochastic recurrence equation, and Appendix E gives analytical expressions for the information matrix. We provide a Monte Carlo study in Appendix F to analyze the finite sample performance of the one-step and two-step estimation procedures, and additional tables and figures are collected in Appendix G.

2 Matrix Logarithm of a Correlation matrix

The matrix logarithm of a correlation matrix Θ is defined as $\log \Theta = \Gamma(\log \Lambda)\Gamma^\intercal$, where Γ is the matrix containing the eigenvectors of Θ in its columns, and Λ is a diagonal matrix containing the corresponding eigenvalues on the diagonal, which are all strictly positive. The diagonal matrix $\log \Lambda$ replaces the eigenvalues on the diagonal by their logarithm. The matrix $\log \Theta$ has a complicated structure, its diagonal elements can take any non-positive values (see Lemma 2 of Archakov and Hansen (2021)) and its off-diagonals can take any values. As an illustration, suppose that

$$\Theta = \begin{pmatrix} 1 & 0.8 & 0.5 \\ 0.8 & 1 & 0.2 \\ 0.5 & 0.2 & 1 \end{pmatrix}, \quad \text{then} \quad \log \Theta = \begin{pmatrix} -0.75 & 1.18 & 0.64 \\ 1.18 & -0.55 & -0.07 \\ 0.64 & -0.07 & -0.17 \end{pmatrix}.$$

There are $N(N+1)/2$ parameters in $\log \Theta$; we call these *log parameters*. On the other hand, Θ has only $N(N-1)/2$ original parameters. For each Θ, its $N(N-1)/2$ original parameters completely identify its $N(N+1)/2$ log parameters. In other words, there exists a function $f : \mathbb{R}^{N(N-1)/2} \to \mathbb{R}^{N(N+1)/2}$ which maps the original parameters to the log parameters. However, when $N > 4$, f does *not* have a closed form because when $N > 4$ the continuous functions which map elements of a matrix to its eigenvalues have no closed form. When $N = 2$, we can solve f by hand (see Example 1).

Example 1. *Suppose*

$$\Theta = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The eigenvalues of Θ *are* $1+\rho$ *and* $1-\rho$*, respectively. The corresponding eigenvectors are* $(1,1)^\intercal/\sqrt{2}$ *and* $(1,-1)^\intercal/\sqrt{2}$*, respectively. Therefore*

$$\begin{aligned}\log\Theta &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \log(1+\rho) & 0 \\ 0 & \log(1-\rho) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{1}{2} \\ &= \begin{pmatrix} \frac{1}{2}\log(1-\rho^2) & \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) \\ \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) & \frac{1}{2}\log(1-\rho^2) \end{pmatrix}.\end{aligned}$$

Thus

$$f(\rho) = \left(\frac{1}{2}\log(1-\rho^2), \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{2}\log(1-\rho^2)\right)^\intercal,$$

and we see that ρ *generates two distinct entries for* $\log\Theta$*. The off-diagonal element* $\frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right)$ *is the well-known Fisher's z-transformation of* ρ*. We also see that* $\log\Theta$ *is not only symmetric about the diagonal, but also symmetric about the cross-diagonal (from the upper right to the lower left).*

The idea of Archakov and Hansen (2021) is to identify a subset of unrestricted log parameters and then fill in the remainder afterwards. They propose the following procedure. Let θ be the vector containing only the lower triangular portion of $\log\Theta$, excluding the diagonal. Then, the diagonal elements of $\log\Theta$ are uniquely determined by some function $\phi : \mathbb{R}^{N(N-1)/2} \to \mathbb{R}^N$, which can be obtained numerically (in the case $N = 2$ there is a closed form, see Example 1). Archakov and Hansen (2021) gave a concrete algorithm to do this and established its validity. We will use this idea in our model, starting from a dynamic parameter θ_t, to first reconstruct the full log correlation matrix $\log\Theta_t$, and then by matrix exponentiation the original correlation matrix Θ_t.

3 The model and its estimation

We introduce a model that combines conditional heteroskedasticity, dynamic conditional correlations, and conditional leptokurtosis in a dynamic score-driven framework. Consider

the return vector for N assets, $y_t = (y_{1t}, \ldots, y_{Nt})^\intercal$. To simplify the presentation and discussion of the model, we will first assume that the conditional distribution of y_t given the past is a multivariate student-t distribution, but later allow for marginal distributions with different degrees of freedom parameters.

We further assume that the parameter vector ϑ can be separated into a vector τ that characterizes the marginal distributions and the volatility part, a vector δ which characterizes the correlation part, and the degrees-of-freedom parameter η, and hence, $\vartheta = (\tau^\intercal, \delta^\intercal, \eta)^\intercal$.

The core of the model is the process generating the conditional correlation matrix Θ_t. Before presenting this part of the model, we introduce the model framework for y_t including the marginal dynamics and the distribution conditional on the information set $\mathcal{F}_{t-1}$,

$$
\begin{aligned}
y_t|\mathcal{F}_{t-1} &\sim F_{\eta,H_t}, \quad \eta > 2 && (3.1)\\
H_t(\vartheta) &= D_t(\tau)\Theta_t(\delta)D_t(\tau) && (3.2)\\
D_t(\tau) &= \operatorname{diag}((\exp(h_{1t}), \ldots, \exp(h_{Nt}))^\intercal) && (3.3)\\
h_{it}(\tau_i) &= \omega_i^v + \phi_i^v h_{i,t-1} + \kappa_i^v u_{i,t-1}^v, \quad i = 1, \ldots, N && (3.4)\\
u_{it}^v &= \frac{(\eta+1)y_{it}^2}{(\eta-2)\exp(2h_{it}) + y_{it}^2} - 1, \quad i = 1, \ldots, N && (3.5)
\end{aligned}
$$

where F_{η,H_t} is the standardized multivariate student-t distribution with η d.o.f., $\eta > 2$, and dispersion matrix H_t, i.e. its log density is given by

$$
f_{\eta,H_t}(y_t) = k(\eta) - \frac{1}{2}\left\{\log|H_t| + (\eta+N)\log\left(1 + \frac{1}{\eta-2}y_t^\intercal H_t^{-1} y_t\right)\right\}.
$$

where $k(\eta) = \log\Gamma\left(\frac{\eta+N}{2}\right) - \log\Gamma\left(\frac{\eta}{2}\right) - \frac{N}{2}\log\big((\eta-2)\pi\big)$. The structure of the model implies that the conditional mean of y_t is zero, the conditional variances are given by $\exp(2h_{it})$, and the conditional correlation matrix is given by Θ_t. Each one of D_t, Θ_t and H_t are assumed to be measurable functions of the information set $\mathcal{F}_{t-1}$.

Each asset's volatility is assumed to follow a univariate Beta-t-EGARCH process, where the score with respect to log volatility is given by (3.5). The only difference with

respect to the standard formulation of this model in Harvey (2013) is that we do not consider the classical but the standardized student-t distribution, but the score is the same up to a scaling factor that depends on the degrees-of-freedom. It is straightforward to generalize this model to allow for volatility spillover, perhaps following some test for Granger causality in volatility. To keep the notation simple and estimation feasible in high dimensions, we refrain from doing so here as it is not directly relevant for our main focus, the modelling of the correlation part.

We denote by $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\intercal$ the vector of standardized residuals obtained after estimation of the volatility model, i.e., $\varepsilon_t = D_t(\tau)^{-1} y_t$, so that the components of ε_t have conditional mean zero and conditional variance unity. However, the components of ε_t are conditionally correlated, and the conditional correlation matrix is given by $\Theta_t = \mathrm{E}[\varepsilon_t \varepsilon_t^\intercal | \mathcal{F}_{t-1}]$.

Now consider the following score-driven model for the conditional correlation matrix $\Theta_t(\delta)$. As in Archakov and Hansen (2021) we model the lower triangular part of $\log \Theta_t(\delta)$, excluding the diagonal, i.e., $\theta_t(\delta) := \mathrm{vecl}(\log \Theta_t(\delta))$, and then fill in the diagonal elements to ensure that Θ_t is a correlation matrix. The model is a first order dynamic conditional score model for the log correlation parameter, i.e.

$$\theta_t(\delta) = \omega + \Phi \theta_{t-1}(\delta) + K u_{t-1} \tag{3.6}$$

where u_t is the score of the likelihood function with respect to θ_t, and $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_{N^*})$ and $K = \mathrm{diag}(\kappa_1, \ldots, \kappa_{N^*})$ are $N^* \times N^*$ diagonal parameter matrices, where $N^* = N(N-1)/2$. We restrict Φ and K to be diagonal, as otherwise the number of parameters would be of order $O(N^4)$, which is infeasible in high dimensions. However, similar to the volatility case, one might consider "log correlation spillover", and tests thereof, but interpretation is not obvious since the link between the log parameters and correlations is not trivial.

The specific form of the score vector u_t depends of course on the conditional distribution, which will be discussed in detail in the following in the context of maximum likelihood estimation. We distinguish between a multivariate student-t distribution, which is used

for one-step estimation, and a Meta-t distribution which is used for two-step estimation. The score vectors in both cases are very similar but not exactly the same. Furthermore, one may use standardized versions of the score vector, by pre-multiplying u_t by the inverse of the (square root of the) information matrix, see e.g. the model of Harvey (2013), Section 7.3.1., which is motivated in particular by cases where the standardized score is an i.i.d. random vector. However, this will generally not be the case in our setting, so we will not pursue this strategy here to keep things simple.

Although one of the main advantages of our model specification lies in the separation of volatility and correlation parts, just as in Engle (2002) and related works, it would be more efficient from a statistical viewpoint to do the estimation of all model parameters in one step. In the following, we therefore present both approaches, one-step and two-step estimation.

3.1 One step estimation

For one-step estimation we assume for simplicity that $\eta_1 = \ldots = \eta_N = \eta$, so that the conditional joint distribution of y_t is a multivariate student-t. The log likelihood function for a sample of T observations as a function of the parameter vector $\vartheta = (\tau^\intercal, \delta^\intercal, \eta)^\intercal$, $\delta := (\omega^\intercal, \mathrm{diag}(\Phi)^\intercal, \mathrm{diag}(K)^\intercal)^\intercal$ takes the form

$$L(\vartheta) = \sum_{t=1}^{T} l_t(\vartheta) = T \log \Gamma\left(\frac{\eta+N}{2}\right) - T \log \Gamma\left(\frac{\eta}{2}\right) - \frac{TN}{2} \log\big((\eta-2)\pi\big) \\ - \frac{1}{2}\sum_{t=1}^{T}\left\{2\log|D_t(\tau)| + \log|\Theta_t(\delta)| + (\eta+N)\log\left(1 + \frac{1}{\eta-2}\varepsilon_t^\intercal \Theta_t(\delta)^{-1}\varepsilon_t\right)\right\} \tag{3.7}$$

The spectral decomposition of the log correlation matrix is denoted by $\log \Theta_t = \Gamma_t \Lambda_t \Gamma_t^\intercal$, where Γ_t is the $(N \times N)$ matrix whose columns are the eigenvectors of $\log \Theta$, and Λ_t is the $(N \times N)$ diagonal matrix containing the corresponding eigenvalues $\lambda_{it}, i = 1, \ldots, N$ on the diagonal. Define the matrix

$$A_t := (\Gamma_t \otimes \Gamma_t)\Xi_t(\Gamma_t \otimes \Gamma_t)^\intercal, \tag{3.8}$$

where Ξ_t is the $N^2 \times N^2$ matrix with elements given by

$$\xi_{ij} = \Xi_{(i-1)N+j,(i-1)N+j} = \begin{cases} \exp(\lambda_{it}), & \text{if } \lambda_{it} = \lambda_{jt} \\ \frac{\exp(\lambda_{it})-\exp(\lambda_{jt})}{\lambda_{it}-\lambda_{jt}}, & \text{if } \lambda_{it} \neq \lambda_{jt} \end{cases}$$

for $i, j = 1, \ldots, N$. Furthermore, define the following elimination matrices: E_d $(N \times N^2)$, E_l $(N(N-1)/2 \times N^2)$, and E_u $(N(N-1)/2 \times N^2)$ such that $\text{vecl}(Q) = E_l \text{vec}(Q)$, $\text{vecl}(Q^\intercal) = E_u \text{vec}(Q)$ and $\text{diag}(Q) = E_d \text{vec}(Q)$ for any square matrix Q, and $E_s := E_l + E_u$. Finally, define the permutation matrix P as $P(\text{diag}(Q)^\intercal, \text{vecl}(Q)^\intercal)^\intercal = \text{vec}(Q)$ for any symmetric conformable matrix Q. With this notation, the following theorem provides the analytic form of the score with respect to the log correlation parameter.

Theorem 1. *The score with respect to the log correlation parameter θ_t is given by*

$$u_t = \frac{\partial l_t}{\partial \theta_t} = \frac{1}{2} M_t vec\left(\frac{\eta + N}{\eta - 2 + \varepsilon_t^\intercal \Theta_t^{-1} \varepsilon_t} \Theta_t^{-1} \varepsilon_t \varepsilon_t^\intercal \Theta_t^{-1} - \Theta_t^{-1} \right) \tag{3.9}$$

where

$$M_t := \begin{pmatrix} -(E_d A_t E_d^\intercal)^{-1} E_d A_t E_s^\intercal \\ I_{N(N-1)/2} \end{pmatrix}^\intercal P^\intercal A_t \tag{3.10}$$

Remark 1. *In the limiting case $\eta \to \infty$, we obtain the Gaussian distribution as a special case, for which we let $\vartheta = (\tau^\intercal, \delta^\intercal)^\intercal$, and the log likelihood function takes the form*

$$L(\vartheta) = -\frac{TN}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} \left\{ 2 \log |D_t(\tau)| + \log |\Theta_t(\delta)| + \varepsilon_t^\intercal \Theta_t^{-1}(\delta) \varepsilon_t \right\} \tag{3.11}$$

and the score reduces to

$$u_t = \frac{\partial l_t}{\partial \theta_t} = \frac{1}{2} \begin{pmatrix} -(E_d A_t E_d^\intercal)^{-1} E_d A_t E_s^\intercal \\ I_{N(N-1)/2} \end{pmatrix}^\intercal P^\intercal A_t vec(\Theta_t^{-1} \varepsilon_t \varepsilon_t^\intercal \Theta_t^{-1} - \Theta_t^{-1}) \tag{3.12}$$

Remark 2. *Note that the score can be written as $u_t = g(\theta_t) Z_t$ where $g(\theta_t) = \frac{1}{2} M_t (\Theta_t^{-1/2} \otimes \Theta_t^{-1/2})$, and $Z_t = vec\left(\frac{\eta+N}{\eta-2+\xi_t^\intercal \xi_t} \xi_t \xi_t^\intercal - I_N \right)$ is an i.i.d. random vector. The score vector u_t is a martingale difference sequence, which can be shown using results of Fiorentini et al. (2003). However, it is not i.i.d. in general, although it is i.i.d. in the univariate case because, in that case, $g(\theta_t) = 1/2$.*

Remark 3. *The process θ_t has an SRE representation as*

$$\theta_t = \psi_t(\theta_{t-1}) \tag{3.13}$$

where $\psi_t(s) = \omega + \Phi s + Kg(s)Z_t$. Appendix A.1 shows that, under our assumptions, this is a Lipschitz map with maximum Lipschitz constant $\Lambda(\psi) < \infty$ a.s. We derive an analytical expression for $\Lambda(\psi)$ in the bivariate case, while in higher dimensions, we show that there exists an upper bound. See Appendix A.1 and Appendix D for details.

The volatility part of the likelihood, i.e., $D_t(\tau)$ in (3.7), is straightforward to calculate and exactly the same as in univariate Beta-t-EGARCH models. For the correlation part we use the following algorithm to compute $\Theta_t(\delta)$ and the score u_t, conditional on the volatility parameter τ:

1. Initialization: Set $u_0 = 0$ and $\theta_0 = (I_{N(N-1)/2} - \Phi)^{-1}\omega$.
2. For $t = 1, \ldots, T$:
   (a) Calculate θ_t as a function of past observations according to (3.6).
   (b) Reconstruct $\text{diag}(\log \Theta_t)$ using the algorithm of Archakov and Hansen (2021).
   (c) Calculate the matrix exponential of $\log \Theta_t$ to obtain Θ_t.
   (d) Calculate the score u_t according to (3.9).
3. Calculate the log likelihood (3.7).

The MLE is then defined as the maximizer of the log likelihood function over some parameter space Ω,

$$\hat{\vartheta} = \arg\max_{\vartheta \in \Omega} L(\vartheta). \tag{3.14}$$

No analytical solution is available, but standard numerical optimization routines can be employed. For the estimation theory we use the following assumptions.

(A1) The conditional distribution of y_t is a standardized multivariate student-t distribution with η degrees of freedom, $\eta > 2$ and dispersion matrix H_t.

(A2) The parameter vector $\vartheta \in \Omega$, where Ω is a compact subset of $\mathbb{R}^J$, $J = \dim(\vartheta)$, such that $|\phi_i^v| < 1, i = 1, \ldots, N$, $|\phi_j| < 1, j = 1, \ldots, N^*$, $\kappa_i^v \neq 0, i = 1, \ldots, N$, and $\kappa_j \neq 0, j = 1, \ldots, N^*$.

(A3) The true parameter ϑ_0 lies in the interior of Ω.

(A4) $|(\phi_i^v)^2 + 2\kappa_i^v \phi_i^v \mathbb{E}(\partial u_{it}^v / \partial h_{it})| + (\kappa_i^v)^2 \mathbb{E}[(\partial u_{it}^v / \partial h_{it})^2] < 1, \quad i = 1, \ldots, N$

(A5) $|\phi_j^2 + 2\kappa_j \phi_j \mathbb{E}(\partial u_{jt} / \partial \theta_{jt})| + \kappa_j^2 \mathbb{E}[(\partial u_{jt} / \partial \theta_{jt})^2] < 1, \quad j = 1, \ldots, N^*$

(A6) The smallest eigenvalue of Θ_t is bounded away from zero by a positive constant, $\lambda_{min}(\Theta_t) \geq \epsilon > 0$, a.s. We call $\mathcal{T}_{\epsilon,N}$ the set of log-parameters θ_t for which this assumption holds, i.e. we assume $\theta_t \in \mathcal{T}_{\epsilon,N} \subset \mathbb{R}^{N(N-1)/2}$ a.s.

(A7) $\mathbb{E}[\log^+(\Lambda(\psi_t)] < \infty$ and $\mathbb{E}[\log(\Lambda(\psi_t)] < 0$.

Assumption (A2) is standard and necessary for identifiability and stationarity of the volatility and log correlation processes. (A3) is also standard and is needed for asymptotic normality. (A4) and (A5) ensure that the scores with respect to log volatility and log correlation, respectively, and their first derivatives, have finite second moments. They are also standard in this literature, see e.g. the conditions of Theorem 1 of Harvey (2013). Assumption (A6) ensures that the stochastic recurrence equation is Lipschitz continuous, as shown in Appendix D. It is standard in the literature on modelling high-dimensional covariance matrices, see e.g. Fan et al. (2011) and Chang et al. (2018). Assumption (A7) is a contraction condition which ensures that the process θ_t is a stationary and ergodic sequence. For given parameters and lower bound ϵ for the smallest eigenvalue of Θ_t, it can be verified by numerical simulation, see Appendix D for details.

Let $\vartheta = (\tau^\intercal, \delta^\intercal, \eta)^\intercal$, and decompose the information matrix as

$$I = \begin{pmatrix} I_{\tau\tau} & I_{\tau\delta} & I_{\tau\eta} \\ I_{\tau\delta}^\intercal & I_{\delta\delta} & I_{\delta\eta} \\ I_{\tau\eta}^\intercal & I_{\delta\eta}^\intercal & I_{\eta\eta} \end{pmatrix} \tag{3.15}$$

We then have the following result.

Theorem 2. *Under Assumptions (A1)-(A7),* $\hat{\vartheta} \to_p \vartheta$ *and the asymptotic distribution of the estimator in (3.14) is given by*

$$\sqrt{T}(\hat{\vartheta} - \vartheta) \xrightarrow{\mathcal{L}} N(0, I^{-1}) \tag{3.16}$$

where I *is given in (3.15) with*

$$\begin{aligned}
I_{\delta\delta} &= \mathbb{E}\left[\frac{\partial\theta_t^T}{\partial\delta} M_t \left\{\frac{N+\eta}{2(N+\eta+2)}\left(\Theta_t^{-1}\otimes\Theta_t^{-1}\right) - \frac{1}{2(N+\eta+2)} vec(\Theta_t^{-1}) vec(\Theta_t^{-1})^\intercal\right\} M_t^\intercal \frac{\partial\theta_t}{\partial\delta^T}\right] \\
I_{\eta\eta} &= \frac{\eta^4}{4}\left[\psi'\left(\frac{\eta}{2}\right) - \psi'\left(\frac{N+\eta}{2}\right)\right] - \frac{N\eta^4[\eta^2 + N(\eta-4) - 8]}{2(\eta-2)^2(N+\eta)(N+\eta+2)} \\
I_{\delta\eta} &= -\frac{N+2)\eta^2}{(\eta-2)(N+\eta)(N+\eta+2)} \mathbb{E}\frac{\partial\theta_t^\intercal}{\partial\delta} M_t vec(\Theta_t^{-1}) \\
I_{\tau\tau} &= \mathbb{E}\left[\frac{\partial h_t^\intercal}{\partial\tau}\left\{\frac{N+\eta}{2(N+\eta+2)}\left(H_t^{-1}\otimes H_t^{-1}\right) - \frac{1}{2(N+\eta+2)} vec(H_t^{-1}) vec(H_t^{-1})^\intercal\right\}\frac{\partial h_t}{\partial\tau^\intercal}\right] \\
I_{\tau\delta} &= \mathbb{E}\left[\frac{\partial h_t^\intercal}{\partial\tau}\left\{\frac{N+\eta}{2(N+\eta+2)}\left(H_t^{-1}\otimes H_t^{-1}\right) - \frac{1}{2(N+\eta+2)} vec(H_t^{-1}) vec(H_t^{-1})^\intercal\right\}(D_t\otimes D_t) M_t^\intercal \frac{\partial\theta_t}{\partial\delta^\intercal}\right] \\
I_{\tau\eta} &= -\frac{N+2)\eta^2}{(\eta-2)(N+\eta)(N+\eta+2)} \mathbb{E}\frac{\partial vec^\intercal(H_t)}{\partial\tau} vec(H_t^{-1})
\end{aligned}$$

The information matrix further depends on the derivative of the log correlation parameter θ_t with respect to δ, which is given by

$$\frac{\partial\theta_t}{\partial\delta^\intercal} = z_{t-1} + x_{t-1}\frac{\partial\theta_{t-1}}{\partial\delta^\intercal}, \tag{3.17}$$

where

$$z_t = \begin{pmatrix} I_{N^*} & \text{diag}(\theta_{1t},\ldots,\theta_{N^*t}) & \text{diag}(u_{1t},\ldots,u_{N^*t}) \end{pmatrix}, \quad x_t = \Phi + K\frac{\partial u_t}{\partial\theta_t^\intercal}. \tag{3.18}$$

The information matrix can be estimated consistently by evaluating it at $\hat{\vartheta}$ and replacing the expectation operators by sample averages.

3.2 Two-step estimation

For two-step estimation we allow for different d.o.f. parameters of the marginal distributions, $\eta_1,\ldots,\eta_N$. Thus we now assume a conditional Meta-t distribution with degrees

of freedom parameter η and scale matrix Θ_t, meaning that the conditional dependence structure of y_t is determined by a student-t copula with scale matrix Θ_t, see e.g. Demarta and McNeil (2005) for details on Meta-t distributions and the student-t copula.

As before, the conditional marginal distributions are given by score-driven volatility processes and student-t distributions, but now with potentially different degrees of freedom η_i. If all d.o.f. parameters happen to be the same and equal to η, then this model reduces to the one considered for one-step estimation, and in this case two-step estimation would be less efficient than one-step estimation, although it might be the only feasible estimation procedure computationally, at least in high dimensions.

Formally, for the theory we replace Assumption (A1) by the following.

(A1') The conditional distribution of y_t is given by

$$F_t(y_{1t}, \ldots, y_{Nt}) = C_{\eta,\Theta_t}(F_{\eta_1,h_{1t}}(y_{1t}), \ldots, F_{\eta_N,h_{Nt}}(y_{Nt}))$$

with $\eta > 2$ and $\eta_i > 2$, $i = 1, \ldots, N$, where $F_{\eta_i,h_{it}}$ are the standardized student-t marginal cdf's with log scale h_{it}, and C_{η,Θ_t} is the cdf of a standardized student-t copula with dispersion matrix Θ_t.

Note that the copula is invariant to standardizations of the marginal distributions, so that the student-t copula with scale matrix H_t is the same as the student-t copula with scale matrix Θ_t. Note also that if all marginal degrees of freedom are the same and equal to the copula parameter η, then Assumption (A1') collapses to (A1).

We augment the vector of parameters characterizing the marginal distributions as $\tau_i = (\omega_i^v, \phi_i^v, \kappa_i^v, \eta_i)^\intercal, i = 1, \ldots, N$, and $\tau = (\tau_1^\intercal, \ldots, \tau_N^\intercal)^\intercal$. The copula density has the form

$$c_{\eta,\Theta_t}(u) = \frac{f_{\eta,\Theta_t}(t_\eta^{-1}(u_1), \ldots, t_\eta^{-1}(u_N))}{\prod_{i=1}^N g_\eta(t_\eta^{-1}(u_i))}$$

see e.g. Demarta and McNeil (2005), where $f_{\eta,\Theta_t}(\cdot)$ is the joint pdf of a multivariate standardized student-t distribution with scale matrix Θ_t, and $t_\eta(\cdot)$ and $g_\eta(\cdot)$ are standardized univariate student-t cdf and pdf, respectively. The conditional joint density of y_t is then

given by

$$f_t(y_t) = c_{\eta,\Theta_t}(F_{\eta_1,h_{1t}}(y_{1t}),\ldots,F_{\eta_N,h_{Nt}}(y_{Nt}))\prod_{i=1}^{N} g_{\eta_i,h_{it}}(y_{it})$$

which is used to construct the likelihood function.

The log likelihood function for a sample of T observations as a function of the parameter vector $\vartheta = (\tau^\intercal, \delta^\intercal, \eta)^\intercal$, $\delta := (\omega^\intercal, \phi^\intercal, \kappa^\intercal)^\intercal$ takes the form

$$\begin{aligned}
L(\vartheta) &= L^V(\tau) + L^C(\vartheta) \\
L^V(\tau) &= \sum_{i=1}^{N} l_i^V(\tau) \\
l_i^V(\tau) &= c(\eta_i,1) - \frac{1}{2}\sum_{t=1}^{T}\left\{2h_{it} + (\eta_i+1)\log(1+\frac{y_{it}^2\exp(-2h_{it})}{\eta_i-2})\right\} \\
L^C(\vartheta) &= c(\eta,N) - Nc(\eta,1) - \frac{1}{2}\sum_{t=1}^{T} l_t^C(\vartheta) \\
l_t^C(\vartheta) &= \log|\tilde{\Theta}_t(\delta)| + (\eta+N)\log\left(1+\frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t(\delta)^{-1}\tilde{\varepsilon}_t}{\eta-2}\right) - \sum_{i=1}^{N}(\eta+1)\log\left(1+\frac{\tilde{\varepsilon}_{it}^{\,2}}{\eta-2}\right) \\
c(\eta,N) &= T\log\Gamma\left(\frac{\eta+N}{2}\right) - T\log\Gamma\left(\frac{\eta}{2}\right) - \frac{TN}{2}\log\big((\eta-2)\pi\big)
\end{aligned}$$

where $\tilde{\varepsilon}_{it} := s_\eta(\varepsilon_{it},\eta_i) = t_\eta^{-1}(t_{\eta_i}(\varepsilon_{it}))$ is a mapping of the quantiles of a standardized student-t distribution with η_i d.o.f. into the quantiles of a standardized student-t with η d.o.f. The mapping function $s_\eta(\cdot)$ is strictly monotone and its inverse is given by $s_\eta^{-1}(\cdot,\eta_i) = t_{\eta_i}^{-1}(t_\eta(\cdot))$. When applied to a vector, we define $s_\eta(\cdot)$ as $s_\eta(\varepsilon_t,\underline{\eta}) = (s_\eta(\varepsilon_{1t},\eta_1),\ldots,s_\eta(\varepsilon_{Nt},\eta_N))^\intercal$, where $\underline{\eta} = (\eta_1,\ldots,\eta_N)^\intercal$, and similarly for $s_\eta^{-1}(\varepsilon_t,\underline{\eta})$.

Note that the likelihood reduces to that of the one-step estimation when imposing the restriction $\eta_1 = \ldots = \eta_N = \eta$.

The matrix $\tilde{\Theta}_t(\delta)$ is such that $\text{vecl}(\log\tilde{\Theta}_t(\delta)) =: \tilde{\theta}_t(\delta)$ follows the same process (3.6) as above for the one-step method, but in the score u_t replacing $\Theta_t(\delta)$ by $\tilde{\Theta}_t(\delta)$ and ε_t by $\tilde{\varepsilon}_t$, i.e.

$$\begin{aligned}
\tilde{\theta}_t(\delta) &= \omega + \Phi\tilde{\theta}_{t-1}(\delta) + K\tilde{u}_{t-1} \\
\tilde{u}_t &= \frac{1}{2}\tilde{M}_t\text{vec}\left(\frac{\eta+N}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1} - \tilde{\Theta}_t^{-1}\right)
\end{aligned}$$

and where $\tilde{M}_t$ is constructed as M_t above but replacing Θ_t by $\tilde{\Theta}_t(\delta)$.

Now, due to the nonlinear transformation $s_\eta(\cdot)$, $\tilde{\Theta}_t(\delta)$ is not exactly equal to the conditional correlation matrix $\Theta_t(\delta)$. It can however be reconstructed easily by numerical methods. For example, for a given $\tilde{\Theta}_t$, generate a large number K of r.v. X with standardized multivariate student-t distribution with η d.o.f. and correlation matrix $\tilde{\Theta}_t$. Then, obtain the sample correlation matrix of the transformed data as

$$\Theta_t = \frac{1}{K}\sum_{k=1}^{K} s_\eta^{-1}(X_k, \underline{\eta})(s_\eta^{-1}(X_k, \underline{\eta}))^\intercal$$

The diagonal elements of Θ_t should be very close to one if K is sufficiently large. To illustrate this mapping of correlations, we plot in Figure 3.1 the correlation ρ as a function of $\tilde{\rho}$ for the case $\eta = 7$, $\eta_1 = 4$ and $\eta_2 = 10$, obtained by simulating $K = 10^7$ bivariate student-t random variables. If all degrees-of-freedom parameters were the same, then this would be the identity function. Here this function is different, but very close to the identity function, with the difference shown on the right of the Figure 3.1. Hence, the correlations change after transformation but only slightly so. In our empirical application of Section 4, where the marginal degrees of freedom are of a similar range as in the numerical example, we therefore decided to skip this rescaling and set $\Theta_t = \tilde{\Theta}_t$, because it practically makes no difference and alleviates the computational burden.

The sum of the marginal log likelihood functions, $L^V(\tau)$, only depends on the marginal parameters, not on the copula parameters. It is in fact equal to the sum of the univariate volatility log likelihoods that are equal to the Beta-t-EGARCH model log likelihoods. In the first step, the inference function for margins approach proposed by Joe (1997) maximizes $L^V(\tau)$ with respect to τ to obtain $\hat{\tau}$, which is then used in a second step to maximize $L^C(\hat{\tau}^\intercal, \delta, \eta)^\intercal$ with respect to δ and η, to give $\hat{\delta}$ and $\hat{\eta}$. Then, $\hat{\vartheta} = (\hat{\tau}^\intercal, \hat{d}^\intercal)^\intercal$, where $d = (\delta^\intercal, \eta)^\intercal$ is the parameter characterizing the copula, and $\hat{d} = (\hat{\delta}^\intercal, \hat{\eta})^\intercal$.

We now have the following result.

Theorem 3. *Under Assumptions (A1')-(A7),* $\hat{\vartheta} \to_p \vartheta$ *and*

$$\sqrt{T}(\hat{\vartheta} - \vartheta) \to_d N(0, (-F)^{-1}G(-F)^{-1})$$

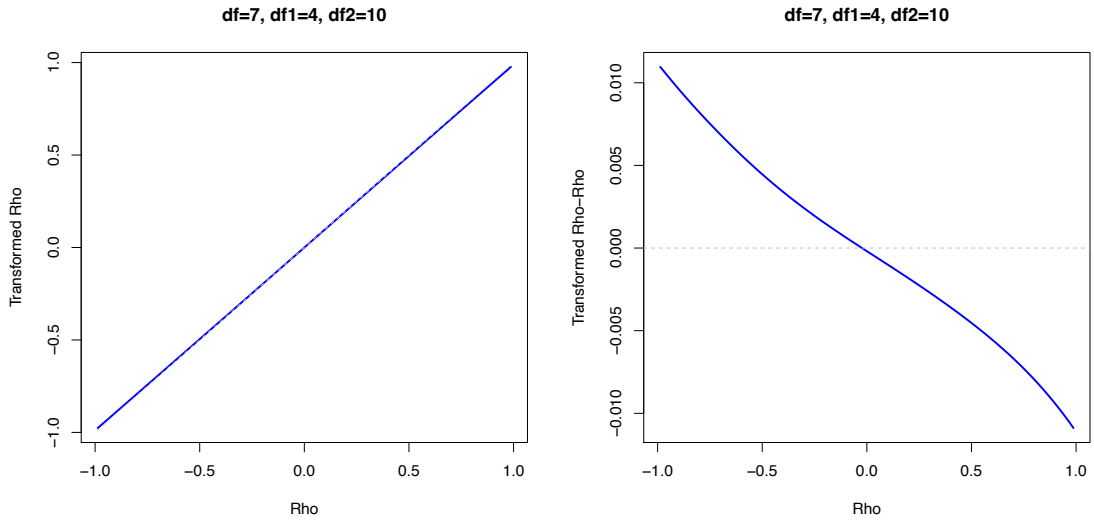

Figure 3.1: *The left panel shows the correlation ρ as a function of $\tilde{\rho}$, where $\tilde{\rho}$ is the correlation parameter of a bivariate standardized student-t r.v. X with $\eta = 7$ d.o.f., and ρ is the correlation of transformed r.v. $s_\eta^{-1}(X_1, \eta_1)$ and $s_\eta^{-1}(X_2, \eta_2)$ with $\eta_1 = 4$ and $\eta_2 = 10$. The right panel shows the difference of this function from the identity function.*

where

$$-F = \begin{pmatrix} \mathcal{J}_{11} & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \mathcal{J}_{NN} & 0 \\ \mathcal{I}_{d1} & \cdots & \mathcal{I}_{dN} & \mathcal{I}_{dd} \end{pmatrix}, \quad G = \begin{pmatrix} \mathcal{J}_{11} & \cdots & \mathcal{J}_{1N} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \mathcal{J}_{N1} & \cdots & \mathcal{J}_{NN} & 0 \\ 0 & \cdots & 0 & \mathcal{I}_{dd} \end{pmatrix}$$

and where

$$\begin{aligned} \mathcal{J}_{ij} &= \mathbb{E}\left[\frac{\partial l_i^V(\tau_i)}{\partial \tau_j}\frac{\partial l_j^V(\tau)}{\partial \tau^\intercal}\right], \quad i,j = 1,\ldots,N \\ \mathcal{I}_{dd} &= -\mathbb{E}\left[\frac{\partial^2 l(\vartheta)}{\partial d \partial d^\intercal}\right], \qquad (3.19) \\ \mathcal{I}_{di} &= -\mathbb{E}\left[\frac{\partial^2 l(\vartheta)}{\partial d \partial \tau_i^\intercal}\right], \quad i = 1,\ldots,N \qquad (3.20) \end{aligned}$$

Note that the matrices $J_{ii}, i = 1,\ldots,N$, are the information matrices of the marginal

likelihoods, so well-known results of univariate Beta-t-EGARCH models can be used for their calculation, see e.g. Proposition 20 of Harvey (2013).

Analytical expressions for the components of $\mathcal{I}$ in (3.19) and (3.20) are given in Appendix E. They can be used, as for one-step estimation, to estimate consistently the asymptotic covariance matrix by evaluating it at $\hat{\vartheta}$ and replacing the expectation operators by sample averages. We will use this two-step estimation approach in our empirical application, to which we turn in the following.

4 Empirical Application

Modelling the day-to-day changes in the correlation structure of financial assets plays an important role in the financial decision making process. We take the standpoint of a representative investor who, in a context of near-zero interest rates, considers not only classical assets such as stocks and bonds, but also commodities (oil), safe-haven investments (gold), and alternative investments (cryptocurrencies). For each market we take a representative index and then model the set of market indices using our multivariate framework. The rest of this section is organized as follows: Section 4.1 presents the data description and sample coverage for our in-sample and out-of-sample analyses; Section 4.2 covers the results for the estimation of the conditional volatility dynamics; Section 4.3 covers the results for the estimation of the conditional correlation dynamics. We conduct an in-depth benchmarking of our proposed model with alternative approaches introduced in the literature both for the in-sample and out-of-sample exercises.

4.1 Data Description

We use historical data on daily log-returns for six representative indices of different asset classes, including both traditional financial assets (i.e. stocks and bonds) as well as commodities and alternative investments (i.e. oil, gold and cryptocurrencies). The sample period spans from August 1, 2014 until December 11, 2020, with the starting date determined by the first available data point of the CRIX cryptocurrency index, see Trimborn

and Härdle (2018) for details on the construction of CRIX. The symbols and description for the indices are shown in Table G.1.

The time series of prices and returns are shown in Figure B.1 and Figure G.3, respectively. We divide the whole sample series into the in-sample period from 2014 to the end of 2018 and the out-of-sample period from 2019 to 2020. We notice that the CRIX index exhibited bubble-like features in 2017 and possibly end of 2020. Table G.2 presents the main descriptive statistics for our historical return series. Alternative investments generally have larger mean returns than stocks or bonds, with higher associated volatilities. The equity index EMU has much larger skewness and kurtosis than the SPX index, which are primarily driven by a single abnormal return in 2016 and by more negative returns during the period of the COVID-19 crisis. Equity and bond returns are generally negatively skewed, corresponding to a long tail in negative returns, whereas the oil index is positively skewed. In addition, the return series for the gold index exhibits approximate symmetry and the CRIX index has a significantly negative skewness, which may be explained by the market downturn of 2018. Furthermore, normality is clearly rejected for all assets by classical tests at conventional significance levels.

We report the sample correlation matrix of historical returns for the six market indices in Table G.3. For financial indices, SPX and EMU are highly correlated with a Pearson correlation coefficient of 0.58. The negative or low coefficients between pairs of SPX-SBW (-0.14) and SPX-GOLD (0.08) confirm the common view that bonds and gold often serve as an effective hedge against stock market risk. This characteristic also results in the fact that the correlation between gold and bonds is relatively high with a coefficient of 0.55. Oil is more closely correlated with the equity indices (0.28 with SPX and 0.25 with EMU), while correlation with other assets is negligible.

4.2 Volatility estimation results

We follow the two-step estimation procedure outlined in Section 3.2, where the first step consists in fitting volatility models to the return series. For each series, we fit a Beta-

t-EGARCH(1,1) model and the model is estimated by maximum likelihood under the assumption of a conditional student-t distribution. In order to improve the convergence of the optimization algorithm, we estimate the inverse of the degrees of freedom parameter, i.e. $\bar{\eta} = 1/\eta$. The results are shown in Table G.4. All parameter estimates are significantly different from zero using estimated asymptotic standard errors. Moreover, the estimates for the autoregressive coefficient ϕ_i indicate high persistence in volatility for all time series and thus confirm the presence of volatility clustering. Note that the CRIX index has the lowest estimated d.o.f. $\eta_i = 1/\bar{\eta}_i$ which corresponds to its observed high kurtosis and thick tails. In Figure G.4, we plot the estimated volatility series estimated using the Beta-t-EGARCH model. We observe high persistence in the estimated volatility series, which is consistent with the estimates for the autoregressive coefficient being close to unity.

We now focus on assessing the validity of the assumptions for the Beta-t-EGARCH model in terms of the residuals distribution and in terms of the score series. Firstly, we test whether the score series is a martingale difference sequence (MDS) using a Box-Pierce test for autocorrelation. The p-values are displayed in Table G.5. Since the p-values are larger than 5% for all six indices, we fail to reject the null hypothesis that the score series have zero autocorrelation up to a lag of 30. This result provides support to the validity of the martingale difference property. Furthermore, we apply the Box-Pierce test to the standardized residuals and their squares. The results, also reported in Table G.5, indicate that significant autocorrelation exists only for the residuals of the CRIX index, whereas the squared residuals for all assets are not serially correlated at the 5% significance level.

The descriptive statistics of standardized residuals are summarized in Table G.6. As expected, the sample means are close to zero and the standard deviations close to one. The only exception is for the CRIX index whose residuals have a standard deviation of 0.78. Again, this may be caused by some remaining serial correlation in the residuals and/or outlier observations. There is remaining excess kurtosis in the standardized residual series, which implies that a conditionally leptokurtic distribution might provide a better fit to the data than the standard normal distribution.

Finally, we use the probability integral transformation (PIT) to check how well the

assumed conditional distribution fits the data. In Figure G.5, we display the histograms of PITs of residual series, assuming respectively a student-t distribution in the left-hand column and a normal distribution in the right-hand column. Most of the histograms for the normal distribution have a hump in the middle and pronounced wings in the tails, which corresponds to our finding of leptokurtic standardized residuals in Table G.6. Although not perfect, the histograms of the PITs for the student-t distribution are much closer to a uniform distribution than the Gaussian PITs. Introducing skewness may be beneficial in improving the fit of the conditional distribution of the data. However, we did not pursue this avenue in our analysis and leave it as a potential extension for future research.

In the next section, we move on to the second step of the two-step estimation and focus on the correlation part of the model, taking the standardized residual series of the univariate volatility models as inputs for the dynamic correlation model.

4.3 Results for the conditional correlations

We now turn to the results for the second step estimation regarding the conditional correlation matrix. First, we want to test the reliability of our assumption of diagonal K and Φ matrices. Estimating a model with six assets and full matrices K and Φ is computationally infeasible as it would entail the estimation of 466 parameters. Instead, we chose five subsets composed of three out of six assets with relatively high correlations and perform likelihood-ratio tests to assess the goodness of fit for the diagonal case (H_0) with respect to these alternative specifications. We fail to reject the null hypothesis H_0 at the 5% significance level for the five specifications considered. This result corroborates our choice of diagonal parameter matrices for the analysis.[1]

The estimated parameters for the models with diagonal K and Φ are reported in Table G.7. All estimated κ_i coefficients are positive and close to zero, while most of the estimated ϕ_i are close to one. Similar to the volatility case, the correlation series

[1]Results for the specification analysis are available from the authors upon request.

are characterized by stationary dynamics but high persistence. Furthermore, we notice that the inverse of the degrees of freedom parameter is approximately 0.029. This result indicates that the t-copula modelling the residual series after de-volatilization is close to, but significantly different from, the Gaussian copula.

Although diagonality constraints of the parameter matrices highly reduce the complexity of the model, we still have 46 parameters to estimate. To further simplify the model, we consider five additional model restrictions. We notice from the results in Table G.7 that the parameter estimates for ω, K, and Φ have quite similar values. This observation motivates alternative specifications in which parameters are restricted to be equal. Table G.8 provides a summary of the different combinations of restrictions considered with the corresponding number of parameters. In the interest of space, we do not provide the estimates of the restricted models and only provide a summary of model performance measures in the following paragraph.

In terms of model diagnostics, we first apply Portmanteau tests to the correlation score series, the correlation residual series, and the outer product of the correlation residual series. The results are reported in Table B.1. For all the log correlation series, there is no evidence against the martingale difference property for the score series at conventional significance levels. The p-values of the test for the correlation residual series indicate that we cannot reject the null hypothesis of no serial autocorrelation at the 1% significance level, whereas we do not find evidence supporting the presence of serial autocorrelation in the outer product of the correlation residual series at all conventional significance levels. Furthermore, we perform likelihood-ratio tests of the restricted models with respect to the diagonal one. It can be observed that we reject the null hypothesis of equal likelihood between the restricted models and Model 1 at the 5% significance level. We reach the same conclusion when considering the 1% significance level except in the case of Model 3. We additionally report the Akaike and Bayesian information criteria (AIC and BIC), to quantify the model performance in terms of goodness-of-fit and model complexity. We conclude that model 2 with 18 parameters should be selected according to the minimum BIC and AIC criteria. In this case, the evidence is in favor of restricting the diagonal

elements of K and those of Φ to be equal, which reduces the parameter dimension substantially, while allowing variation for ω across different series. This result also suggests that the dynamics of the log correlation series θ_i can be assumed to be the same across different components, and that the only difference occurs in the level of the series. In higher dimensions, we conjecture that similar parameter restrictions are reasonable and allow us to obtain computationally feasible results in cases of several hundreds of assets. This actually corresponds to a similar observation made by Caporin and McAleer (2012) for the classes of BEKK and DCC models, for which the "scalar" versions of the models are preferred in high dimensions to avoid the curse of dimensionality.

4.3.1 In sample performance

In this part, we conduct an empirical study for model performance comparison considering the GARCH-cDCC, the GAS class of models proposed by Creal et al. (2011) and our model. We categorize the 8 models into two classes. The Meta-t models on the one hand are estimated using a two-step estimation procedure. In this case, the volatilities are all estimated by a Beta-t-EGARCH model assuming different degrees of freedom for each asset. The dynamic correlations are modelled with different specifications, namely, our log correlation matrix model (Meta-t-log-corr), the cDCC model (Meta-t-cDCC), the GAS model introduced in Creal et al. (2011) (Meta-t-GAS), and its variation using hyperspherical coordinates decomposition (Meta-t-GAS-h). The second group of models, on the other hand, contains the multivariate t models estimated by a one-step procedure. In particular, the t-cDCC, t-GAS, t-GAS with log-volatility (t-GAS-l) specification are estimated using variance and correlation targeting techniques. We also consider the t-GAS model with a hyperspherical coordinates decomposition (t-GAS-hl) for comparison.

Note that all models based on the Meta-t distribution face the problem of rescaling of Θ_t that is discussed in Section 3.2, because for a Meta-t distribution, $\tilde{\Theta}_t$ is not exactly equal to the correlation matrix. However, as we have shown in our example with marginal d.o.f. equal to 4 and 10, and a copula d.o.f. of 7, the approximation is very accurate. The range of marginal d.o.f. is similar for our data. We therefore decided to skip the rescaling

and set $\Theta_t = \tilde{\Theta}_t$, which practically makes no difference but simplifies the computation of all considered Meta-t models.

The parameters estimated from the group of Meta-t models are reported in Table G.9.[2] It can be observed that the autoregressive parameter values Φ for all models are close to unity indicating high persistence in the correlation series. Moreover, the inverse of the degrees of freedom parameter is close to zero (around 0.033) but significantly different from zero. Table G.10 presents the parameter estimation of the group of multivariate t models. For each model, the parameters α_i and β_i with $i = 1, 2, \ldots, 6$ are the volatility parameters while α_7 and β_7 are the correlation parameters with different specifications. Similarly, all the autoregressive parameters are close to one such that the estimated volatility series and correlation series are highly persistent.

The Meta-t log correlation specification has the best model fit in terms of log likelihood values, reported in Table B.2. The differences between the likelihood values of the four Meta-t variations are quite small and would not be significant using likelihood ratio tests, if the models were nested. However, the difference between the Meta-t models and the multivariate student-t models are quite substantial. Therefore, the added flexibility in the Meta-t specification allowing for different degrees of freedom for individual assets outperforms the specification using a multivariate t distribution with a single degrees of freedom parameter by a larger log likelihood of around 80 to 100. Furthermore, since the only difference among the four Meta-t models comes from the parameterization of the correlation matrix, the results indicate that the different specifications of the correlation dynamics after the devolatilization appear to make a smaller difference. This observation is consistent with the empirical results in Creal et al. (2011). Based on the third and fourth rows in Table B.2, the Meta-t log correlation model would be selected using the classical model selection criteria AIC and BIC. The superior performance of the Meta-t models is confirmed by the AIC and BIC results although the multivariate t models have fewer parameters.

Figure B.2 plots the estimated correlation series for four selected pairs of asset indices,

[2]To save space, we only report the K, Φ and $\bar{\eta}$ parameters and leave out the ω vectors.

namely, SPX–SBW, SPX–EMU, SPX–OIL, and SBW–GOLD. We compare the results for our Meta-t-log-corr model and the best performing multivariate t model, i.e. the t-GAS-hl. While the overall level of estimated correlations is about the same for both models, the Meta-t-log-corr model is more volatile and less persistent over time. This is consistent with the fact that the Meta-t-log-corr model has a slightly lower estimated autoregressive parameter than the t-GAS-hl model.

4.3.2 Out of sample performance

To verify the short-term forecasting ability of the dynamic log correlation model, we compare the performance of different approaches in a Global Minimum Variance (GMV) portfolio allocation exercise. To this end, we forecast the 1-day ahead covariance matrices using the better performing models, namely the group of Meta-t models. Our out-of-sample exercise is conducted over a period ranging from the start of 2019 to the end of 2020. Market conditions are rather stable for the 2019 period and we thus only re-estimate the models twice over the year. The year 2020, on the other hand, featured extremely volatile market conditions which culminated at the onset of the COVID-19 crisis in March. We address the potential impact of these exceptional market conditions on our out-of-sample results by re-estimating the models every two months to incorporate new information on these rapidly evolving financial conditions. The 1-day ahead volatility is forecast using the Beta-t-EGARCH model and the correlation matrices are forecast respectively from the Meta-t-log-corr model, the Meta-t-cDCC model, the Meta-t-GAS model, and the Meta-t-GAS-h model. Based on the daily forecasts, we construct the conditional covariance matrix and use it as an input to compute the optimal GMV portfolio allocation for each day.

We present in Table B.3 the out-of-sample performance results in terms of the annualized portfolio standard deviation. We observe that the Meta-t-log-corr model has similar portfolio volatility to the Meta-t-GAS and Meta-t-GAS-h models. Furthermore, all three score-driven models exhibit better performances than the Meta-t-cDCC model. This result highlights the fact that score driven models are more robust to outliers and,

as a consequence, exhibit superior performance under extreme market conditions. Furthermore, we can draw similar conclusions for model comparisons in terms of the Sharpe ratio of the considered portfolios. More specifically, the Sharpe ratios of the dynamic conditional score models are close to each other and higher than the corresponding Sharpe ratio for the cDCC model specification.

5 Conclusions

The proposed score-driven correlation model ensures positivity of the conditional correlation matrix while offering high flexibility in the modelling of the conditional distribution and dynamics. In the empirical application we have shown that the model fits well a set of financial and alternative assets, passing various diagnostic tests for the distribution and dynamics. We also show that our model performs well compared to alternative correlation parameterizations in terms of in-sample fit. Our out-of-sample analysis highlights the superior performance of score driven models compared to the cDCC model in terms of correlation forecasting applied to minimum variance portfolio selection.

Many generalizations are possible, for example using generalized student-t distributions as in Harvey and Lange (2017), and distributions allowing for skewness as in Bauwens and Laurent (2005) or Harvey and Sucarrat (2014). Rather than models based on the student-t distribution, one may also generalize the Gamma-GED-EGARCH to the multivariate case. Future work may also address specification tests in high dimensions, for example clustering of parameters characterizing the dynamics of the log correlation matrix. We have discussed two extreme cases in the empirical example, the unrestricted case and the one setting all parameters equal, but developing strategies to specify intermediate cases such as the factor structure used in Archakov et al. (2020) would seem reasonable as well.

Acknowledgements

We are grateful for discussions with participants at the INET Cambridge workshop on score driven time series models in March 2019. The work was done partially while the first author was visiting the Institute for Mathematical Sciences, National University of Singapore in 2019. The visit was supported by the Institute. Both authors acknowledge financial support of the ARC research contract 18/23-089 of the Belgian Federal Science Policy Office.

Appendices

A Theoretical material

A.1 Lyapounov exponent of the SRE in the bivariate case

In the bivariate case, the SRE is given by (3.13) where $\psi_t(s) = \omega + \phi s + \kappa g(s) Z_t$. Since $g(s)$ is continuously differentiable, we are looking for

$$\Lambda(\psi) = \sup_{s \in \mathbb{R}} |\partial \psi(s) / \partial s|.$$

Straightforward calculations show that $g(s) = \frac{1}{2}(-\rho, 1, 1, -\rho)^\intercal$, with $\rho = \frac{\exp(2s)-1}{\exp(2s)+1}$, and $Z_t = \text{vec}(w_t \xi_t \xi_t^\intercal - I_N)$, with $w_t = \frac{\eta+N}{\eta-2+\xi_t^\intercal \xi_t}$. Hence,

$$\frac{\partial \psi(s)}{\partial s} = \phi - \kappa \frac{\exp(2s)}{(\exp(2s)+1)^2} \left\{ w_t(\xi_{1t}^2 + \xi_{2t}^2) - 2 \right\}$$

and because $\inf_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 0$ and $\sup_{s \in \mathbb{R}} \frac{\exp(2s)}{(\exp(2s)+1)^2} = 1$, we have

$$\Lambda(\psi) = \max \left(\phi,\ \phi - \kappa \left[w_t(\xi_{1t}^2 + \xi_{2t}^2) - 2 \right] \right).$$

Using this expression, we can simulate a large number of i.i.d. ξ_t and estimate the Lyapounov exponent $\mathbb{E}[\log \Lambda(\psi)]$ by the sample mean of $\log \Lambda(\psi)$. For a student-t with 5 and 15 degrees of freedom, Figure G.1 shows the parameter space of (ϕ, κ) for which the Lyapounov exponent is negative. As ϕ is approaching 1, κ is constrained to a shrinking region around zero.

For the general case $N \geq 2$ and $\theta_t \in \mathcal{T}_{N,\epsilon} \subset \mathbb{R}^{N(N-1)/2}$, no simple analytical expression is available for $\Lambda(\psi)$, but Appendix D shows that there exists an upper bound. Furthermore, we can obtain numerical approximations to estimate the Lyapounov exponent for a given set of parameters, see Appendix D for details.

A.2 Proof of Theorem 1

First, we have

$$\begin{aligned} dL(\delta) &= -\frac{1}{2}\sum_{t=1}^{T}\left\{ d\log|\Theta_t| - \frac{\eta+N}{\eta-2+\varepsilon_t^{\intercal}\Theta_t^{-1}\varepsilon_t} Tr\left(\varepsilon_t\varepsilon_t^{\intercal} d\Theta_t^{-1}\right)\right\} \\ &= -\frac{1}{2}\sum_{t=1}^{T} Tr\left\{ \Theta_t^{-1}d\Theta_t - \frac{\eta+N}{\eta-2+\varepsilon_t^{\intercal}\Theta_t^{-1}\varepsilon_t}\varepsilon_t\varepsilon_t^{\intercal}\Theta_t^{-1}d\Theta_t\Theta_t^{-1}\right\} \\ &= \frac{1}{2}\sum_{t=1}^{T} Tr\left\{ \left(\frac{\eta+N}{\eta-2+\varepsilon_t^{\intercal}\Theta_t^{-1}\varepsilon_t}\Theta_t^{-1}\varepsilon_t\varepsilon_t^{\intercal}\Theta_t^{-1} - \Theta_t^{-1}\right) d\Theta_t\right\} \\ &= \frac{1}{2}\sum_{t=1}^{T} \mathrm{vec}\left(\frac{\eta+N}{\eta-2+\varepsilon_t^{\intercal}\Theta_t^{-1}\varepsilon_t}\Theta_t^{-1}\varepsilon_t\varepsilon_t^{\intercal}\Theta_t^{-1} - \Theta_t^{-1}\right)^{\intercal} d\mathrm{vec}\Theta_t \end{aligned}$$

Now, using a result of Linton and McCrorie (1995) for the differential of the matrix exponential, we can write

$$d\mathrm{vec}\Theta_t = A_t d\mathrm{vec}\log\Theta_t,$$

and with the definition of the permutation matrix P,

$$d\mathrm{vec}\Theta_t = A_t P \begin{pmatrix} d\mathrm{diag}(\log\Theta_t) \\ d\mathrm{vecl}(\log\Theta_t) \end{pmatrix}$$

Now, $\mathrm{vecl}(\log\Theta_t) = \theta_t$, and by Archakov and Hansen (2021), $\mathrm{diag}(\log\Theta_t)$ is a unique function of θ_t, for which no analytical form exists for $N > 4$. However, using the proof of Proposition 3 of Archakov and Hansen (2021), we have an analytical expression for the differential of this function, given by

$$d\mathrm{diag}(\log\Theta_t) = -(E_d A_t E_d^{\intercal})^{-1} E_d A_t E_s^{\intercal} d\theta_t$$

which completes the stated result.

The expressions given in Remark 1 for the Gaussian distribution, i.e. for the limiting

case $\eta \to \infty$, can be obtained easily. First, using Stirling's approximation, note that

$$\begin{aligned}
\lim_{\eta\to\infty} \quad & \left[T \log\left(\Gamma\left(\frac{\eta+N}{2}\right)\right) - T\log\left(\Gamma\left(\frac{\eta}{2}\right)\right) - \frac{TN}{2}\log\left((\eta-2)\pi\right)\right] \\
= \; & \lim_{\eta\to\infty} T \log \frac{\Gamma\left(\frac{\eta+N}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)\left((\eta-2)\pi\right)^{N/2}} \\
= \; & T\log\left(2\pi\right)^{-N/2} = -\frac{TN}{2}\log(2\pi)
\end{aligned}$$

And for the score function, it can be easily shown that

$$\begin{aligned}
p\lim_{\eta\to\infty} \quad & -\frac{1}{2}\sum_{t=1}^{T}\left\{\log|\Theta_t| + (\eta+N)\log\left(1+\frac{1}{\eta-2}\varepsilon_t^{\intercal}{\Theta_t}^{-1}\varepsilon_t\right)\right\} \\
= \; & p\lim_{\eta\to\infty} -\frac{1}{2}\sum_{t=1}^{T}\left\{\log|\Theta_t| + \frac{\eta+N}{\eta-2}\varepsilon_t^{\intercal}{\Theta_t}^{-1}\varepsilon_t\right\} \\
= \; & -\frac{1}{2}\sum_{t=1}^{T}\left\{\log|\Theta_t| + \varepsilon_t^{\intercal}{\Theta_t}^{-1}\varepsilon_t\right\}
\end{aligned}$$

Thus, we arrive at the expression of the log likelihood function for the multivariate normal distribution. Furthermore, the score can be obtained directly by taking the probability limit of $\frac{\eta+N}{\eta-2+\varepsilon_t^{\intercal}\Theta_t^{-1}\varepsilon_t}$ as $\eta \to \infty$, which is equal to 1.

A.3 Proof of Theorem 2

We first need to show that, under our conditions, θ_t and its derivatives up to order 3, evaluated at the true parameters, are stationary and ergodic processes. The process θ_t satisfies the stochastic recurrence equation (SRE) in (3.13) which by Theorem 2.8 of Straumann and Mikosch (2006) is stationary and ergodic under the contraction condition of Assumption (A6). The compactness of Ω then implies consistency by Lemma 1 of Jensen and Rahbek (2004).

The process $\partial\theta_t/\partial\delta$ permits an SRE as in (3.17). It is stationary and ergodic if $\mathbb{E}[\log|x_t|] < 0$, see Straumann and Mikosch (2006). This is implied by Assumption (A5), i.e. $\mathbb{E}[x_t^2] < 1$, using Jensen's inequality.

The process $\frac{\partial^2\theta_t}{\partial\delta\partial\delta^\intercal}$ also permits an SRE:

$$\frac{\partial^2\theta_t}{\partial\delta\partial\delta^\intercal} = C_{t-1} + x_{t-1}\frac{\partial^2\theta_{t-1}}{\partial\delta\partial\delta^\intercal}$$

where $C_t = \frac{\partial z_t}{\partial\delta^\intercal} + \frac{\partial\theta_t}{\partial\delta}\frac{\partial x_t}{\partial\delta^\intercal}$. Again, stationary and ergodicity are implied by Assumption (A5). The same argument applies to the third derivatives of θ_t.

We then obtain asymptotic normality for the score evaluated at the true parameter, invoking the central limit theorem for martingale difference sequences of Brown (1971), i.e.

$$T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_t(\vartheta_0)}{\partial\vartheta} \to_d N(0, I(\vartheta_0)).$$

This requires that $\mathbb{E}[(\partial\theta_{it}/\partial\vartheta_j)^2] < \infty$ for $i = 1,\dots,N^*$ and $j = 1\dots,\dim(\vartheta)$, which is implied by Assumption (A5), see Theorem 1 of Harvey (2013).

Moreover, we have

$$-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\vartheta_0)}{\partial\vartheta\partial\vartheta^\intercal} \to_p I(\vartheta_0)$$

as in Proposition 5 of Harvey (2013). It suffices to check that, for a neighborhood $N(\vartheta_0)$ of the true parameter,

$$\mathbb{E}\sup_{\vartheta\in N(\vartheta_0)}\left\|\frac{\partial\text{vec}}{\partial\vartheta'}\frac{\partial^2 l_t(\vartheta)}{\partial\vartheta\partial\vartheta'}\right\| < \infty. \tag{A.1}$$

For the volatility part this has been shown before, see e.g. Theorem 3 of Harvey (2013), so we concentrate on the correlation part. We have

$$\begin{aligned}\frac{\partial\text{vec}}{\partial\delta^\intercal}\frac{\partial^2 l_t(\delta)}{\partial\delta\partial\delta^\intercal} &= (I_m\otimes u_t^\intercal\otimes I_m)\frac{\partial\text{vec}}{\partial\delta^\intercal}\frac{\partial\text{vec}}{\partial\delta^\intercal}\frac{\partial\theta_t^\intercal}{\partial\delta} + (\frac{\partial\text{vec}^\intercal}{\partial\delta}\frac{\partial\theta_t^\intercal}{\partial\delta}\otimes I_m)(I_{N^*}\otimes\text{vec}(I_m))\frac{\partial u_t}{\partial\delta^\intercal}\\ &+ (\frac{\partial u_t^\intercal}{\partial\delta}\otimes I_m)\frac{\partial\text{vec}}{\partial\delta^\intercal}\frac{\partial\theta_t^\intercal}{\partial\delta} + (I_m\otimes\frac{\partial\theta_t^\intercal}{\partial\delta})\frac{\partial\text{vec}}{\partial\delta^\intercal}\frac{\partial u_t}{\partial\delta^\intercal}\end{aligned} \tag{A.2}$$

with $m = \dim(\delta)$. The first term on the right hand side of (A.2) depends on the third order derivatives of the correlation parameter with respect to δ. A typical element of this matrix is given by

$$\begin{aligned}\frac{\partial^3\theta_t(\delta)}{\partial\delta_i\partial\delta_j\partial\delta_k} &= \frac{\partial^2 z_{i,t-1}}{\partial\theta_j\partial\theta_k} + \frac{\partial x_{t-1}}{\partial\delta_j}\frac{\partial^2\theta_{t-1}}{\partial\delta_i\partial\delta_k} + \frac{\partial^2 x_{t-1}}{\partial\delta_j\partial\delta_k}\frac{\partial\theta_{t-1}}{\partial\delta_i} + \frac{\partial x_{t-1}}{\partial\delta_k}\frac{\partial^2\theta_{t-1}}{\partial\delta_i\partial\delta_j} + x_{t-1}\frac{\partial^3\tilde{\theta}_{t-1}(\delta)}{\partial\delta_i\partial\delta_j\partial\delta_k}\\ &= \sum_{j=1}^{\infty}\zeta_{t-j}\prod_{i=1}^{j-1}x_{t-i}\end{aligned} \tag{A.3}$$

where z_{it} denotes the i-th column of z_t, and $\zeta_t = \frac{\partial^2 z_{i,t-1}}{\partial\theta_j\partial\theta_k} + \frac{\partial x_{t-1}}{\partial\delta_j}\frac{\partial^2\theta_{t-1}}{\partial\delta_i\partial\delta_k} + \frac{\partial^2 x_{t-1}}{\partial\delta_j\partial\delta_k}\frac{\partial\theta_{t-1}}{\partial\delta_i} + \frac{\partial x_{t-1}}{\partial\delta_k}\frac{\partial^2\theta_{t-1}}{\partial\delta_i\partial\delta_j}$. Each term of ζ_t can be uniformly bounded in a neighborhood of δ_0 and a repeated application of the c_r inequality yields $\mathbb{E}\sup_{\delta\in N(\delta_0)}|\zeta_t| < \infty$. Note that $|x_t|$ is a stationary ergodic sequence of nonnegative random variables with $\mathbb{E}\log|x_t| < 0$, which is implied by Assumption (A5). Thus, by Lemma 2.4 of Straumann and Mikosch (2006), $\prod_{i=1}^{t-1}|x_{t-i}| \to_{e.a.s.} 0$, meaning that there exists a $\gamma > 1$ s.t. $\gamma^t\prod_{i=1}^{t-1}|x_{t-i}| \to_{a.s.} 0$. This implies that $\mathbb{E}\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| = O(\rho^t)$, which in turn implies that the infinite sum in (A.3) is convergent and we obtain

$$\mathbb{E}\sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3\theta_t(\delta)}{\partial\delta_i\partial\delta_j\partial\delta_k}\right| < \infty.$$

The remaining terms on the right hand side of (A.2) can be bounded in a similar way. Together with corresponding results for the volatility part, this implies (A.1).

Finally, tedious but straightforward calculations yield the expression of the information matrix using results of Fiorentini et al. (2003). The details are omitted here but available upon request.

A.4 Proof of Theorem 3

The proof of consistency and asymptotic normality is essentially the same as that of Theorem 2, noting that stationarity and ergodicity of the process $\tilde{\theta}_t$ hold under the same conditions as for the process θ_t. The form of the asymptotic covariance matrix of the inference function for margins approach is based on Joe (2005), and the components of $\mathcal{I}$ follow by straightforward calculations, which are again omitted to economize on space, but available upon request.

B Tables and Figures

Table B.1: Diagnostics and model selection criteria for multivariate models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
k	46	18	32	31	4	3
T	845	845	845	845	845	845
score test	0.5014	0.3598	0.3228	0.3910	0.4150	0.3134
ξ_t	0.0653	0.0489	0.0818	0.0810	0.0513	0.0461
$vech(\xi_t\xi_t^\intercal)$	0.9457	0.9463	0.9526	0.9723	0.9585	0.9539
$\ln \mathcal{L}(\hat{\theta})$	**446.59**	421.01	432.77	424.91	368.12	361.92
LR		51.17	27.65	43.37	156.95	169.35
d.o.f.		28	14	15	42	43
p-value		0.0048	0.0158	0.0001	0.0000	0.0000
BIC	-583.18	**-720.70**	-649.88	-640.90	-709.28	-703.62
AIC	-801.19	**-806.01**	-801.54	-787.82	-728.23	-717.83

Note: k is the number of parameters in the model. T is the number of data points in the sample. Score test, ξ_t and $vech(\xi_t\xi_t^\intercal)$ are the p-values of a multivariate Portmanteau test applied to the correlation score series, the correlation residual series and the outer product of the correlation residual series. The number of lags is 30 and degrees of freedom are $N_{\text{lag}} \times N^{*2}$ with $N_{\text{lag}} = 30$ and $N^* = 15$. The likelihood-ratio statistic is computed as $LR = -2 \ln \frac{\mathcal{L}_i(\hat{\theta})}{\mathcal{L}_1(\hat{\theta})}$, with $\mathcal{L}_i$ corresponding to the likelihood for model i. The number of degrees of freedom for the LR test is computed as the difference in the number of parameters between Model 1 and Model i. The Akaike and Bayesian information criteria are respectively computed as $AIC = -2 \ln \mathcal{L}(\hat{\theta}) + 2k$, and $BIC = -2 \ln \mathcal{L}(\hat{\theta}) + k \ln T$.

Table B.2: Modelling comparison for dynamic correlation matrix

	Meta-t-log-corr	Meta-t-cDCC	Meta-t-GAS	Meta-t-GAS-h
k	42	42	42	42
$\ln \mathcal{L}(\hat{\theta})$	**16330.63**	16329.34	16327.67	16329.73
AIC	**-32577.26**	-32574.69	-32571.33	-32575.47
BIC	**-32378.21**	-32375.64	-32372.28	-32376.41
	t-cDCC	t-GAS	t-GAS-l	t-GAS-hl
k	36	36	36	36
$\ln \mathcal{L}(\hat{\theta})$	16244.69	16230.82	16231.58	16247.45
AIC	-32417.39	-32389.63	-32391.15	-32422.89
BIC	-32246.77	-32219.02	-32220.54	-32252.28

Note: k is the number of parameters. The Akaike and Bayesian information criteria are respectively computed as $AIC = -2\ln \mathcal{L}(\hat{\theta}) + 2k$, and $BIC = -2\ln \mathcal{L}(\hat{\theta}) + k \ln T$. The first four models are estimated following a two-step estimation procedure with dynamic volatility estimated using the Beta-t-EGARCH model and the dynamic correlation matrix modelled respectively by the Meta-t-log-corr, Meta-t-cDCC, Meta-t-GAS, and Meta-t-GAS-h models. The last four models are estimated following a one-step estimation procedure with dynamic volatility and correlation matrix modelled respectively by t-cDCC, t-GAS, t-GAS-l and t-GAS-hl.

Table B.3: Out-of-sample annualized portfolio standard deviation (in percentage points) and Sharpe ratio

Meta-t-log-corr	Meta-t-cDCC	Meta-t-GAS	Meta-t-GAS-h
Annualized standard deviation			
5.8203	6.0859	5.8416	5.8112
Sharpe ratio			
0.9403	0.6730	0.9591	0.9409

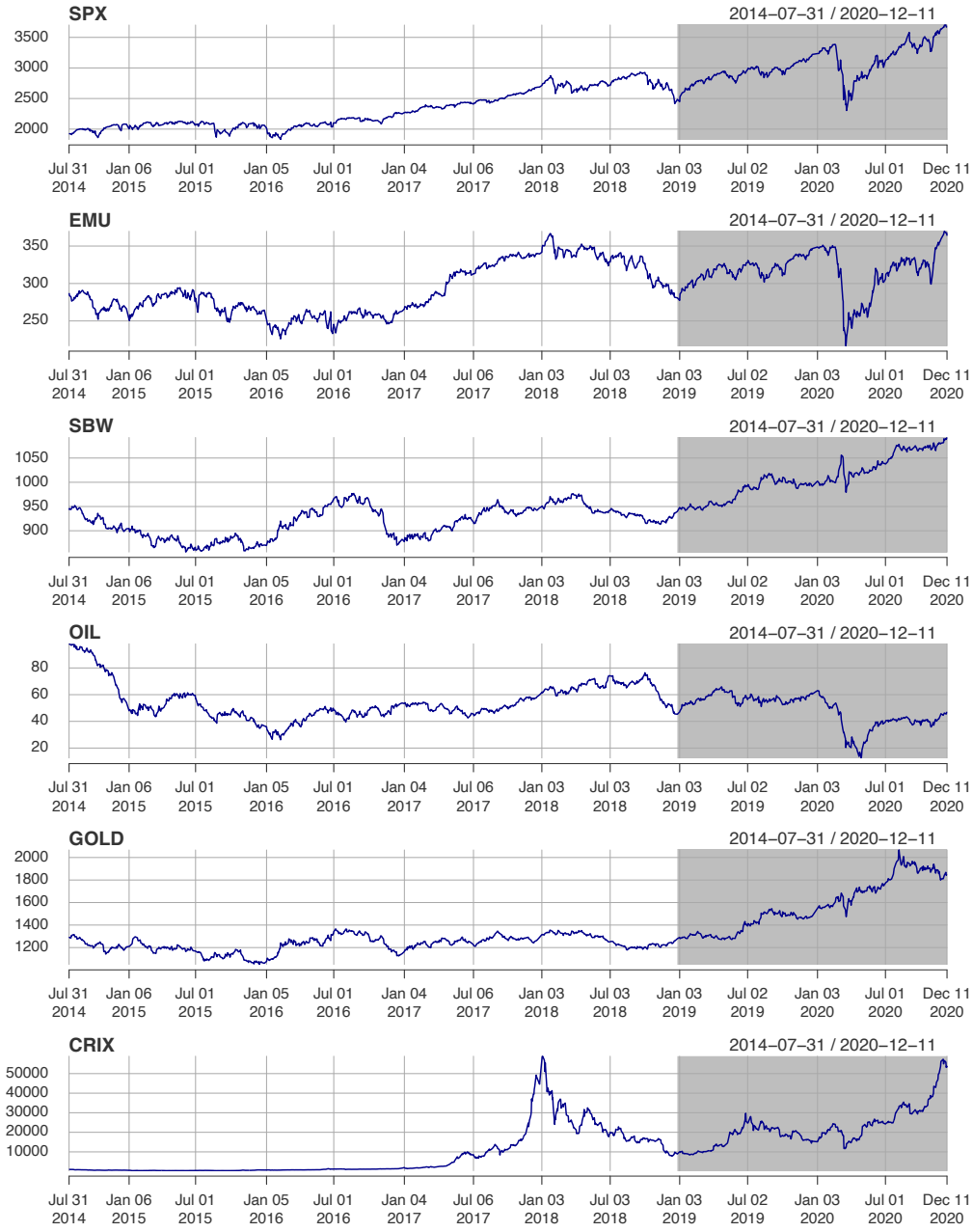

Figure B.1: Time series of daily asset prices, August 1, 2014 to December 12, 2020, for (from top to bottom) SPX, SBW, EMU, OIL, GOLD, and CRIX. The grey shaded area is the out-of-sample period ranging from the beginning of 2019 to the end of 2020.

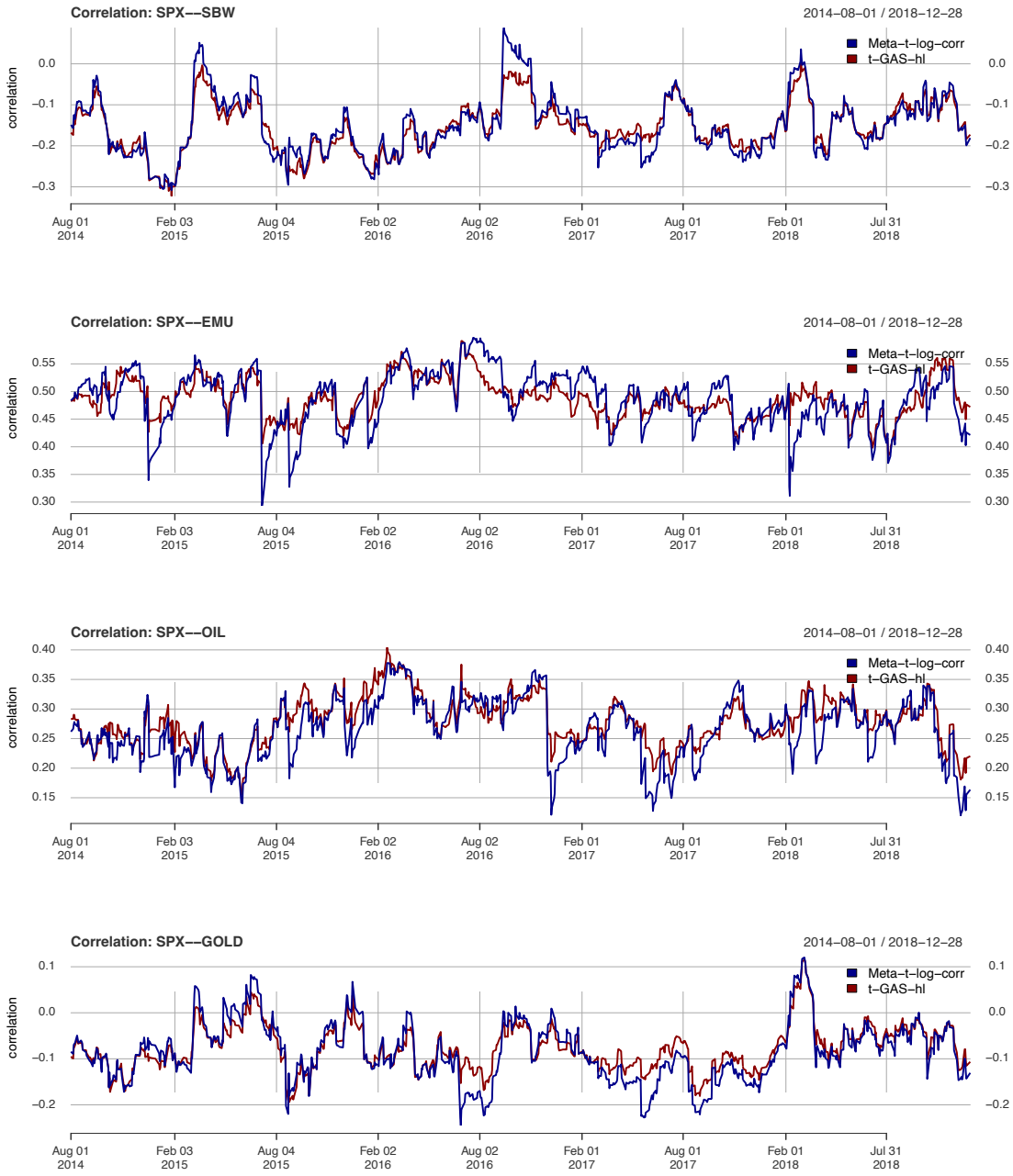

Figure B.2: Estimated conditional correlation series for four selected pairs (from top to bottom): SPX-SBW, SPX-EMU, SPX-OIL, SPX-GOLD, using two models: the Meta-t-log-corr model and the t-GAS-hl model.

References

Archakov, I. and Hansen, P. R. (2020): "A canonical representation of block matrices with applications to covariance and correlation matrices". *arXiv preprint arXiv:2012.02698*.

Archakov, I. and Hansen, P. R. (2021): "A new parametrization of correlation matrices". *Econometrica* 89.4, pp. 1699–1715.

Archakov, I., Hansen, P. R. and Lunde, A. (2020): "A multivariate realized GARCH model". *arXiv preprint arXiv:2012.02708*.

Bauwens, L., Hafner, C. M. and Laurent, S. (2012): *Handbook of volatility models and their applications*. Vol. 3. John Wiley & Sons.

Bauwens, L. and Laurent, S. (2005): "A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models". *Journal of Business & Economic Statistics* 23.3, pp. 346–354.

Bauwens, L., Laurent, S. and Rombouts, J. V. (2006): "Multivariate GARCH models: A survey". *Journal of Applied Econometrics* 21.1, pp. 79–109.

Brown, B. M. (1971): "Martingale central limit theorems". *The Annals of Mathematical Statistics* 42.1, pp. 59–66.

Caporin, M. and McAleer, M. (2012): "Do we really need both BEKK and DCC? A tale of two multivariate GARCH models". *Journal of Economic Surveys* 26.4, pp. 736–751.

Chang, J., Qiu, Y., Yao, Q. and Zou, T. (2018): "Confidence regions for entries of a large precision matrix". *Journal of Econometrics* 206.1, pp. 57 –82.

Creal, D., Koopman, S. J. and Lucas, A. (2011): "A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations". *Journal of Business & Economic Statistics* 29.4, pp. 552–563.

Creal, D., Koopman, S. J. and Lucas, A. (2013): "Generalized autoregressive score models with applications". *Journal of Applied Econometrics* 28.5, pp. 777–795.

Demarta, S. and McNeil, A. J. (2005): "The t copula and related copulas". *International Statistical Review* 73.1, pp. 111–129.

Engle, R. (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". *Journal of Business & Economic Statistics* 20.3, pp. 339–350.

Fan, J., Liao, Y. and Mincheva, M. (2011): "High dimensional covariance matrix estimation in approximate factor models". *The Annals of Statistics* 39.6, 3320–3356.

Fiorentini, G., Sentana, E. and Calzolari, G. (2003): "Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with Student t innovations". *Journal of Business & Economic Statistics* 21.4, pp. 532–546.

Francq, C. and Zakoian, J.-M. (2019): *GARCH models: Structure, statistical inference and financial applications*. John Wiley & Sons.

Gorgi, P, Hansen, P., Janus, P and Koopman, S. (2018): "Realized Wishart-GARCH: A score-driven multi-asset volatility model". *Journal of Financial Econometrics* 17.1, pp. 1–32.

Hafner, C. M., Linton, O. B. and Tang, H. (2020): "Estimation of a multiplicative correlation structure in the large dimensional case". *Journal of Econometrics* 217.2, pp. 431–470.

Hansen, P. R., Lunde, A. and Voev, V. (2014): "Realized beta GARCH: A multivariate GARCH model with realized measures of volatility". *Journal of Applied Econometrics* 29.5, pp. 774–799.

Harvey, A. and Lange, R.-J. (2017): "Volatility modeling with a generalized t distribution". *Journal of Time Series Analysis* 38.2, pp. 175–190.

Harvey, A. and Sucarrat, G. (2014): "EGARCH models with fat tails, skewness and leverage". *Computational Statistics & Data Analysis* 76, pp. 320–338.

Harvey, A. C. (2013): *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press.

Jensen, S. T. and Rahbek, A. (2004): "Asymptotic inference for nonstationary GARCH". *Econometric Theory* 20.6, pp. 1203–1226.

Joe, H. (1997): *Multivariate models and multivariate dependence concepts*. Chapman & Hall.

Joe, H. (2005): "Asymptotic efficiency of the two-stage estimation method for copula-based models". *Journal of Multivariate Analysis* 94, pp. 401–419.

Kawakatsu, H. (2006): "Matrix exponential GARCH". *Journal of Econometrics* 134.1, pp. 95–128.

Nelson, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach". *Econometrica* 59.2, pp. 347–370.

Opschoor, A., Janus, P., Lucas, A. and Van Dijk, D. (2018): "New HEAVY models for fat-tailed realized covariances and returns". *Journal of Business & Economic Statistics* 36.4, pp. 643–657.

Straumann, D. and Mikosch, T. (2006): "Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach". *The Annals of Statistics* 34.5, pp. 2449–2495.

Trimborn, S. and Härdle, W. K. (2018): "CRIX an Index for cryptocurrencies". *Journal of Empirical Finance* 49, pp. 107–122.

Vassallo, D., Buccheri, G. and Corsi, F. (2021): "A DCC-type approach for realized covariance modeling with score-driven dynamics". *International Journal of Forecasting* 37.2, pp. 569–586.

Supplementary Material for "A dynamic conditional score model for the log correlation matrix"

Christian M. Hafner* Linqi Wang†

February 23, 2022

Appendices

C Example: The bivariate Gaussian case

This section shows that our model reduces to that of Harvey (2013) (Section 7.3.1) in the bivariate Gaussian case. It also serves to illustrate the structure of the involved matrix expressions. Let $N = 2$. For Θ and $\log \Theta$, see Example 1 above. We have $A_t = (\Gamma_t \otimes \Gamma_t)\Xi_t(\Gamma_t \otimes \Gamma_t)^\intercal$ with $\text{diag}(\Xi_t) = (1+\rho, 2\rho/\log\frac{1+\rho}{1-\rho}, 2\rho/\log\frac{1+\rho}{1-\rho}, 1-\rho)$ and

$$\Gamma_t \otimes \Gamma_t = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

*ISBA and Louvain Institute of Data Analysis and Modelling, UCLouvain, christian.hafner@uclouvain.be

†LFIN and Louvain Institute of Data Analysis and Modelling, UCLouvain, linqi.wang@uclouvain.be.

which gives

$$A_t = \frac{1}{2}\begin{pmatrix} 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \rho & \rho & 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \\ \rho & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \rho \\ \rho & 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \rho \\ 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \rho & \rho & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \end{pmatrix}$$

The elimination matrices are $E_l = (0,1,0,0)$, $E_u = (0,0,1,0)$, and

$$E_d = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

such that

$$E_d A_t E_d^\intercal = \frac{1}{2}\begin{pmatrix} 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \\ 1-\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \end{pmatrix}, \quad E_d A_t E_s^\intercal = \begin{pmatrix} \rho \\ \rho \end{pmatrix}$$

$$(E_d A_t E_d^\intercal)^{-1} = \frac{\log\frac{1+\rho}{1-\rho}}{4\rho}\begin{pmatrix} 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \frac{2\rho}{\log\frac{1+\rho}{1-\rho}}-1 \\ \frac{2\rho}{\log\frac{1+\rho}{1-\rho}}-1 & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \end{pmatrix}$$

$$-(E_d A_t E_d^\intercal)^{-1} E_d A_t E_s^\intercal = -\frac{\log\frac{1+\rho}{1-\rho}}{4\rho}\begin{pmatrix} 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} & \frac{2\rho}{\log\frac{1+\rho}{1-\rho}}-1 \\ \frac{2\rho}{\log\frac{1+\rho}{1-\rho}}-1 & 1+\frac{2\rho}{\log\frac{1+\rho}{1-\rho}} \end{pmatrix}\begin{pmatrix} \rho \\ \rho \end{pmatrix} = \begin{pmatrix} -\rho \\ -\rho \end{pmatrix}$$

Note that this can be obtained directly from Example 1, where the diagonal elements of $\log\Theta$ are $\frac{1}{2}\log(1-\rho^2)$, and the off-diagonal element is $\gamma := \frac{1}{2}\log\frac{1+\rho}{1-\rho}$, or $\rho = \frac{\exp(2\gamma)-1}{\exp(2\gamma)+1}$. Hence,

$$\frac{\partial}{\partial\rho}\frac{1}{2}\log(1-\rho^2)\frac{\partial\rho}{\partial\gamma} = \frac{-\rho}{1-\rho^2}\frac{4\exp(2\gamma)}{(\exp(2\gamma)+1)^2} = \frac{-2\rho}{1-\rho^2}(1-\rho^2) = -2\rho.$$

Together with the factor $1/2$ in u_t, we obtain the same result.

The permutation matrix P is given by

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Then,

$$\begin{pmatrix} -(E_d A_t E_d^\intercal)^{-1} E_d A_t E_s^\intercal \\ I_{N(N-1)/2} \end{pmatrix}^\intercal P^\intercal = \begin{pmatrix} -\rho & -\rho & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix}$$

and

$$\begin{aligned} \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix} A_t &= \frac{1}{2} \begin{pmatrix} -\rho & 1 & 1 & -\rho \end{pmatrix} \begin{pmatrix} 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \\ \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ \rho & 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho \\ 1 - \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} & \rho & \rho & 1 + \frac{2\rho}{\log \frac{1+\rho}{1-\rho}} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1-\rho^2 & 1-\rho^2 & 0 \end{pmatrix} \end{aligned}$$

Finally,

$$\mathrm{vec}(\Theta^{-1} \varepsilon_t \varepsilon_t^\intercal \Theta^{-1}) = \frac{1}{(1-\rho^2)^2} \begin{pmatrix} \varepsilon_{1t}^2 - 2\rho\varepsilon_{1t}\varepsilon_{2t} + \rho^2 \varepsilon_{2t}^2 \\ (1+\rho^2)\varepsilon_{1t}\varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \\ (1+\rho^2)\varepsilon_{1t}\varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \\ \rho^2 \varepsilon_{1t}^2 - 2\rho\varepsilon_{1t}\varepsilon_{2t} + \varepsilon_{2t}^2 \end{pmatrix}, \quad \mathrm{vec}(\Theta^{-1}) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 \\ -\rho \\ -\rho \\ 1 \end{pmatrix}$$

such that

$$\begin{pmatrix} 0 & 1-\rho^2 & 1-\rho^2 & 0 \end{pmatrix} \mathrm{vec}(\Theta^{-1} \varepsilon_t \varepsilon_t^\intercal \Theta^{-1} - \Theta^{-1}) = \frac{2}{1-\rho^2} \left\{ (1+\rho^2)\varepsilon_{1t}\varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \right\} + 2\rho$$

and

$$u_t = \frac{1}{1-\rho^2} \left\{ (1+\rho^2)\varepsilon_{1t}\varepsilon_{2t} - \rho(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) \right\} + \rho$$

which is the score given by Harvey (2013, Section 7.3.1) for the bivariate Gaussian case. He uses the parametrization $\gamma = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$ and

$$g = \frac{\exp \gamma + \exp(-\gamma)}{2}, \quad \bar{g} = \frac{\exp \gamma - \exp(-\gamma)}{2}$$

so that

$$u_t = (2g^2 - 1)\varepsilon_{1t}\varepsilon_{2t} - g\bar{g}(\varepsilon_{1t}^2 + \varepsilon_{2t}^2) + \frac{\bar{g}}{g}.$$

For the information matrix of Theorem 2, the expressions reduce to

$$\begin{aligned} I_{\delta\delta} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial\theta_t^\intercal}{\partial\delta} M_t \left(\Theta_t^{-1}\otimes\Theta_t^{-1}\right) M_t^\intercal \frac{\partial\theta_t}{\partial\delta^\intercal}\right] \\ I_{\tau\tau} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial \mathrm{vec}(H_t)^\intercal}{\partial\tau}\left(H_t^{-1}\otimes H_t^{-1}\right)\frac{\partial\mathrm{vec}(H_t)}{\partial\tau^\intercal}\right] \\ I_{\tau\delta} &= \frac{1}{2}\mathbb{E}\left[\frac{\partial \mathrm{vec}(H_t)^\intercal}{\partial\tau}\left(H_t^{-1}\otimes H_t^{-1}\right)(D_t\otimes D_t) M_t^\intercal \frac{\partial\theta_t}{\partial\delta^\intercal}\right] \end{aligned}$$

With $M_t = (1-\rho^2)\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix}$, straightforward calculations show that $\frac{1}{2} M_t \left(\Theta_t^{-1}\otimes\Theta_t^{-1}\right) M_t^\intercal = (1+\rho^2) = 2 - 1/g^2$, and

$$I_{\delta\delta} = \mathbb{E}\left[(2-1/g^2)\frac{\partial\theta_t^\intercal}{\partial\delta}\frac{\partial\theta_t}{\partial\delta^\intercal}\right]$$

Furthermore,

$$H_t = \begin{pmatrix} \exp(2h_1) & \rho\exp(h_1+h_2) \\ \rho\exp(h_1+h_2) & \exp(2h_2) \end{pmatrix}, \quad H_t^{-1} = \frac{1}{(1-\rho^2)}\begin{pmatrix} \exp(-2h_1) & -\rho\exp(-h_1-h_2) \\ -\rho\exp(-h_1-h_2) & \exp(-2h_2) \end{pmatrix}$$

and

$$\frac{\partial\mathrm{vec}(H_t)}{\partial\tau^\intercal} = \frac{\partial\mathrm{vec}(H_t)}{\partial\varpi^\intercal}\frac{\partial\varpi}{\partial\tau^\intercal}, \quad \frac{\partial\mathrm{vech}(H_t)}{\partial\varpi^\intercal} = \begin{pmatrix} 2\exp(2h_1) & 0 & 0 \\ \rho\exp(h_1+h_2) & \rho\exp(h_1+h_2) & \exp(h_1+h_2) \\ 0 & 2\exp(2h_2) & 0 \end{pmatrix}$$

where $\varpi := (h_1, h_2, \rho)^\intercal$. Straightforward calculations show that

$$\frac{1}{2}\frac{\partial\mathrm{vec}(H_t)^\intercal}{\partial\varpi}\left(H_t^{-1}\otimes H_t^{-1}\right)\frac{\partial\mathrm{vec}(H_t)}{\partial\varpi^\intercal} = \begin{pmatrix} 1+g^2 & 1-g^2 & -\bar{g}/g \\ 1-g^2 & 1+g^2 & -\bar{g}/g \\ -\bar{g}/g & -\bar{g}/g & 2-1/g^2 \end{pmatrix} =: J$$

which is the matrix given in Proposition 40 of Harvey (2013). Thus,

$$I_{\tau\tau} = \mathbb{E}\left[\frac{\partial\varpi^\intercal}{\partial\tau} J \frac{\partial\varpi}{\partial\tau^\intercal}\right]$$

Similarly, we obtain

$$I_{\tau\delta} = \mathbb{E}\left[\frac{\partial\varpi^\intercal}{\partial\tau} J_3 \frac{\partial\theta}{\partial\delta^\intercal}\right]$$

where J_3 is the third column of J.

D Lyapounov exponent of the SRE

In the general case $N \geq 2$ and $\theta_t \in \mathcal{T}_{N,\epsilon} \subset \mathbb{R}^{N(N-1)/2}$, we denote the spectral (operator) matrix norm by $\|\cdot\|$, and the L_2 norm by $\|\cdot\|_2$. At several occasions we will use that

$$\|\Theta_t\| \leq \|\Theta_t\|_2 \leq N, \quad a.s. \tag{D.1}$$

because Θ_t is a correlation matrix and, hence, $\sup_{i,j} |\Theta_{t,ij}| = 1$.

For the differential of the impact function $g(\cdot)$ we have

$$dg(\theta_t) = \frac{1}{2}\left\{(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2})dM_t + M_t d(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2})\right\} \tag{D.2}$$

Let $Y_t = \Theta_t \otimes \Theta_t$, and Then,

$$\begin{aligned}
\|dY_t^{-1/2}\| &\leq \|d\text{vec}Y_t^{-1/2}\|_2 && \text{(D.3)}\\
&= \|(Y_t^{-1/2} \otimes Y_t^{-1/2})d\text{vec}Y_t^{1/2}\|_2 && \text{(D.4)}\\
&= \|(Y_t^{-1/2} \otimes Y_t^{-1/2})(Y_t^{1/2} \otimes I_N + I_N \otimes Y_t^{1/2})^{-1}d\text{vec}Y_t\|_2 && \text{(D.5)}\\
&= \|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1}d\text{vec}Y_t\|_2 && \text{(D.6)}\\
&\leq \|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1}\|\ \|d\text{vec}Y_t\|_2 && \text{(D.7)}
\end{aligned}$$

where in (D.7) we used that the spectral norm $\|\cdot\|$ is induced by the L_2 vector norm $\|\cdot\|_2$, see e.g. Magnus and Neudecker (2019).

Now,

$$\begin{aligned}
\|(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1}\| &\leq \lambda_{max}(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t)^{-1} && \text{(D.8)}\\
&= \lambda_{min}^{-1}(Y_t \otimes Y_t^{1/2} + Y_t^{1/2} \otimes Y_t) && \text{(D.9)}\\
&\leq \lambda_{min}^{-1}(Y_t \otimes Y_t^{1/2}) && \text{(D.10)}\\
&\leq \epsilon^{-3} < \infty, \quad a.s. && \text{(D.11)}
\end{aligned}$$

by Assumption (A6), and

$$\begin{aligned}
d\text{vec}(Y_t) &= (I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})d\text{vec}(\Theta_t) && \text{(D.12)}\\
&= (I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})M_t^{\intercal}d\theta_t && \text{(D.13)}
\end{aligned}$$

with commutation matrix C_{NN}. We have

$$\begin{aligned}\|d\text{vec}(Y_t)\|_2 &\leq \|(I_N \otimes C_{NN} \otimes I_N)(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})M_t\| \, \|d\theta_t\|_2 \\ &\leq \|(I_N \otimes C_{NN} \otimes I_N)\| \, \|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| \, \|M_t\| \, \|d\theta_t\|_2 \\ &= \|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| \, \|M_t\| \, \|d\theta_t\|_2\end{aligned}$$

where the first inequality follows by the fact that the spectral norm is induced by the L_2 vector norm, the second inequality by submultiplicativity of the spectral norm, and the last equality by the fact that $\lambda_{max}(C_{NN}) = 1$. Now,

$$\begin{aligned}\|(I_{N^2} \otimes \text{vec}(\Theta_t) + \text{vec}(\Theta_t) \otimes I_{N^2})\| &\leq 2\|I_{N^2} \otimes \text{vec}(\Theta_t)\| \\ &= 2\lambda_{max}^{1/2}\left\{(I_{N^2} \otimes \text{vec}(\Theta_t))^\intercal (I_{N^2} \otimes \text{vec}(\Theta_t))\right\} \\ &= 2\lambda_{max}^{1/2}\left\{I_{N^2}\text{vec}(\Theta_t)^\intercal \text{vec}(\Theta_t)\right\} \\ &= 2\left\{\text{vec}(\Theta_t)^\intercal \text{vec}(\Theta_t)\right\}^{1/2} \leq 2N, \quad a.s.\end{aligned}$$

where the last inequality follows by (D.1). Furthermore, note that M_t can be written as $M_t = M_{1t} + M_{2t}$, where $M_{2t} = P_l^\intercal A_t$, $M_{1t} = -E_s A_t E_d^\intercal (E_d A_t E_d^\intercal)^{-1} P_d^\intercal A_t$, and P_d and P_l are defined by $P_d \text{diag}(Q) = \text{vec}(Q)$ and $P_l \text{vecl}(Q) = \text{vec}(Q)$, respectively, for any symmetric conformable matrix Q.

$$\begin{aligned}\|M_t\| &\leq \|M_{1t}\| + \|M_{2t}\| \qquad \text{(D.14)} \\ &\leq \left(\left\|(E_d A_t E_d^\intercal)^{-1} E_d A_t E_s^\intercal\right\| + 1\right) \|A_t\| \\ &\leq (\lambda_{min}^{-1}(A_t)\lambda_{max}(A_t) + 1)\lambda_{max}(A_t) \\ &\leq (N/\epsilon + 1)N < \infty \qquad \text{(D.15)}\end{aligned}$$

where we have used that $E_d A_t E_d^\intercal$ and $E_d A_t E_s$ are principal submatrices of A_t, and hence $\lambda_{min}(E_d A_t E_d^\intercal) \geq \lambda_{min}(A_t)$ and $\lambda_{max}(E_d A_t E_s^\intercal) \leq \lambda_{max}(A_t)$. Moreover, $\lambda_{min}(A_t) = \lambda_{min}(\Theta_t) \geq \epsilon > 0$ a.s. by Assumption (A6), and $\lambda_{max}(A_t) = \lambda_{max}(\Theta_t) \leq N$ a.s. because $\|\Theta_t\| \leq \|\Theta_t\|_2$, which is due to (D.1). Thus, $\|d\text{vec}(Y_t^{-1/2})\|_2 \leq 2N^2\epsilon^{-3}(N/\epsilon + 1)\|d\theta_t\|_2$, and the norm of the second term in (D.2) is

$$\frac{1}{2}\|M_t dY_t^{-1/2}\| \leq (N/\epsilon)^3(N/\epsilon + 1)^2\|d\theta_t\|_2.$$

As for the first term, note that $\left\|\Theta_t^{-1/2} \otimes \Theta_t^{-1/2}\right\| \leq \lambda_{min}^{-1}(\Theta) \leq \epsilon^{-1}$ by Assumption (A6), and therefore

$$\frac{1}{2}\|(\Theta_t^{-1/2} \otimes \Theta_t^{-1/2})dM_t\| \leq \frac{1}{2}\epsilon^{-1}\|dM_t\|$$

by submultiplicativity of the spectral norm.

Note that $dM_t = dM_{1t} + dM_{2t}$, where

$$\begin{aligned} dM_{1t} &= -E_s dA_t E_d^\intercal (E_d A_t E_d^\intercal)^{-1} P_d^\intercal A_t \\ &+ E_s A_t E_d^\intercal (E_d A_t E_d^\intercal)^{-1}(E_d dA_t E_d^\intercal)(E_d A_t E_d^\intercal)^{-1} P_d A_t \\ &- E_s A_t E_d^\intercal (E_d A_t E_d^\intercal)^{-1} P_d^\intercal dA_t. \end{aligned}$$

The matrix A_t in (3.8) has an equivalent representation as $A_t = \int_0^1 \Theta_t^r \otimes \Theta_t^{1-r} dr$, see Linton and McCrorie (1995) and Higham (2008). Thus,

$$dA_t = d\int_0^1 \Theta_t^r \otimes \Theta_t^{1-r} dr \tag{D.16}$$

$$= \int_0^1 d\Theta_t^r \otimes \Theta_t^{1-r} dr + \int_0^1 \Theta_t^r \otimes d\Theta_t^{1-r} dr \tag{D.17}$$

$$= \int_0^1 U_r \otimes \Theta_t^{1-r} dr + \int_0^1 \Theta_t^r \otimes V_r dr \tag{D.18}$$

with

$$U_r := \int_0^1 \Theta_t^{(1-s)r} d(r \log \Theta_t) \Theta_t^{rs} ds \tag{D.19}$$

$$V_r := \int_0^1 \Theta_t^{(1-s)(1-r)} d((1-r) \log \Theta_t) \Theta_t^{(1-r)s} ds \tag{D.20}$$

This yields, using the triangle inequality,

$$\|dA_t\| \leq \left\|\int_0^1 U_r \otimes \Theta_t^{1-r} dr\right\| + \left\|\int_0^1 \Theta_t^r \otimes V_r dr\right\| \tag{D.21}$$

$$\leq \int_0^1 \|U_r \otimes \Theta_t^{1-r}\| dr + \int_0^1 \|\Theta_t^r \otimes V_r\| dr \tag{D.22}$$

and

$$\begin{aligned}
\|U_r\| &\leq \int_0^1 e^{\lambda_{max}(1-s)r} e^{\lambda_{max} rs} ds\ r\|d\log\Theta_t\| && \text{(D.23)}\\
&= \int_0^1 e^{\lambda_{max} r} ds\ r\|d\log\Theta_t\| && \text{(D.24)}\\
&= re^{\lambda_{max} r}\|d\log\Theta_t\| && \text{(D.25)}\\
\|V_r\| &\leq (1-r)e^{\lambda_{max}(1-r)}\|d\log\Theta_t\| && \text{(D.26)}
\end{aligned}$$

Therefore, $\|U_r \otimes \Theta_t^{1-r}\| \leq re^{\lambda_{max}}\|d\log\Theta_t\|$ and $\|\Theta_t^r \otimes V_r\| \leq (1-r)e^{\lambda_{max}}\|d\log\Theta_t\|$, and thus, $\|dA_t\| \leq (\int_0^1 rdr + \int_0^1 (1-r)dr)e^{\lambda_{max}}\|d\log\Theta_t\| = e^{\lambda_{max}}\|d\log\Theta_t\|$. Now,

$$\begin{aligned}
\|d\log\Theta_t\| &\leq \|d\text{vec}\log\Theta_t\|_2 && \text{(D.27)}\\
&\leq \|A_t^{-1}M_t^\intercal d\theta_t\|_2 && \text{(D.28)}\\
&\leq (N/\epsilon+1)\|d\theta_t\|_2 && \text{(D.29)}
\end{aligned}$$

and therefore $\|dA_t\| \leq N(N/\epsilon+1)\|d\theta_t\|_2$. Finally, $\|dM_{1t}\| \leq \|dA_t\|(2N/\epsilon + N^2/\epsilon^2)$ and $\|dM_t\| \leq \|dM_{1t}\| + \|dM_{2t}\| \leq \|dA_t\|(2N/\epsilon + N^2/\epsilon^2 + 1) \leq (2N^2/\epsilon + N^3/\epsilon^2 + N)(N/\epsilon + 1)\|d\theta_t\|_2$. To conclude,

$$\begin{aligned}
\sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \left\| \frac{\partial \text{vec}g(\theta_t)}{\partial\theta_t^\intercal} \right\| &= \sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \frac{\|d\text{vec}g(\theta_t)\|_2}{\|d\theta_t\|_2} && \text{(D.30)}\\
&\leq (N/\epsilon)(N/\epsilon+1)^2\left\{(N/\epsilon)^2 + (N/\epsilon+1)/2\right\} && \text{(D.31)}\\
&=: B(\epsilon,N) < \infty && \text{(D.32)}
\end{aligned}$$

and the Lipschitz constant is given by

$$\Lambda(\psi) = \sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \left\| \frac{\partial\psi(\theta_t)}{\partial\theta_t^\intercal} \right\| = \sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \frac{\|d\psi(\theta_t)\|_2}{\|d\theta_t\|_2} = \sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \frac{\|\Phi d\theta_t + Kdg(\theta_t)Z_t\|_2}{\|d\theta_t\|_2}$$

Note that

$$\Lambda(\psi) \leq \bar{\Lambda}(\psi) := \|\Phi\| + \|K\| \sup_{\theta_t\in\mathcal{T}_{N,\epsilon}} \left\| \frac{\partial \text{vec}g(\theta_t)}{\partial\theta_t^\intercal} \right\| \|Z_t\|_2 < \infty$$

where the upper bound $\bar{\Lambda}(\psi)$ is equal to $\phi_{max} + \kappa_{max}B(\epsilon,N)\|Z_t\|_2$ if Φ and K are diagonal with maximum diagonal elements given by ϕ_{max} and κ_{max}, respectively. This shows that the SRE is Lipschitz continuous with an upper bound for the Lipschitz constant given

by $\bar{\Lambda}(\psi)$. For given parameters, we can simulate a large number of Z_t, calculate $\bar{\Lambda}(\psi)$ for each draw, and then estimate $\mathbb{E}\log|\bar{\Lambda}(\psi)|$ by the simulation average of $\log|\bar{\Lambda}(\psi)|$. Then, $\mathbb{E}\log|\bar{\Lambda}(\psi)|<0$ implies that the Lyapounov exponent is negative, i.e. $\mathbb{E}\log|\Lambda(\psi)|<0$.

In large dimensions, or when ϵ is very close to zero, this is unlikely to hold, so it is better to work with $\Lambda(\psi)$ directly and solve for given Z_t the problem

$$\Lambda(\psi)=\sup_{\theta_t\in\mathcal{T}_{N,\epsilon}}\left\|\Phi+(Z_t^{\intercal}\otimes K)\frac{\partial \text{vec}g(\theta_t)}{\partial\theta_t^{\intercal}}\right\|$$

where the solution for θ_t is found numerically over the space $\mathcal{T}_{N,\epsilon}$, and where (D.2) is used for the differential of $g(\cdot)$. Then, one repeats this for many draws of Z_t and takes the average of $\log|\Lambda(\psi)|$ as an estimate of the Lyapounov exponent. We have done this for our six-dimensional empirical example (Model 2), using the estimated parameters in Table G.9 and 100 draws of a multivariate student-t distribution. For the lower bound of the eigenvalues of Θ_t we choose $\epsilon=0.1$, motivated by Figure G.2 that shows the minimum eigenvalue of simulated Θ_t for the given parameters by increasing the sample size up to $T=10^6$, which is not smaller than 0.2. We obtain an estimate of the Lyapounov exponent of -0.013, which suggests that Assumption (A7) holds for the estimated model.

E Expressions for the information matrix

We give analytical expressions for the second derivatives in (3.19) and (3.20). Note first that

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial\delta\partial\tau_i^\intercal} &= \frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta}\frac{\partial\tilde{u}_t}{\partial\tau_i^\intercal} + (\tilde{u}_t^\intercal\otimes I_m)\frac{\partial \mathrm{vec}}{\partial\tau_i^\intercal}\frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta}\\ \frac{\partial^2 l_t^C}{\partial\delta\partial\delta^\intercal} &= \frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta}\frac{\partial\tilde{u}_t}{\partial\delta^\intercal} + (\tilde{u}_t^\intercal\otimes I_m)\frac{\partial \mathrm{vec}}{\partial\delta^\intercal}\frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta},\end{aligned}$$

where $m = \dim(\delta)$ and $\partial\tilde{\theta}_t^\intercal/\partial\delta$ is given in (3.17), replacing θ_t by $\tilde{\theta}_t$, u_t by $\tilde{u}_t$, z_t by $\tilde{z}_t$ and x_t by $\tilde{x}_t$. Furthermore,

$$\begin{aligned}\frac{\partial \mathrm{vec}}{\partial\tau_i^\intercal}\frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta} &= \frac{\partial \mathrm{vec}(\tilde{z}_{t-1}^\intercal)}{\partial\tau_i^\intercal} + (\tilde{x}_{t-1}\otimes I_m)\frac{\partial \mathrm{vec}}{\partial\tau_i^\intercal}\frac{\partial\tilde{\theta}_{t-1}^\intercal}{\partial\delta} + (I_{N^*}\otimes\frac{\partial\tilde{\theta}_{t-1}^\intercal}{\partial\delta})\frac{\partial \mathrm{vec}(\tilde{x}_{t-1}^\intercal)}{\partial\tau_i^\intercal}\\ \frac{\partial \mathrm{vec}}{\partial\delta^\intercal}\frac{\partial\tilde{\theta}_t^\intercal}{\partial\delta} &= \frac{\partial \mathrm{vec}(\tilde{z}_{t-1}^\intercal)}{\partial\delta^\intercal} + (\tilde{x}_{t-1}\otimes I_m)\frac{\partial \mathrm{vec}}{\partial\delta^\intercal}\frac{\partial\tilde{\theta}_{t-1}^\intercal}{\partial\delta} + (I_{N^*}\otimes\frac{\partial\tilde{\theta}_{t-1}^\intercal}{\partial\delta})\frac{\partial \mathrm{vec}(\tilde{x}_{t-1}^\intercal)}{\partial\delta^\intercal}\end{aligned}$$

Next we calculate the derivatives involving the degrees of freedom parameter η. The first derivative is given by

$$\begin{aligned}\frac{\partial l}{\partial\eta} &= \frac{\partial c(\eta,N)}{\partial\eta} - N\frac{\partial c(\eta,1)}{\partial\eta} - \frac{1}{2}\sum_{t=1}^T\frac{\partial l_t^C}{\partial\eta}\\ \frac{\partial l_t^C}{\partial\eta} &= \tilde{u}_t^\intercal\frac{\partial\tilde{\theta}_t^\intercal}{\partial\eta} + \log\left(1+\frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\eta-2}\right) + \frac{\eta+N}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left\{2\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial\tilde{\varepsilon}_t}{\partial\eta} - \frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\eta-2}\right\}\\ &\quad - \sum_{i=1}^N\log\left(1+\frac{\tilde{\varepsilon}_{it}^2}{\eta-2}\right) + \frac{\eta+1}{\eta-2+\tilde{\varepsilon}_{it}^2}\left\{2\tilde{\varepsilon}_{it}\frac{\partial\tilde{\varepsilon}_{it}}{\partial\eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta-2}\right\}\end{aligned}$$

The second derivatives can be obtained by straightforward calculations. First, we have

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial\eta\partial\tau_i^\intercal} &= \frac{\partial\tilde{\theta}_t^\intercal}{\partial\eta}\frac{\partial\tilde{u}_t}{\partial\tau_i^\intercal} + \tilde{u}_t^\intercal\frac{\partial^2\tilde{\theta}_t}{\partial\eta\partial\tau_i^\intercal}\\ &\quad + \left\{\frac{\eta+N}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left(2\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial\tilde{\varepsilon}_t}{\partial\eta} - \frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\eta-2}\right) - \frac{N+2}{\eta-2}\right\}\left(\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t\right)^{-1}\frac{\partial\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial\tau_i^\intercal}\\ &\quad - \left\{\frac{\eta+1}{\eta-2+\tilde{\varepsilon}_{it}^2}\left(2\tilde{\varepsilon}_{it}\frac{\partial\tilde{\varepsilon}_{it}}{\partial\eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta-2}\right) - \frac{3}{\eta-2}\right\}\left(\eta-2+\tilde{\varepsilon}_{it}^2\right)^{-1}2\tilde{\varepsilon}_{it}\frac{\partial\tilde{\varepsilon}_{it}}{\partial\tau_i^\intercal}\\ &\quad + \frac{2(\eta+N)}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left(\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial^2\tilde{\varepsilon}_t}{\partial\eta\partial\tau_i^\intercal} + \frac{\partial\tilde{\varepsilon}_t^\intercal}{\partial\eta}\frac{\partial\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial\tau_i^\intercal}\right) - \frac{2(\eta+1)}{\eta-2+\tilde{\varepsilon}_{it}^2}\left(\tilde{\varepsilon}_{it}\frac{\partial^2\tilde{\varepsilon}_{it}}{\partial\eta\partial\tau_i^\intercal} + \frac{\partial\tilde{\varepsilon}_{it}}{\partial\eta}\frac{\partial\tilde{\varepsilon}_{it}}{\partial\tau_i^\intercal}\right)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \tilde{\varepsilon}_t^\intercal \tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial \tau_i^\intercal} &= 2\tilde{\varepsilon}_t^\intercal \tilde{\Theta}_t^{-1}\frac{\partial \tilde{\varepsilon}_t}{\partial \tau_i^\intercal} - 2(\tilde{\varepsilon}_t^\intercal \tilde{\Theta}_t^{-1}\otimes \tilde{\varepsilon}_t^\intercal \tilde{\Theta}_t^{-1})\tilde{M}_t\frac{\partial \tilde{\theta}_t}{\partial \tau_i^\intercal}\\ \frac{\partial \tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial \tau_i^\intercal} &= \tilde{\Theta}_t^{-1}\frac{\partial \tilde{\varepsilon}_t}{\partial \tau_i^\intercal} - 2(\tilde{\varepsilon}_t^\intercal \tilde{\Theta}_t^{-1}\otimes \tilde{\Theta}_t^{-1})\tilde{M}_t\frac{\partial \tilde{\theta}_t}{\partial \tau_i^\intercal}\end{aligned}$$

Then

$$\begin{aligned}\frac{\partial^2 l_t^C}{\partial\eta\partial\delta^\intercal} &= \frac{\partial\tilde{\theta}_t^\intercal}{\partial\eta}\frac{\partial\tilde{u}_t}{\partial\delta^\intercal} + \tilde{u}_t^\intercal\frac{\partial^2\tilde{\theta}_t}{\partial\eta\partial\delta^\intercal}\\ &+ \left\{\frac{\eta+N}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left(2\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial\tilde{\varepsilon}_t}{\partial\eta} - \frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\eta-2}\right) - \frac{N+2}{\eta-2}\right\}\left(\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t\right)^{-1}\frac{\partial\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial\delta^\intercal}\\ &+ \frac{2(\eta+N)}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left(\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial^2\tilde{\varepsilon}_t}{\partial\eta\partial\delta^\intercal} + \frac{\partial\tilde{\varepsilon}_t^\intercal}{\partial\eta}\frac{\partial\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial\delta^\intercal}\right)\end{aligned}$$

and finally,

$$\begin{aligned}\frac{\partial l^2}{\partial\eta^2} &= \frac{\partial^2 c(\eta,N)}{\partial\eta^2} - N\frac{\partial^2 c(\eta,1)}{\partial\eta^2} - \frac{1}{2}\sum_{t=1}^{T}\frac{\partial^2 l_t^C}{\partial\eta^2}\\ \frac{\partial^2 l_t^C}{\partial\eta^2} &= \frac{\partial\tilde{\theta}_t^\intercal}{\partial\eta}\frac{\partial\tilde{u}_t}{\partial\eta} + \tilde{u}_t^\intercal\frac{\partial^2\tilde{\theta}_t}{\partial\eta\partial\eta} + (\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t)^{-1}\times\\ &\times \left\{\left(2\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial\tilde{\varepsilon}_t}{\partial\eta} - \frac{\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\eta-2}\right)\left[1-\frac{N+2}{\eta-2}-\frac{N+\eta}{\eta-2+\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}\left(1+\frac{\partial\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\tilde{\varepsilon}_t}{\partial\eta}\right)\right]\right.\\ &+ \left. 2\frac{N+2}{\eta-2}(\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\otimes\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1})\tilde{M}_t\frac{\partial\tilde{\theta}_t}{\partial\eta} + 2(\eta+N)\left[\frac{\partial\tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}}{\partial\eta}\frac{\partial\tilde{\varepsilon}_t}{\partial\eta} + \tilde{\varepsilon}_t^\intercal\tilde{\Theta}_t^{-1}\frac{\partial^2\tilde{\varepsilon}_t}{\partial\eta^2}\right]\right\}\\ &- \sum_{i=1}^{N}(\eta-2+\tilde{\varepsilon}_{it}^2)^{-1}\times\\ &\times \left\{\left(2\tilde{\varepsilon}_t\frac{\partial\tilde{\varepsilon}_{it}}{\partial\eta} - \frac{\tilde{\varepsilon}_{it}^2}{\eta-2}\right)\left[\frac{\eta-5}{\eta-2}-\frac{1+\eta}{\eta-2+\tilde{\varepsilon}_{it}^2}\left(1+\frac{\partial\tilde{\varepsilon}_{it}^2}{\partial\eta}\right)\right] + 2(\eta+1)\left[\left(\frac{\partial\tilde{\varepsilon}_{it}}{\partial\eta}\right)^2 + \tilde{\varepsilon}_{it}\frac{\partial^2\tilde{\varepsilon}_{it}}{\partial\eta^2}\right]\right\}\end{aligned}$$

where

$$
\begin{aligned}
\frac{\partial \tilde{\varepsilon}_t^{\intercal} \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \eta} &= 2\tilde{\varepsilon}_t^{\intercal} \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - 2(\tilde{\varepsilon}_t^{\intercal} \tilde{\Theta}_t^{-1} \otimes \tilde{\varepsilon}_t^{\intercal} \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \eta} \\
\frac{\partial \tilde{\Theta}_t^{-1} \tilde{\varepsilon}_t}{\partial \eta} &= \tilde{\Theta}_t^{-1} \frac{\partial \tilde{\varepsilon}_t}{\partial \eta} - 2(\tilde{\varepsilon}_t^{\intercal} \tilde{\Theta}_t^{-1} \otimes \tilde{\Theta}_t^{-1}) \tilde{M}_t \frac{\partial \tilde{\theta}_t}{\partial \eta} \\
\frac{\partial \tilde{\theta}_t}{\partial \eta} &= \Phi \frac{\partial \tilde{\theta}_{t-1}}{\partial \eta} + K \frac{\partial \tilde{u}_{t-1}}{\partial \eta} \\
\frac{\partial^2 \tilde{\theta}_t}{\partial \eta^2} &= \Phi \frac{\partial^2 \tilde{\theta}_{t-1}}{\partial \eta^2} + K \frac{\partial^2 \tilde{u}_{t-1}}{\partial \eta^2}
\end{aligned}
$$

F Monte Carlo simulation

We carry out a Monte Carlo study to investigate the finite sample performance of the two estimation methods. We simulate three return series ($N = 3$) of length $T = 1000$ and $T = 5000$ observations. The parameters used in the data generating process are taken from our empirical example, using the three assets SPX, SBX, and EMU. For the degrees of freedom, we take the average of the degrees of freedom parameters from the marginal distributions and the one from the copula. The values are provided in Table F.1.

Table F.1: Parameters used in the data generating process

Volatility parameters			Correlation parameters			degrees of freedom
ω^v	κ^v	ϕ^v	ω			$\bar{\eta}$
-0.3133	0.1164	0.9355	-0.0250	0.0725	0.0202	0.1335
-0.0348	0.0200	0.9938	K	Φ		
-0.0909	0.0453	0.9803	0.0675	0.8715		

We generate 100 replications of the return series and estimate the model using both the one-step and the two-step estimation procedures. Our two measures of accuracy are the Bias and the root mean squared error (RMSE). For each parameter s, the measures are computed as Bias $= \frac{1}{100}\sum_{i=1}^{100} \hat{s}_i - s$, and RMSE $= \frac{1}{100}\sum_{i=1}^{100} \sqrt{(\hat{s}_i - s)^2}$.

The results are presented in Table F.2. The left-hand panel compares the evolution of the bias in the parameter estimates across different sample sizes for both the one-step and two-step estimation procedures. For the volatility dynamics, we observe downward biases for ω^v and ϕ^v, and upward biases for κ^v. For the correlation dynamics, the estimates ω and κ are only slightly biased, while the estimates for the mean reversion parameter ϕ are downward biased, especially for the one-step method. In all cases, the absolute biases are reduced when increasing the sample size. For the degrees of freedom parameter $\bar{\eta}$, we observe a slightly higher upward bias for the two-step method, which might be due to the aggregation of the marginal degrees of freedom and of the copula one into a single

parameter. The bias remains small, however, and decreases with larger samples. When the sample size is increased from $T = 1000$ to $T = 5000$, we observe an overall reduction in the magnitude of the estimation bias for both the one-step and two-step estimation procedures.

The right-hand panel compares the evolution of the RMSE in the parameter estimates across different sample sizes for both the one-step and two-step estimation procedures. For the volatility dynamics, we notice that the one-step and two-step estimation procedures have comparable RMSE of the parameter estimates. For the correlation dynamics, the two-step estimation procedure exhibits lower RMSE than the one-step estimation case, which is mainly due to the lower biases. When the sample size is increased from $T = 1000$ to $T = 5000$, we observe an overall reduction in the RMSE of the parameter estimates for both the one-step and two-step estimation procedures.

The general picture is that the two-step method is performing better for the considered parameterization and sample sizes. It has faster computation times (the average estimation times are indicated in the last row of Table F.2), and generally better performances than the one-step method in terms of RMSE.

Table F.2: Bias and RMSE of parameter estimates

	Bias				RMSE			
	T=1000		T=5000		T=1000		T=5000	
	one-step	two-step	one-step	two-step	one-step	two-step	one-step	two-step
ω_1^v	-0.0413	-0.0384	-0.0012	0.0007	0.0753	0.0748	0.0270	0.0273
ω_2^v	-0.0770	-0.0749	-0.0090	-0.0072	0.0844	0.0842	0.0131	0.0129
ω_3^v	-0.0416	-0.0479	-0.0053	-0.0047	0.0506	0.0585	0.0149	0.0159
κ_1^v	0.0019	0.0009	0.0008	-0.0000	0.0106	0.0114	0.0047	0.0053
κ_2^v	0.0014	0.0012	0.0002	0.0001	0.0053	0.0056	0.0022	0.0023
κ_3^v	0.0006	0.0010	0.0001	-0.0005	0.0069	0.0075	0.0033	0.0037
ϕ_1^v	-0.0084	-0.0078	-0.0002	0.0001	0.0155	0.0154	0.0056	0.0056
ϕ_2^v	-0.0137	-0.0133	-0.0016	-0.0013	0.0150	0.0150	0.0023	0.0023
ϕ_3^v	-0.0090	-0.0103	-0.0011	-0.0010	0.0109	0.0127	0.0032	0.0034
ω_1	-0.0019	-0.0013	0.0002	-0.0019	0.0186	0.0110	0.0142	0.0069
ω_2	0.0014	0.0001	-0.0048	0.0027	0.0531	0.0243	0.0430	0.0123
ω_3	0.0028	-0.0020	0.0015	0.0004	0.0154	0.0131	0.0118	0.0034
K	0.0056	-0.0035	0.0101	-0.0029	0.0857	0.0123	0.0612	0.0067
Φ	-0.0665	-0.0131	-0.0621	-0.0074	0.1164	0.0375	0.0819	0.0197
$\bar{\eta}$	0.0023	0.0148	0.0045	0.0081	0.0111	0.0267	0.0070	0.0142
Avg. t	1.9223	0.4938	8.6614	2.4342	1.9223	0.4938	8.6614	2.4342

Note: Avg. t is the average running time of each estimation over 100 replications expressed in hour units.

G Additional Tables and Figures

Table G.1: Description of the assets considered in the empirical application

	SECURITY	**DESCRIPTION**
SPX	SPX Index	S&P 500 Index
SBW	SBWGU Index	FTSE World Government Bond Index (WGBI) USD
EMU	NDDUEMU Index	MSCI EMU Net Total Return Index
OIL	CL1 Comdty	Generic 1st Crude Oil; WTI
GOLD	GOLDLNPM Index	LBMA Gold Price PM USD
CRIX	CRIX	Cryptocurrency Index

Table G.2: Descriptive statistics for historical log-returns

	SPX	**SBW**	**EMU**	**OIL**	**GOLD**	**CRIX**
N.Obs	1220	1220	1220	1220	1220	1220
Minimum	-0.0999	-0.0225	-0.1514	-0.2799	-0.0526	-0.4466
Maximum	0.0897	0.0181	0.0857	0.3196	0.0513	0.1985
Mean	0.0005	0.0001	0.0003	0.0014	0.0002	0.0011
Stdev	0.0111	0.0038	0.0122	0.0311	0.0089	0.0426
Skewness	-0.1796	-0.3471	-1.8953	1.5560	0.0410	-1.3486
Kurtosis	15.7469	3.8632	25.1767	23.0595	4.1170	13.8111
Sharpe Ratio	0.0470	0.0132	0.0269	0.0448	0.0204	0.0267
Beta	1.1666	0.1127	0.9668	3.1250	0.4091	2.5494

Note: The kurtosis reported in the table corresponds to the excess kurtosis.

Table G.3: The sample correlation matrix of returns

	SPX	**SBW**	**EMU**	**OIL**	**GOLD**	**CRIX**
SPX	1.0000	-0.1380	0.5753	0.2836	0.0766	-0.0898
SBW	-0.1380	1.0000	0.0831	-0.0288	0.5476	0.0600
EMU	0.5753	0.0831	1.0000	0.2473	0.1097	0.0385
OIL	0.2836	-0.0288	0.2473	1.0000	0.0578	0.0103
GOLD	0.0766	0.5476	0.1097	0.0578	1.0000	0.0515
CRIX	-0.0898	0.0600	0.0385	0.0103	0.0515	1.0000

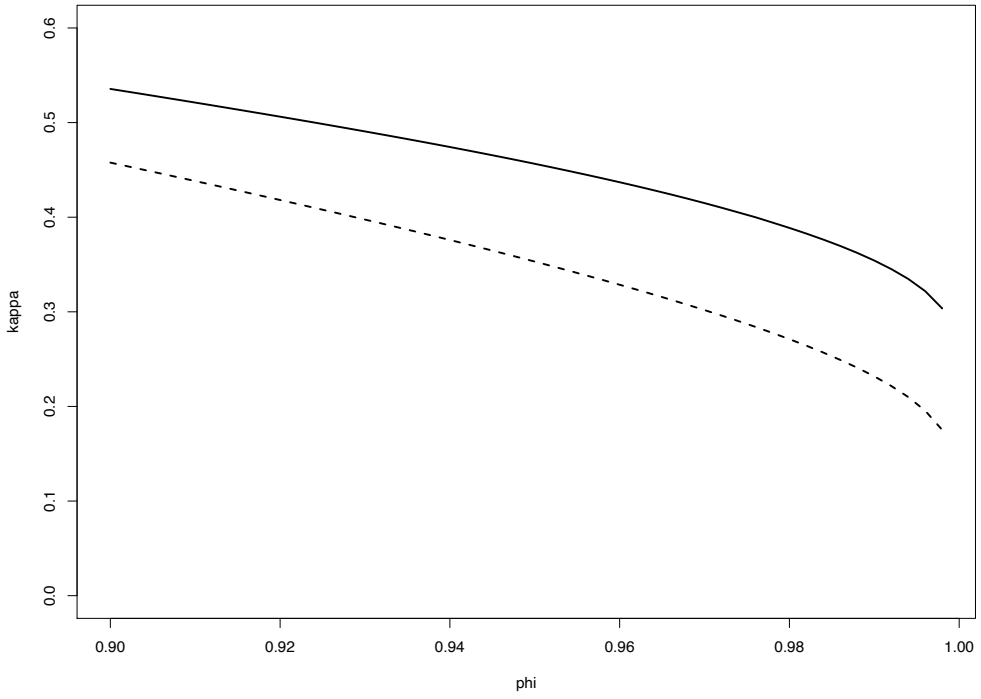

Figure G.1: The lines show the combinations of (ϕ, κ) for which $\mathrm{E}[\log \Lambda(\psi)] = 0$. Solid line: student-t with 5 df, dashed: 15 df. South of these functions is the region for which the Lyapounov exponent is negative.

Table G.4: Parameter estimates of the volatility using Beta-t-EGARCH(1,1)

	ω_i	κ_i	ϕ_i	$\bar{\eta}_i$
SPX	-0.31327	0.11642	0.93555	0.19126
	(0.00398)	(0.00069)	(0.00081)	(0.00121)
SBW	-0.03482	0.02004	0.99383	0.11880
	(0.00146)	(0.00029)	(0.00026)	(0.00120)
EMU	-0.09091	0.04531	0.98032	0.12895
	(0.00191)	(0.00046)	(0.00041)	(0.00097)
OIL	-0.09397	0.04746	0.97529	0.11695
	(0.00164)	(0.00037)	(0.00042)	(0.00104)
GOLD	-0.01888	0.01576	0.99613	0.18421
	(0.00076)	(0.00021)	(0.00016)	(0.00131)
CRIX	-0.17057	0.17687	0.94220	0.40950
	(0.00196)	(0.00102)	(0.00065)	(0.00150)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$.

Table G.5: Volatility model diagnostics

	SPX	SBW	EMU	OIL	GOLD	CRIX
score	0.4224	0.8021	0.1333	0.1478	0.1643	0.9878
ε	0.7751	0.6110	0.4500	0.3886	0.3072	0.0026
ε^2	0.6223	0.0623	0.7234	0.3237	0.2229	0.0577

Note: The table reports the p-values of Box-Pierce test applied to the score with respect to log volatility, the standardized residuals (ε) and squared standardized residuals (ε^2). The number of lags being tested is 30.

Table G.6: Descriptive statistics for standardized residuals

	SPX	**SBW**	**EMU**	**OIL**	**GOLD**	**CRIX**
N.Obs	845	845	845	845	845	845
Minimum	-6.0957	-4.2156	-6.7406	-4.0590	-4.2477	-5.8446
Maximum	3.3762	3.9590	3.7776	4.1949	4.5662	3.8765
Mean	0.0294	-0.0079	0.0122	0.0018	-0.0139	0.0183
Stdev	1.0268	0.9936	1.0077	0.9954	0.9882	0.7815
Skewness	-1.1159	-0.0773	-0.4852	-0.2146	0.1064	-0.9837
Kurtosis	4.8808	0.8876	2.8387	1.1157	1.6320	8.4533

Note: The kurtosis reported in the table corresponds to the excess kurtosis.

Table G.7: Two-step parameter estimates for the correlation part of the model. The diagonal elements of K and Φ are not restricted to be the same (Model 1).

	ω_{i*5+1}	ω_{i*5+2}	ω_{i*5+3}	ω_{i*5+4}	ω_{i*5+5}
	-0.03331	0.56013	0.12427	-0.00364	-0.01029
ω	(0.00990)	(0.02140)	(0.02359)	(0.00187)	(0.00833)
	0.00527	-0.01730	0.01239	0.00103	0.00208
	(0.00804)	(0.00460)	(0.03263)	(0.00079)	(0.00618)
	0.00201	0.03624	0.09152	0.00221	0.00022
	(0.00094)	(0.00805)	(0.01145)	(0.00199)	(0.00183)
	κ_{i*5+1}	κ_{i*5+2}	κ_{i*5+3}	κ_{i*5+4}	κ_{i*5+5}
	0.05506	0.05420	0.02531	0.00771	0.00000
K	(0.03334)	(0.03482)	(0.03438)	(0.03439)	(0.03440)
	0.04870	0.02947	0.01731	0.00551	0.01755
	(0.02108)	(0.03419)	(0.02156)	(0.03439)	(0.03233)
	0.04583	0.02583	0.03543	0.01497	0.03328
	(0.03431)	(0.03441)	(0.03447)	(0.03419)	(0.03441)
	ϕ_{i*5+1}	ϕ_{i*5+2}	ϕ_{i*5+3}	ϕ_{i*5+4}	ϕ_{i*5+5}
	0.82329	0.00000	0.52003	0.92355	0.28273
Φ	(0.04426)	(0.03440)	(0.04968)	(0.03543)	(0.03440)
	0.96677	0.59401	0.98322	0.91835	0.98210
	(0.04748)	(0.03699)	(0.04363)	(0.03539)	(0.04079)
	0.89026	0.23319	0.02115	0.94376	0.95013
	(0.06085)	(0.03451)	(0.03440)	(0.04146)	(0.05646)
$\bar{\eta}$	0.02863				
	(0.00670)				

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates, and $i = 0, 1, 2$. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta}_i = \frac{1}{\eta_i} \in (0, 1)$.

Table G.8: Model descriptions

	DESCRIPTION	**# Pars**
Model 1	full ω, diagonal K, diagonal Φ	46
Model 2	full ω, scalar K, scalar Φ	18
Model 3	scalar ω, diagonal K, diagonal Φ	32
Model 4	$\omega = 0$, diagonal K, diagonal Φ	31
Model 5	scalar ω, scalar K, scalar Φ	4
Model 6	$\omega = 0$, scalar K, scalar Φ	3

Note: # Pars is the number of parameters.

Table G.9: Parameter estimates for the dynamic correlation models

	Meta-t-log-corr	**Meta-t-cDCC**	**Meta-t-GAS**	**Meta-t-GAS-h**
κ	0.01901	0.02065	0.01919	0.01862
	(0.00015)	(0.00016)	(0.00015)	(0.00015)
ϕ	0.95527	0.93750	0.96162	0.96098
	(0.00057)	(0.00061)	(0.00047)	(0.00051)
$\bar{\eta}$	0.03258	0.03495	0.03326	0.03312
	(0.00026)	(0.00028)	(0.00027)	(0.00027)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta_i} = \frac{1}{\eta_i} \in (0, 1)$.

Table G.10: Parameter estimates for the dynamic volatility and correlation models

	t-cDCC	t-GAS	t-GAS-l	t-GAS-hl
α_1	0.16976	0.16711	0.13417	0.13107
	(0.05958)	(0.00234)	(0.00078)	(0.02919)
α_2	0.01799	0.01880	0.02004	0.01854
	(0.05958)	(0.03680)	(0.00020)	(0.01904)
α_3	0.06530	0.04153	0.04015	0.04063
	(0.05959)	(0.00708)	(0.00045)	(0.03273)
α_4	0.07391	0.07273	0.06185	0.06301
	(0.05964)	(0.00341)	(0.00051)	(0.04186)
α_5	0.02753	0.02004	0.02191	0.02218
	(0.05958)	(0.00874)	(0.00019)	(0.01375)
α_6	0.21436	0.15464	0.16283	0.15848
	(0.05966)	(0.00646)	(0.00071)	(0.02246)
α_7	0.01811	0.01003	0.01355	0.01439
	(0.00113)	(0.03440)	(0.00015)	(0.00578)
β_1	0.77216	0.93407	0.91498	0.90985
	(0.05957)	(0.00603)	(0.00082)	(0.05781)
β_2	0.98046	1.00000	0.99854	0.99921
	(0.00097)	(0.03440)	(0.00013)	(0.02684)
β_3	0.91487	0.99212	0.98609	0.98444
	(0.00017)	(0.00281)	(0.00028)	(0.03189)
β_4	0.89709	0.97438	0.97530	0.97599
	(0.00082)	(0.00517)	(0.00042)	(0.06085)
β_5	0.96657	0.99882	0.99736	0.99713
	(0.00063)	(0.00866)	(0.00011)	(0.02868)
β_6	0.73315	0.97121	0.92509	0.91195
	(0.00025)	(0.00726)	(0.00073)	(0.05214)
β_7	0.93653	0.90664	0.97642	0.96149
	(0.00140)	(0.03440)	(0.00046)	(0.04242)
$\bar{\eta}$	0.13152	0.13068	0.13378	0.12944
	(0.00022)	(0.00948)	(0.00044)	(0.00700)

Note: The numbers in parentheses are the asymptotic standard errors of the corresponding parameter estimates. For the degrees of freedom parameter η_i we estimate its inverse, i.e. $\bar{\eta_i} = \frac{1}{\eta_i} \in (0, 1)$. α_i and β_i with $i = 1, \ldots, 6$ are the volatility parameters. α_7 and β_7 are the correlation parameters.

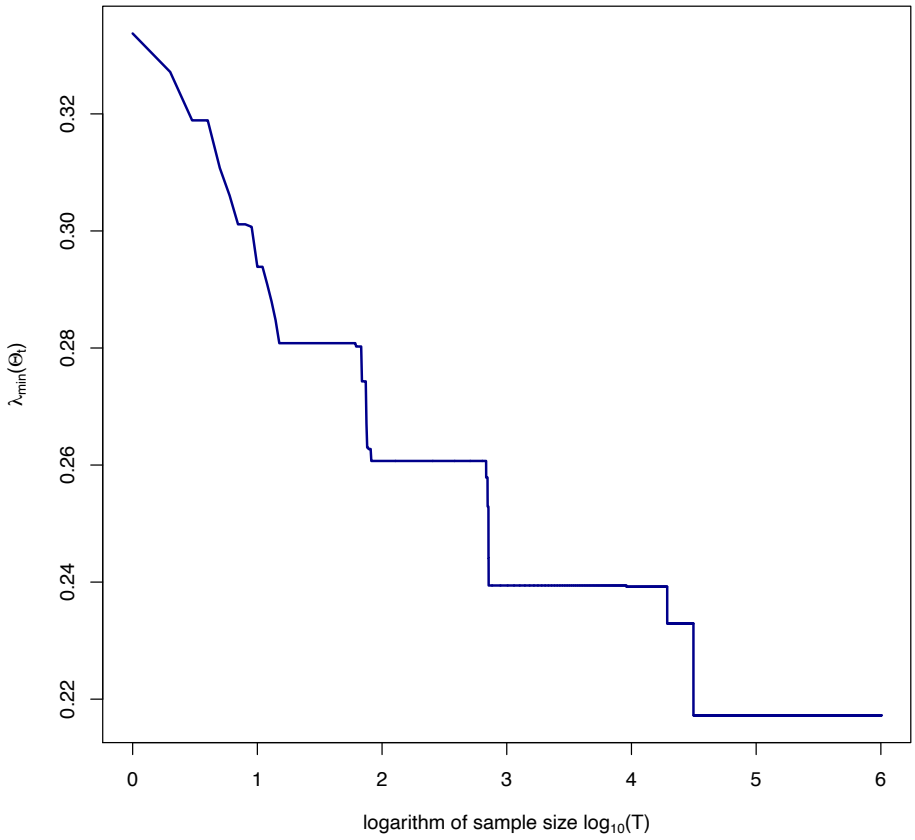

Figure G.2: Simulated $\lambda_{min}(\Theta_t)$ for $N = 6$ and different sample sizes up to 1,000,000 observations. The abscissa is the sample size on a log scale. The parameters of Θ_t are set to the estimates of Model 2 in our empirical example. The innovations are drawn from a multivariate student-t distribution, with the degrees of freedom being the average of the ones from the marginal distributions and from the copula parameter.

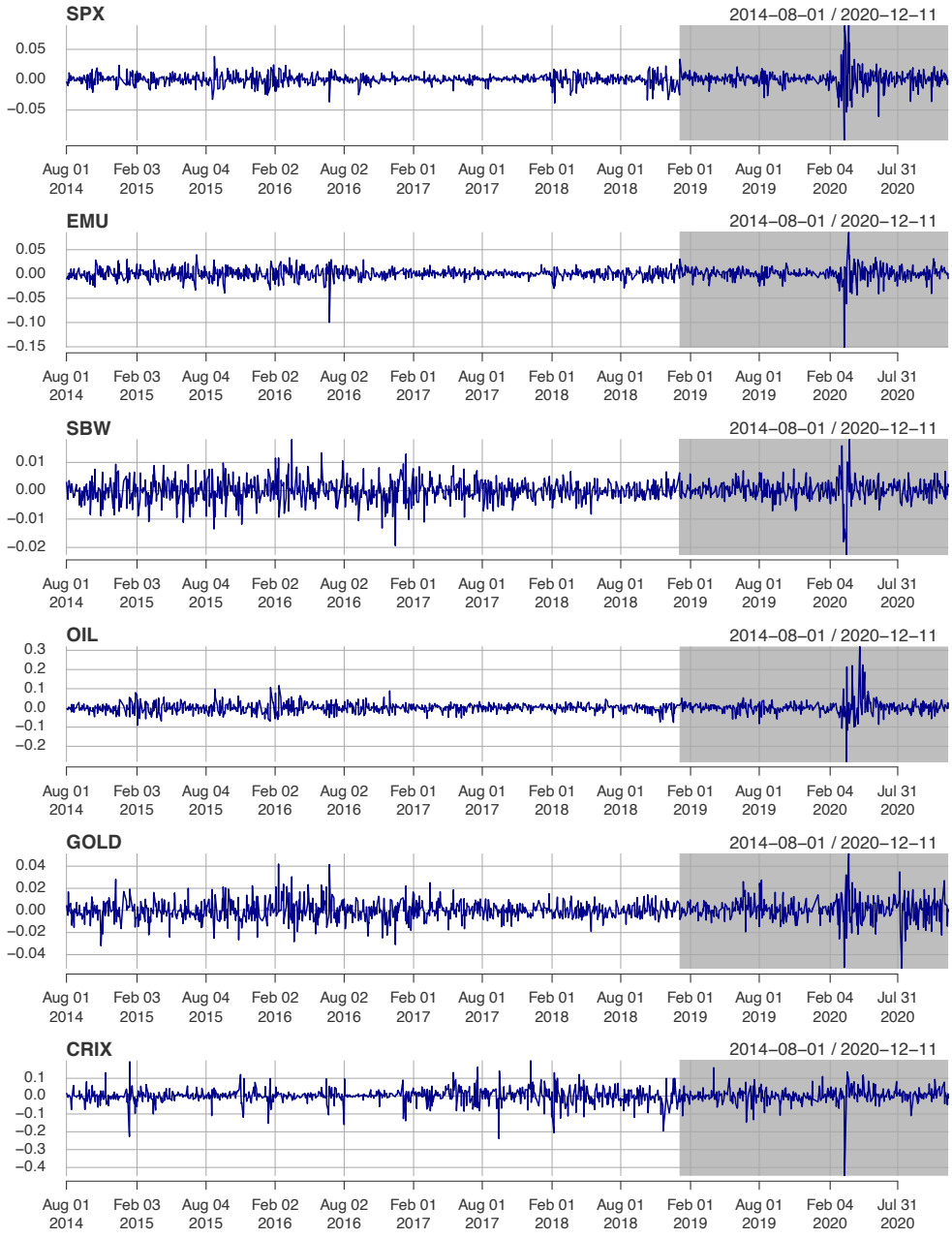

Figure G.3: Time series of daily log-returns, August 1, 2014 to December 12, 2020, for (from top to bottom) SPX, SBW, EMU, OIL, GOLD, and CRIX. The grey shaded area is the out-of-sample period ranging from the beginning of 2019 to the end of 2020.

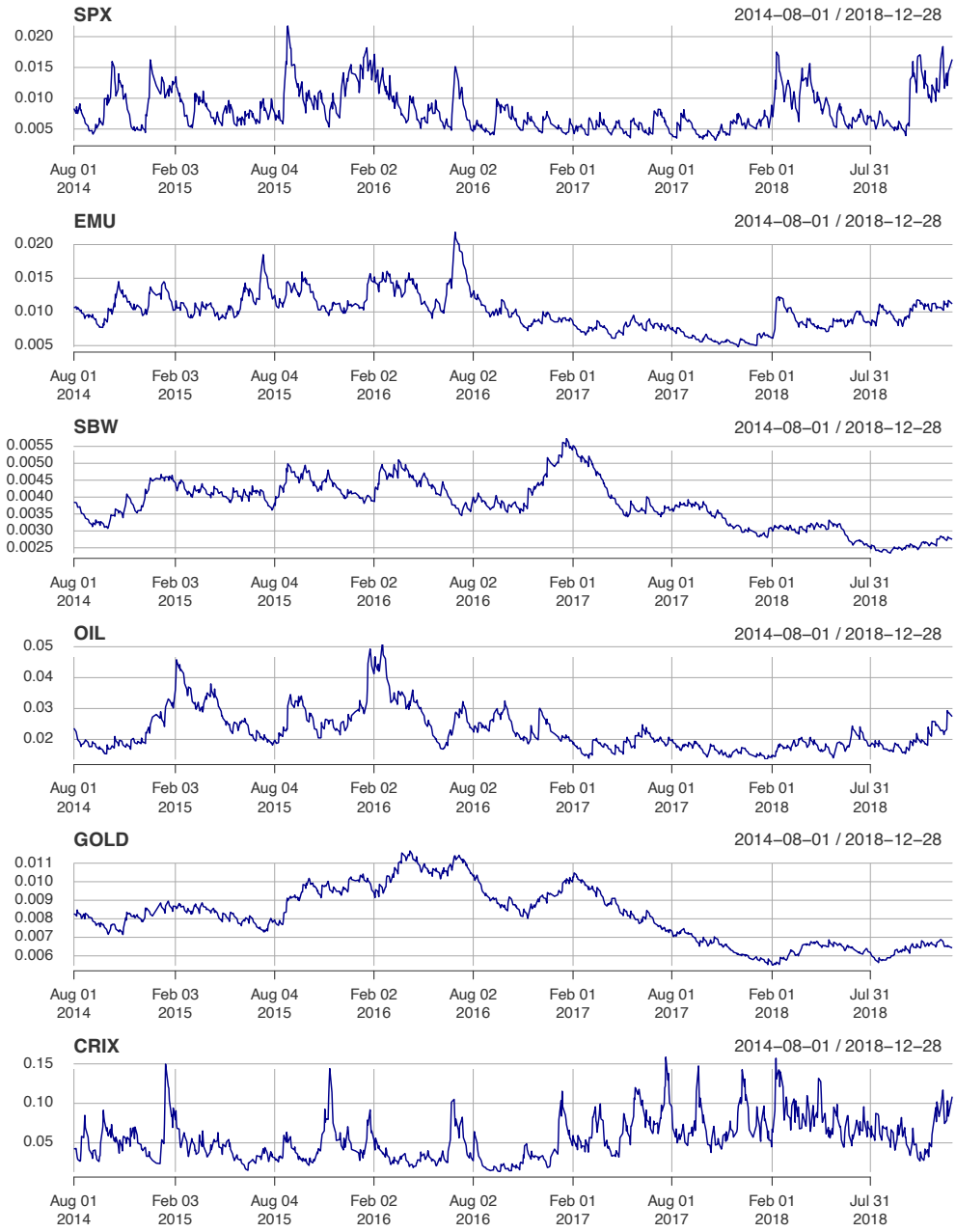

Figure G.4: Estimated volatility series using Beta-t-EGARCH.

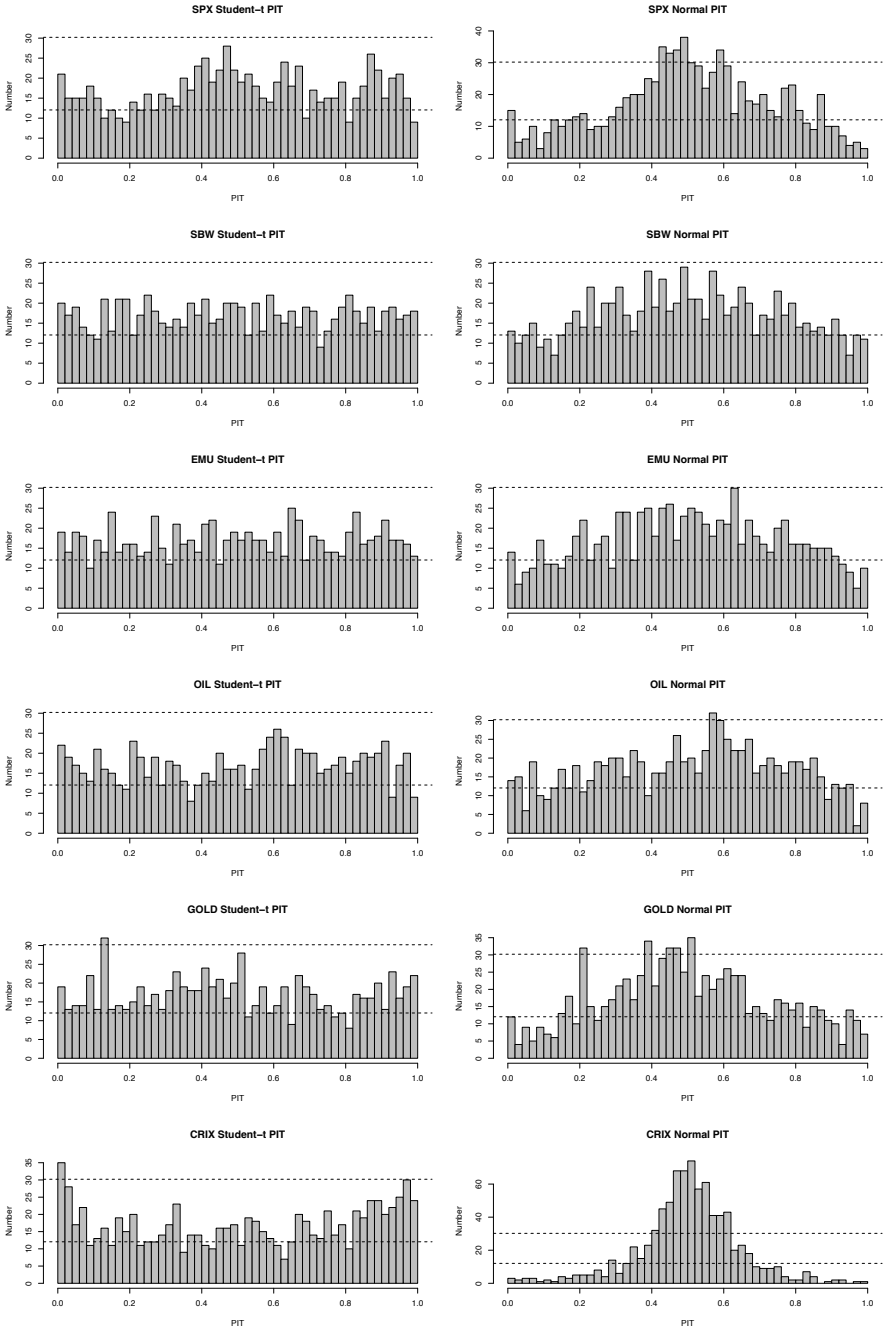

Figure G.5: Histograms of probability integral transforms (PITs) of standardized residuals of the Beta-t-EGARCH model. We use the student-t distribution with the estimated d.o.f. for figures in the left-hand column. For the figures in the right-hand column, we use the Gaussian cdf for the transformation.

References

Harvey, A. C. (2013): *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press.

Higham, N. J. (2008): *Functions of matrices: Theory and computation*. Society for Industrial and Applied Mathematics.

Linton, O. and McCrorie, J. R. (1995): "Differentiation of an exponential matrix function". *Econometric Theory* 11.5, pp. 1182–1185.

Magnus, J. R. and Neudecker, H. (2019): *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons.