Discussion Papers
Department of Economics
University of Copenhagen

No. 19-13

# Dynamic Conditional Eigenvalue GARCH

Simon Hetland, Rasmus Søndergaard Pedersen & Anders Rahbek

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

# Dynamic Conditional Eigenvalue GARCH

Simon Hetland[†], Rasmus Søndergaard Pedersen[†] and Anders Rahbek[†]

Abstract

In this paper we consider a multivariate generalized autoregressive conditional heteroskedastic (GARCH) class of models where the eigenvalues of the conditional covariance matrix are time-varying. The proposed dynamics of the eigenvalues is based on applying the general theory of dynamic conditional score models as proposed by Creal, Koopman and Lucas (2013) and Harvey (2013). We denote the obtained GARCH model with dynamic conditional eigenvalues (and constant conditional eigenvectors) as the $\lambda$-GARCH model. We provide new results on asymptotic theory for the Gaussian QMLE, and for testing of reduced rank of the (G)ARCH loading matrices of the time-varying eigenvalues. The theory is applied to US data, where we find that the eigenvalue structure can be reduced similar to testing for the number in factors in volatility models.

Keywords: Multivariate GARCH; GO-GARCH; Reduced Rank; Asymptotic Theory.
JEL: C32; C51; C58.

[†]Department of Economics, University of Copenhagen, Denmark. Correspondence to: Anders Rahbek: anders.rahbek@econ.ku.dk This research was supported by the Danish Council for Independent Research (DSF Grant 7015-00028B).

# 1 Introduction

In this paper we consider $p$-dimensional multivariate generalized autoregressive conditional heteroskedastic (GARCH) models where the eigenvalues $(\lambda_{1t}, ..., \lambda_{pt})$ of the conditional covariance matrix of the $p$-dimensional vector $X_t$ (of returns) are modelled as time-varying. The proposed dynamics of the eigenvalues $(\lambda_{1t}, ..., \lambda_{pt})$ is based on utilizing the general theory of dynamic conditional score models for time-varying parameters as proposed by Creal, Koopman and Lucas (2013) and Harvey (2013). We denote the obtained GARCH model with dynamic conditional eigenvalues (and constant conditional eigenvectors) as the $\lambda$-GARCH model.

We consider in detail the cases where (the returns) $X_t$ are assumed to be multivariate conditionally Gaussian and Student's $t_v$-distributed respectively, which constitute the conditional distributions most widely applied in empirical modelling of time-varying covariances. By definition, both specifications imply a rich and general dynamic structure for the evolution of the eigenvalues. Specifically, in the conditional Gaussian case, the resulting dynamics of the eigenvalues of the $\lambda$-GARCH model is an extended version of the generalized orthogonal GARCH (GO-GARCH) model of van der Weide (2002). Here the $\lambda$-GARCH model extends the GO-GARCH as the spill-over effects are allowed more degrees of flexibility, similar to the extended version of the constant conditional correlation (ECCC) GARCH model in Jeantheau (1998) which generalizes the CCC-GARCH model of Bollerslev (1990). On the other hand, in the conditionally $t$-distributed case, the dynamics of the $\lambda$-GARCH model generalizes and extends the univariate $\beta$-$t$-GARCH model of Harvey (2013) and Harvey and Chakravarty (2008) to the multivariate case, where the "ARCH" effects are time-varying, while the "GARCH" effects remain constant. One may note that the score approach is also used for considering time-varying correlations – as opposed to time-varying eigenvalues – in Creal, Koopman and Lucas (2011), where the DCC-GARCH model of Engle (2002) is considered under the assumption of a conditional $t$-distribution of returns $X_t$.

As demonstrated in the empirical illustration, the dynamic specification in the $\lambda$-GARCH class allows one to impose hypotheses on the inter-action between linear combinations of the eigenvalues. In particular, for the returns on three major US bank shares, we find that while we reject constancy of all (three) eigenvalues, there is one linear combination of the eigenvalues which appear constant. Equivalently, the implied reduced rank structure of the (G)ARCH loading matrices, indicates that there are two linear combinations of the eigenvalues which drive the conditional volatility of $X_t$. Thus we are able to disentangle time-varying linear combinations of the eigenvalues, or factors, from time-invariant factors which drive the dynamics of the conditional covariance, see also Lanne and Saikkonen (2007) and Dovonon and Renault (2013).

In terms of inference and asymptotic theory, we provide a full asymptotic theory for the Gaussian-based quasi maximum likelihood estimator (QMLE) of the (vector) parameter of the $\lambda$-GARCH model. We provide conditions for strict stationarity, ergodicity, and finite moments of $X_t$, and present primitive sufficient condition for consistency and asymptotic normality of the QMLE relying on only finite second-order moments of $X_t$. Simulations indicate that while sufficient, second-order moments may not even be necessary as the estimator is well-behaved even when relaxing the condition of finite unconditional variance, similar to results in the univariate analysis of GARCH models, see also Jensen and Rahbek (2004). The asymptotic results are new, and while the arguments applied for establishing limiting distributions are based on classic likelihood expansions, a novel result on identification is given, which is needed for establishing consistency of the QMLE estimator.

Moreover, testing reduced rank in the context of multivariate GARCH models is nonstandard as it involves non-identified parameters under the hypothesis – see Pedersen and Rahbek (2019) for a discussion of the univariate case – and we discuss the general theory applicable for our empirical illustration. In particular, we derive the limiting distribution of the sup likelihood ratio (supLR) test statistic for the case of zero rows, and hence reduced rank, of the (G)ARCH loading matrices, while we for the more general case propose a bootstrap based approach, see also Cavaliere, Nielsen, Pedersen and Rahbek (2019).

Existing theory for the classic (non-extended) multivariate GO-GARCH model typically rely on two (or, three) step estimators. For the multiple step estimators, essentially, in a first step the unconditional covariance matrix is estimated, which is then kept fixed in the next estimation step(s), where the remaining dynamic GARCH parameters are estimated, see Fan, Wang and Yao (2008) and Boswijk and van der Weide (2011) and the references therein. In contrast, we consider here joint one-step estimation of all parameters, which in particular requires the mentioned identification result as the unconditional covariance, and hence eigenvectors, are not fixed in a first estimation step. In terms of asymptotic theory for two, or multiple, step estimators in other multivariate GARCH type models, Pedersen and Rahbek (2014) discuss this in terms of covariance targeting for the BEKK-GARCH model, while Francq, Horvath and Zakoïan (2014) discuss variance targeting for the ECCC-GARCH model. Lanne and Saikkonen (2007) consider one-step estimation of their factor GO-GARCH model. By noting that the model has a BEKK-type representation, they argue that the MLE (for the identified parameters) is consistent and asymptotically normal by referring to the theory for BEKK models derived by Comte and Lieberman (2003). We emphasize that this theory rely on the assumption of finite higher-order moments of $X_t$ (specifically, $E\|X_t\|^8 < \infty$) which is typically needed for showing asymptotic normality of the MLE for BEKK models, see also Hafner and Preminger (2009a), Avarucci, Beutner and Zaffaroni (2013), and Pedersen and Rahbek (2014). In contrast, we show that the QMLE for the $\lambda$-GARCH model is asymptotically normal under mild second order moment conditions on $X_t$.

The paper is structured as follows. Section 2 defines the $\lambda$-GARCH model for the case of conditional Gaussianity and conditional Student's $t$ distributed returns. In Section 3, the stochastic properties of the $\lambda$-GARCH process is discussed, and asymptotic theory for the QMLE is given. In Section 4 testing of reduced rank ARCH and GARCH loading matrices is discussed and Section 5 contains an empirical example with US data. The Appendix contains mathematical proofs (Appendix A), details on hypothesis testing (Appendices B and C), and a short simulation study on the finite sample properties of the QMLE (Appendix D).

## 1.1 Notation

Some notation used throughout the paper. For $p \in \mathbb{N}$, $I_p$ denotes the $(p \times p)$ identity matrix and $0_{n\times p}$ denotes a $n \times p$ matrix of zeros (and $0_n = 0_{n\times 1}$). For a $p$-dimensional vector $x$, $\operatorname{diag}(x) = \operatorname{diag}((x_i)_{i=1}^p)$ is a diagonal matrix with $x$ on the diagonal. Furthermore, denote by $\rho(A)$ the spectral radius of any square matrix $A$. We use $||\cdot||$ to denote the Euclidean matrix norm. Moreover, $A \odot B$ denotes the Hadamard product, while $A \otimes B$ denotes the Kronecker product of $A$ and $B$ of suitable dimensions. We set $A^{\odot 2} = A \odot A$ and $A^{\otimes 2} = (A \otimes A)$. Finally, let $\stackrel{p}{\to}$, $\stackrel{d}{\to}$ and $\stackrel{w}{\to}$ denote convergence in probability, in distribution and weakly respectively. Unless stated otherwise, all limits are taken as the sample size $T \to \infty$.

# 2 Score Driven Conditional Eigenvalues | $\lambda$-GARCH

We consider a class of multivariate conditionally heteroskedastic models where the eigenvalues of the conditional covariance matrix are allowed to be time-varying, where we apply the approach of Creal, Koopman and Lucas (2011) and Harvey (2013) to arrive at dynamic specifications of the time-varying eigenvalues under different distributional assumptions on the innovations.

Let $X_t$ be a $p$-dimensional vector of observed variables (returns, say), $X_t \in \mathbb{R}^p$ for $t = 1, ...., T$. Define the information at time $t$, $\mathcal{F}_t$ as the $\sigma$-algebra generated by the past variables, $\mathcal{F}_t = \sigma(X_i : i \leq t)$, and let $f(X_t|\mathcal{F}_{t-1})$ denote the conditional density of $X_t$ given $\mathcal{F}_{t-1}$. Assume with no loss of generality that the conditional mean $E(X_t|\mathcal{F}_{t-1})$ is zero, $E(X_t|\mathcal{F}_{t-1}) = 0$, and that the conditional distribution of $X_t$, or $f(X_t|\mathcal{F}_{t-1})$, can be characterized in terms of the time-varying conditional covariance matrix $\Omega_t = E(X_t X_t|\mathcal{F}_{t-1})$ in additional to (constant) distributional shape parameters.

The conditional covariance matrix $\Omega_t$ is stated in terms of time-varying conditional eigenvalues $(\lambda_{i,t})_{i=1}^p$ and corresponding $p-$dimensional constant conditional eigenvectors $(v_i)_{i=1}^p$. That is,

$$\Omega_t = V\Lambda_t V',$$

with $V = (v_1, ..., v_p)$ and $\Lambda_t = \text{diag}\big((\lambda_{i,t})_{i=1}^p\big)$. By definition the eigenvectors are orthogonal, such that $V'V = VV' = I_p$, while $\lambda_{i,t} > 0$ (almost surely) for $i = 1, ..., p$ and for all $t$. With

$$\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{p,t})'$$

the vector of eigenvalues, we note that $f(X_t|\mathcal{F}_{t-1})$ may be indexed by $\lambda_t$, and we write henceforth

$$f(X_t|\mathcal{F}_{t-1}) = f(X_t|\lambda_t).$$

The dynamics of the time-varying eigenvalues $\lambda_t$ is given by the score updating equation, see Creal *et al.* (2011),

$$\lambda_t = W + \mathcal{A}s_{t-1} + \mathcal{B}\lambda_{t-1}, \tag{1}$$

where $W$ is a $p$-dimensional vector of constants and $\mathcal{A}$ and $\mathcal{B}$ are general $(p \times p)$ coefficient matrices. The $p$-dimensional (score) vector $s_t$ is defined as the score of the log-density $\log f(\cdot|\lambda_t)$ with respect to $\lambda_t$, up to an appropriate scaling. That is, the score contribution in the dynamics is given by,

$$s_t = S_t \frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t}, \tag{2}$$

with $S_t$ an appropriate scaling matrix, which here in line with existing literature on score driven models is set to the inverse of the (conditional) Fisher information matrix, i.e.

$$S_t = \left( E\left[ \frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t} \frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t'} \middle| \mathcal{F}_{t-1} \right] \right)^{-1}. \tag{3}$$

Below we consider the implied $\lambda$-GARCH models when $f(\cdot|\lambda_t)$ is assumed to be one of the two dominating densities in the multivariate GARCH literature; the multivariate Gaussian and Student's $t$ respectively. These yield fundamentally different dynamics of the eigenvalues as clear from the next.

## 2.1 Conditional Gaussian Distribution

Consider the case of conditional normality of $X_t$, such that the conditional density $f(X_t|\lambda_t)$ is given by,

$$f(X_t|\lambda_t) = (2\pi)^{-p/2} \det(\Omega_t)^{-1/2} \exp\left(-X_t'\Omega_t^{-1}X_t/2\right).$$

Using the definitions in (1)–(3), give upon tedious calculations that the implied dynamics of $\lambda_t$ can be represented in a multivariate GARCH-type form,

$$\lambda_t = W + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1},$$

where $W$ is a $p$-dimensional vector, $A = \mathcal{A}$ and $B = \mathcal{B} - \mathcal{A}$, and where we restrict the $(p \times p)$ matrices $A$ and $B$ to have non-negative entries.

Note that, for each $i$, the time-varying positive eigenvalue $\lambda_{i,t}$ is allowed to depend on all of the orthogonal linear combinations $v_j'X_{t-1}$, where $\mathrm{Cov}\big(v_j'X_{t-1}, v_k'X_{t-1}|\mathcal{F}_{t-1}\big) = 0$ (and hence $\mathrm{Cov}\big(v_j'X_{t-1}, v_k'X_{t-1}\big) = 0$) for all $j \neq k$. In addition, our proposed $\lambda$-GARCH model allows $\lambda_{i,t}$ to depend on all entries of $\lambda_{t-1}$. In that sense the Gaussian score-driven eigenvalue model is a generalization the GO-GARCH models considered by Fan *et al.* (2008) and Boswijk and van der Weide (2011). Finally, we stress that our proposed parametrizations appeal to estimating all model parameters simultaneously, and not as is common in two, or more steps.

Specifically, Boswijk and van der Weide (2011) consider the GO-GARCH model

$$X_t = V\Lambda_t^{1/2}\eta_t, \quad \eta_t \sim i.i.d.(0, I_p),$$

with $\Lambda_t = \mathrm{diag}\big((\lambda_{i,t})_{i=1}^p\big)$ satisfying[1], with $B_d$ a $(p \times p)$ diagonal matrix,

$$\lambda_t = (I - A - B_d) + A\left(V'X_{t-1}\right)^{\odot 2} + B_d\lambda_{t-1}. \tag{4}$$

Moreover, Boswijk and van der Weide (2011) assume that the matrix $V = (V_1, \ldots, V_p)$ is non-singular with polar decomposition

$$V = CR,$$

such that $C$ is positive definite and $R$ is orthogonal. Lanne and Saikkonen (2007) considered an identical model, but with the additional restriction that some row in $A$ and (the diagonal) $B_d$ is zero, and hence allowing for constant conditional eigenvalues $\lambda_{i,t}$. We discuss in Section 4 testing for reduced rank of $A$ and $B$ in the $\lambda$-GARCH model, for which the zero row

[1] To be precise, the authors only state that $\lambda_{it}$ is "assumed to follow a GARCH-type structure" [p.119], but the following specification is the one considered in their empirical application.

restriction is a special case. The model in Boswijk and van der Weide (2011) is closely related to the model considered by Fan *et al.* (2008) who, identically to our approach, let $V$ be orthogonal but with $\Lambda_t$ defined by (4) such that the condition that $E(X_t X_t') = I_p$ (or, equivalently, standardized returns) is imposed.

## 2.2 Conditional Student's $t$-Distribution

Consider here the case where the conditional distribution of $X_t$ is a standardized Student's $t$ distribution with $\nu > 2$ degrees of freedom. In this case the conditional density is given by

$$f(X_t|\lambda_t) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} [(\nu-2)\pi]^{p/2} \det(\Omega_t)^{1/2} \left[1 + \frac{X_t'\Omega_t^{-1}X_t}{\nu-2}\right]^{-(\nu+p)/2},$$

where $\Gamma(\cdot)$ is the Gamma function. In line with the Gaussian case in Section 2.1, the bivariate case of the Student's score dynamics can be represented as

$$\lambda_t = W + A_t(V'X_{t-1})^{\odot 2} + B\lambda_{t-1}.$$

Here $W$ is as before while

$$A_t = \frac{2w_t}{(\kappa^2-\gamma^2)}\mathcal{A}\begin{pmatrix} \kappa & -\gamma\frac{\lambda_{1,t-1}}{\lambda_{2,t-1}} \\ -\gamma\frac{\lambda_{2,t-1}}{\lambda_{1,t-1}} & \kappa \end{pmatrix} \quad \text{and} \quad B = \left[\mathcal{B} - \left(\frac{\nu+4}{\nu-2}\right)\mathcal{A}\right],$$

with $\kappa = 3(v+2)/(v+4) - 1$, $\gamma = (v+2)/(v+4) - 1$. Moreover, the time-varying "weights" $w_t$ of the ARCH-loadings are given by

$$w_t = \frac{1+2/\nu}{1+\nu^{-1}[y_{1,t-1}^2/\lambda_{1,t-1} + y_{2,t-1}^2/\lambda_{2,t-1} - 2]},$$

with $(y_{1,t}, y_{2,t}) = X_t'V$.

We note that, similar to the Gaussian case, one may view the dynamics of $\lambda_t$ as GARCH-type dynamics where the "GARCH" coefficients $B$ – as for the Gaussian case – are constant, while the "ARCH" coefficients, $A_t$, are time-varying and stochastic. Note that one obtains the Gaussian case by setting $\nu^{-1} = 0$. Also, note that for the one-dimensional case, $p = 1$, we obtain the Beta-$t$-GARCH considered in Harvey (2013, Ch.4.7). While the bivariate (and univariate) case has a somewhat simple structure, the general case of $p > 2$ has a less transparent representation. Specifically, following Creal *et al.* (2011,Theorem 1), the score and scaling matrix are given respectively by

$$\frac{\partial \log f(X_t|\mathcal{F}_{t-1})}{\partial \lambda_t} = \frac{1}{2}\Psi_t'(V\Lambda_t^{-1}V')^{\otimes 2}\left[w_t X_t^{\otimes 2} - V^{\otimes 2}\text{vec}(\Lambda_t)\right],$$

$$S_t = \frac{1}{4}\Psi_t'(V\Lambda_t^{-1/2}V')^{\otimes 2}\left[gG - \text{vec}(I_p)\text{vec}(I_p)'\right](V\Lambda_t^{-1/2}V')^{\otimes 2}\Psi_t,$$

where $\Psi_t = V^{\otimes 2}\partial\text{vec}(\Lambda_t)/\partial\lambda_t'$, $w_t = (v+p)/(v-2+X_t'V\Lambda_t^{-1}VX_t)$, and $g = (v+p)/(v+2+p)$. Moreover, $G = E[(z_t z_t')^{\otimes 2}]$ with $z_t \sim N(0, I_p)$.

# 3 Properties and Estimation of the $\lambda$-GARCH Model

In the remainder of the paper we focus on the Gaussian case in Section 2.1, and study quasi-likelihood inference. In particular, we provide sufficient conditions for strict stationarity and state primitive conditions for strong consistency and asymptotic normality of the one-step quasi-maximum likelihood estimator (QMLE) for all parameters.

The $\lambda$-GARCH model may be summarized as,

$$X_t = V\Lambda_t^{1/2}\eta_t, \quad \Lambda_t = \text{diag}\left((\lambda_{i,t})_{i=1}^p\right), \quad V'V = VV' = I_p, \tag{5}$$

$$\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{p,t})' = W + A(V'X_{t-1})^{\odot 2} + B\lambda_{t-1}, \tag{6}$$

with $\eta_t$ i.i.d$(0, I_p)$. The parameters of the model are given by the $p$-dimensional vector $W = (\omega_1, ..., \omega_p)'$ with strictly positive entries, $\omega_i > 0$ for $i = 1, 2, ..., p$ and the $(p \times p)$ matrices $A = (\alpha_{ij})_{i,j=1,...,p}$ and $B = \left(\beta_{ij}\right)_{i,j=1,..,p}$ with non-negative entries, $\alpha_{ij}, \beta_{ij} \geq 0$.

Additionally, the constant conditional eigenvectors $V$ are parametrized by $\phi$, which is a $p(p-1)/2$-dimensional vector $\phi = (\phi_{12}, \ \ldots, \ \phi_{(p-1)p})'$. More specifically, for the case of $p = 3$,

$$V(\phi) = \begin{pmatrix} \cos(\phi_{12}) & \sin(\phi_{12}) & 0 \\ -\sin(\phi_{12}) & \cos(\phi_{12}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\phi_{13}) & 0 & \sin(\phi_{13}) \\ 0 & 1 & 0 \\ -\sin(\phi_{13}) & 0 & \cos(\phi_{13}) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi_{23}) & \sin(\phi_{23}) \\ 0 & -\sin(\phi_{23}) & \cos(\phi_{23}) \end{pmatrix},$$

while for the general case the $(p \times p)$ dimensional $V$ matrix is defined in terms of so-called rotation matrices $R(i,j) = (R(i,j)_{kl})_{k,l=1,...,p}$ as applied in van der Weide (2002) for the GO-GARCH model. That is, $V = \prod_{i=1}^{p-1}\prod_{j=i+1}^{p} R(i,j)$, where

$$R(i,j)_{kk} = 1 \text{ if } k \neq i,j, \qquad R(i,j)_{kl} = 0 \text{ if } k \neq l \text{ and } k \neq i,j,$$

$$R(i,j)_{ii} = R(i,j)_{jj} = \cos(\phi_{ij}), \text{ and } R(i,j)_{ij} = -R(i,j)_{ji} = \sin(\phi_{ij}).$$

For the estimation, or statistical analysis, we assume the $\eta_t$ are Gaussian distributed, while this assumption is relaxed when studying the probabilistic properties of $X_t$ as well as the asymptotic properties of the resulting Gaussian-based quasi likelihood estimators (QMLEs).

## 3.1 Stochastic properties

For the stochastic properties of $X_t$ satisfying equations (5)-(6), we note that $V'X_t$ satisfies the stochastic recursion,

$$V'X_t = \Lambda_t^{1/2}\eta_t, \text{ with } \lambda_t = W + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1}, \tag{7}$$

such that the rich literature on stochastic recursions can be applied in order to state conditions for strict stationarity and ergodicity as well as conditions for finite moments of $X_t$. To see this, rewrite the dynamics of $\lambda_t$ in (7) as the stochastic recurrence equation,

$$\lambda_t = W + \Phi_{t-1}\lambda_{t-1} \tag{8}$$

where $\Phi_t$ are i.i.d. random matrices,

$$\Phi_t = A \text{ diag}\left((\eta_{i,t}^2)_{i=1}^p\right) + B, \tag{9}$$

with $\Phi_t$ and $\lambda_t$ independent. By Francq and Zakoïan (2019, Theorem 10.6 and Corollary 10.2) and Pedersen (2017, Lemmas B.5 and B.6) we immediately have the following result.

**Theorem 3.1.** *The process $(X_t : t \in \mathbb{Z})$ obeying (5)-(6) is strictly stationary and ergodic if and only if $\xi < 0$, where $\xi$ is the top Lyapunov coefficient of $(\Phi_t : t \in \mathbb{Z})$ defined by*

$$\xi = \lim_{n\to\infty} n^{-1}E(\log||\prod_{t=1}^{n}\Phi_t||), \tag{10}$$

*with $\Phi_t$ defined in (9). The strictly stationary and ergodic process has $E\|X_t\|^s < \infty$ for some $s > 0$. Moreover, for $k \in \mathbb{N}$, $E||X_t||^{2k} < \infty$ if and only if $\{\rho(E(H_t^{\otimes k})) < 1$ and $E||\eta_t||^{2k} < \infty\}$.*

**Remark 1.** *Notice that a necessary and sufficient condition for finite second order moments, $E||X_t^{\odot 2}|| < \infty$, of the strictly stationarity and ergodic process $(X_t : t \in \mathbb{Z})$ is that $\rho(A+B) < 1$. In this case, the unconditional variance of the process is $E\left(X_tX_t'\right) = E\left(\Omega_t\right) = E(V\Lambda_tV') = V\left(diag\{(I_p - A - B)^{-1}W\}\right)V'$.*

### 3.2 Quasi-Maximum Likelihood Estimation

The parameters of the $\lambda$-GARCH model in (5)-(6) are given by,

$$\theta = (W', \mathrm{vec}(A)', \mathrm{vec}(B)', \phi')',$$

with parameter space $\Theta$,

$$\Theta = \Theta_\omega \times \Theta_A \times \Theta_B \times \Theta_\phi.$$

Here $\Theta_W = [\omega_L, \omega_U]^p$ for some $0 < \omega_L < \omega_U < \infty$, $\Theta_A := [0, \alpha_U]^{p^2}$ for some $0 < \alpha_U < \infty$, $\Theta_B \subset \mathbb{R}_+^{p^2}$ such that $\sup_{\mathrm{vec}(B)\in\Theta_B} \rho(B) < 1$, and $\Theta_\phi = [0, \pi/2]^{(p+1)p/2}$. We make the following standard assumption:

**Assumption 3.1.** *The true value of the parameter vector $\theta_0 \in \Theta$ and $\Theta$ is compact.*

Given a realization $(X_t : t = 0, 1, \ldots, T)$ of the $\lambda$-GARCH process in (5)-(6), the Gaussian quasi-maximum likelihood estimator (QMLE), $\hat{\theta}_T$, for $\theta$ is defined as

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} L_T(\theta),$$

where the log-Gaussian likelihood function is given by,

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = \log\det(\Omega_t(\theta)) + X_t'\Omega_t^{-1}(\theta)X_t, \tag{11}$$

$$\Omega_t(\theta) = V(\theta)\Lambda_t(\theta)V(\theta)', \quad \Lambda_t(\theta) = \mathrm{diag}(\lambda_t(\theta)), \tag{12}$$

$$\lambda_t(\theta) = W(\theta) + A(\theta)\left(V(\theta)'X_{t-1}\right)^{\odot 2} + B(\theta)\,\lambda_{t-1}(\theta), \quad t = 1, \ldots, T, \tag{13}$$

with $\lambda_0(\theta) = \bar{\lambda}_0$ fixed and with strictly positive entries. Throughout we make the following assumption about the data generating process $(X_t : t \in \mathbb{Z})$ where $A_0 = A(\theta_0)$ and similarly for the remaining true parameter values.

**Assumption 3.2.** *Assume that $\xi_0 < 0$, where $\xi$ is defined in (10), such that the process $(X_t : t \in \mathbb{Z})$ is stationary and ergodic with $E\|X_t\|^s < \infty$ for some $s > 0$.*

Lastly, in order to show that the QMLE is strongly consistent, we make the following identification assumptions.

**Assumption 3.3.** *Assume for the i.i.d.$(0, I_p)$ sequence $(\eta_t : t \in \mathbb{Z})$ that $\eta_{it}$ and $\eta_{jt}$ are independent for all $i \neq j$, $i, j = 1, ..., p$. Moreover, assume that $\eta_{i,t}^2$ is non-degenerate for $i = 1, \ldots, p$.*

**Assumption 3.4.** *Assume that the* $(p \times 2p)$ *dimensional matrix* $[A_0, B_0]$ *has full rank* $p$*. Moreover, with* $z \in \mathbb{C}$ *and* $\theta \in \Theta$*, assume that the polynomials* $A(\theta) z$ *and* $I_p - B(\theta) z$ *satisfy that* $(I_p - B(\theta) z)^{-1} A(\theta) z = (I_p - B_0 z)^{-1} A_0 z$ *implies* $\theta = \theta_0$*.*

**Assumption 3.5.** *With* $\theta_0$ *the true value of* $\theta$*, and for any* $\theta \in \Theta$*, let* $\tilde{V} = V_0' V$ *where* $V = V(\theta)$ *and* $V_0 = V(\theta_0)$*. Assume that for some* $j \in \{1, \ldots, p\}$*,* $\tilde{V}_{jj} \neq 0$*. Moreover, with* $A_j$ *and* $A_{j,0}$ *the* $j$*th row of* $A(\theta)$ *and* $A_0$*, and*

$$\gamma_t(j) = \left( A_j (V' X_t)^{\odot 2} - A_{j,0} (V_0' X_t)^{\odot 2} \right), \tag{14}$$

*assume that* $\gamma_t(j)$ *conditional on* $\mathcal{F}_{t-1}^{\eta} = \sigma\{\eta_s, s < t\}$ *is degenerate implies that* $V = V_0$*. Finally, assume that*

$$V = V_0 \text{ implies that } \phi = \phi_0. \tag{15}$$

Assumptions 3.3 and 3.4 are standard and in line with existing literature on two-step estimators as well as theory for ECCC-GARCH type models as in Francq and Zakoïan (2019). Assumption 3.5 is new and specifically ensures that the rotation parameters $\phi$ are identified. For the case where $\phi_0 \in \text{int}\Theta_\phi$ we have the following result:

**Lemma 1.** *With* $\theta_0 \in \text{int}\Theta$ *and* $\eta_t$ *i.i.d.*$N_p(0, I_p)$*, then (14) holds. If* $\phi_0 \in \text{int}\Theta_\phi$*, then (15) holds.*

**Remark 2.** *For our choice of parametrization of* $V$*, we note that the first column of* $V$ *is given by* $(V_{11}, \ldots, V_{p1})'$*, where* $V_{11} = \prod_{i=1}^{p-1} \cos\phi_i$ *and* $V_{j1} = -\prod_{i=1}^{p-j} \cos\phi_{p-i} \sin\phi_{j-1}$*,* $j = 2, \ldots, p$*. Note that for any* $\phi \in \Theta_\phi$ *there exits a* $j$ *such that* $V_{j1} \neq 0$*. Moreover, for any* $j$*,* $V_{j1} \neq 0$ *on* $\text{int}\Theta_\phi$*. Hence,* $\tilde{V}_{11} = \sum_{j=1}^{p} V_{0,j1} V_{j1} \neq 0$ *on* $\Theta_\phi$ *if* $\phi_0 \in \text{int}\Theta_\phi$*.*

We have the following result on strong consistency of the QMLE:

**Theorem 3.2** (Consistency)**.** *Under Assumptions 3.1-3.5,* $\hat{\theta}_T \to \theta_0$ *almost surely.*

In order to show that the QMLE is asymptotically Gaussian, we make some additional assumptions.

**Assumption 3.6.** *The true value of the parameter vector* $\theta_0 \in \text{int}\Theta$*.*

**Assumption 3.7.** *The data-generating process satisfies that* $E\|\eta_t\|^4 < \infty$ *and* $E\|X_t\|^{2+\epsilon} < \infty$ *for some* $\epsilon > 0$*.*

**Assumption 3.8.** *The matrix* $A_0$ *has a row with a unique entry.*

The assumptions that $\theta_0$ is an interior point and that $\eta_t$ has finite fourth-order moments are standard. The assumption of finite second-order moments of $X_t$ is used to show that the expectation of the third-order partial derivatives of the log-likelihood contribution is finite on a (suitable) neighborhood around $\theta_0$ in the proof of Lemma A.5 in Appendix A.5. Specifically, the third-order derivatives contain terms essentially of the form

$$\frac{\dot{\lambda}_{s,t,i}(\theta)\dot{\lambda}_{s,t,j}(\theta)\dot{\lambda}_{s,t,k}(\theta)}{\lambda_{s,t}^3(\theta)} \times \frac{\lambda_{g,t}^{1/2}(\theta_0)\lambda_{h,t}^{1/2}(\theta_0)\eta_{h,t}\eta_{g,t}}{\lambda_{s,t}(\theta)}, \tag{16}$$

where $\eta_{s,t}$ denotes the $s$th entry of the noise $\eta_t$, $\lambda_{s,t}(\theta)$ is the $s$th entry of $\lambda_t(\theta)$ in (13), and $\dot{\lambda}_{s,t,i}(\theta) = \partial\lambda_{s,t}(\theta)/\partial\theta_i$. Any power of the first factor has finite expectation on the neighborhood, whereas for the case where $g \neq s$, it is not obvious that the second factor has finite expectation for $\theta \neq \theta_0$. On the other hand, it is straightforward to show that the fraction is (up to a scaling constant) bounded (uniformly on the neighborhood) by $\|\lambda_t(\theta_0)\|\|\eta_t\|^2$ which has finite expectation provided that $E\|X_t\|^2 < \infty$. By Hölder's inequality it then follows that (16) has finite expectation if $E\|X_t\|^{2+\epsilon} < \infty$. Simulations in Appendix D indicate that, while sufficient, the condition may not be needed in order for the QMLE to be asymptotically normal.

We note that the moment requirement is stronger than for the theory for the Gaussian-based QMLE for the ECCC-GARCH model (Francq and Zakoïan, 2012) and the factor-GARCH (Hafner and Preminger, 2009b), where only $E\|X_t\|^{\epsilon} < \infty$ for some $\epsilon > 0$ is needed. On the other hand, it is milder than the requirements of finite sixth- or eighth-order moments assumed by Hafner and Preminger (2009a) and Comte and Lieberman (2003) for the VEC and BEKK class of models, respectively.

Assumption 3.8 is used in the proof of Lemma A.4 in order to show that the expectation of the Hessian, i.e. the probability limit of $T^{-1}\partial^2 L_T(\theta_0)/\partial\theta\partial\theta'$, is invertible. Typically in the literature, the proof of invertibility relies on showing that there exists no non-zero $\gamma \in \mathbb{R}^{d_\theta}$ such that for all $t$

$$\gamma' \frac{\partial\lambda_t(\theta_0)}{\partial\theta} = 0_{p\times 1} \quad \text{almost surely.} \tag{17}$$

In much of the existing literature on multivariate GARCH models, e.g. Comte and Lieberman (2003) on BEKK models and Francq and Zakoïan (2012) ECCC models, such a property is typically verified by exploiting that, under (17), $\gamma'\partial\lambda_t(\theta_0)/\partial\theta$ is linear in $(V_0'X_{t-1})^{\odot 2}$ and $\lambda_{t-1}(\theta_0)$ and that $\theta_0$ is identified. In our model, we do not have linearity as $\gamma'\partial\lambda_t(\theta_0)/\partial\theta$ contains terms with partial derivatives with respect to the entries of $\phi$. This leads to additional considerations about invertibility of $J$, and we make the additional Assumption 3.8, see the proof of Lemma A.4 for details.

We have the following result:

**Theorem 3.3** (Asymptotic normality). *Under Assumptions 3.1-3.8,*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, J^{-1}\Sigma J^{-1}),$$

*where $J$ is an invertible matrix defined in (A.26) and $\Sigma$ is a non-negative definite matrix defined in (A.11) in the Appendix.*

A small simulation study in Appendix D illustrates that the finite-sample distribution of the QMLE is well-approximated by a normal distribution, and moreover indicate that the sufficient moment conditions can be relaxed.

Next, we consider hypothesis testing in the $\lambda$-GARCH model motivated by the idea that a few conditional time-varying linear combinations of $\lambda_t$ are driving the volatility of the $X_t$ process.

# 4 Reduced Rank of $A$ and $B$

Consider the $\lambda$-GARCH model in (5)-(6) on the form,

$$\lambda_t = W + A(V'X_{t-1})^{\odot 2} + B\lambda_{t-1}.$$

A relevant hypothesis to test is if there are no spillovers between the eigenvalues, that is if the matrices $A$ and $B$ are diagonal, similar to testing for no volatility spillovers in ECCC-GARCH models as considered by Pedersen (2017). We here take another direction and consider testing of the hypothesis that one or more *linear combinations* of $\lambda_t$ are constant. A special case of this is to test if one or more conditional *eigenvalues* are constant, similar to the test for a constant factor in the factor GO-GARCH model by Lanne and Saikkonen (2007).

The hypothesis of $(p-q)$ constant conditional linear combinations of $\lambda_t$ may be parametrized as the hypothesis $\mathsf{H}_q$ of reduced rank $q < p$ of $A$ and $B$,

$$\mathsf{H}_q : A = \gamma\alpha' \quad \text{and} \quad B = \gamma\beta'. \tag{18}$$

Here $\gamma, \alpha$ and $\beta$ are $(p \times q)$ dimensional matrices, such that $A$ and $B$ have non-negative entries. An immediate implication is indeed that the $(p-q)$ combinations $\gamma_c'\lambda_t$ are constant, where $\gamma_c$ is $(p \times p - q)$ dimensional and $\gamma_c'\gamma = 0$ with rank of $(\gamma, \gamma_c)$ equal to $p$. That is, the hypothesis is equivalent to $(p-q)$ constant conditional eigenvalue relations $\gamma_c'\lambda_t$, while the remaining $q$ relations, $\gamma'\lambda_t$ are time-varying.

In terms of testing – apart from standard identification issues related to the reduced rank as well-known from testing reduced rank in e.g. cointegrated vector autoregressive processes, see e.g. Cavaliere, Rahbek and Taylor (2012) – this raises the issue of non-identified parameters under $\mathsf{H}_q$ as addressed in Andrews (2001) for univariate GARCH models, see also Pedersen and Rahbek (2019) for GARCH models with exogenous covariates. In the $\lambda$-GARCH case the non-identified parameters appear in the GARCH loadings matrix $B$, and hence across equations which requires arguments different from the univariate cases mentioned.

To illustrate, we start out by considering in Section 4.1 a $p = 3$ dimensional model with $\gamma$ in (18) known which reduces the testing problem to that of a zero row in $A$ and $B$. Next, in Section 4.2, we discuss testing of $\mathsf{H}_q$, that is, extend the discussion to include an unknown $\gamma$ matrix (and general dimension $p$). In the empirical illustration in Section 5 we consider implementation of both cases.

## 4.1 Testing with $\gamma$ known

Consider the case of a $p = 3$ dimensional system with

$$\gamma' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

This is a special case of $\mathsf{H}_2$, as with the $(3 \times 2)$ matrices $\alpha$ and $\beta$ given by

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{21} \\ \alpha_{12} & \alpha_{22} \\ \alpha_{13} & \alpha_{23} \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \end{pmatrix},$$

one can write $A$ and $B$ as

$$A = \gamma\alpha' = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and } B = \gamma\beta' = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ 0 & 0 & 0 \end{pmatrix}.$$

We denote this hypothesis by $\mathsf{H}_2^{\dagger}$. Observe, that under $\mathsf{H}_2^{\dagger}$ the loading matrices $A$ and $B$ indeed have reduced rank (less than or equal to) $q = 2$, as induced by a zero row. Note also under $\mathsf{H}_2^{\dagger}$, $\gamma_c = (0, 0, 1)'$ such that $\gamma_c'\lambda_t = \lambda_{3t}$ is constant, while the remaining two linear combinations in $\gamma'\lambda_t = (\lambda_{1t}, \lambda_{2t})'$ are time-varying.

**Remark 3.** *The case of testing for a zero row in $A$ and $B$, or $\mathsf{H}_2^{\dagger}$, is similar to testing the* *hypothesis of weak exogeneity known from cointegration analysis, see Harbo* et al. *(1998).*

In terms of testing $\mathsf{H}_2^{\dagger}$, it follows that $\beta_{33}$ in (the unrestricted) $B$ is not identified analogous to testing of conditional homoskedasticity in GARCH models, see Andrews (2001). Moreover, for the two remaining eigenvalues $\lambda_{1t}$ and $\lambda_{2t}$ under $\mathsf{H}_2^{\dagger}$,

$$\begin{aligned}\lambda_{jt} &= \omega_1 + \sum_{i=1}^{3} \alpha_{ji} \left(V_i' X_{t-1}\right)^{\odot 2} + \sum_{i=1}^{3} \beta_{ji} \lambda_{it-1} \\ &= \left(\omega_j + \beta_{j3}\omega_3\right) + \sum_{i=1}^{3} \alpha_{ji} \left(V_i' X_{t-1}\right)^{\odot 2} + \sum_{i=1}^{2} \beta_{ji} \lambda_{it-1},\ j = 1, 2.\end{aligned}$$

Hence, in addition to $\beta_{33}$, we also see that the parameters $\beta_{13}$ and $\beta_{23}$ are non-identified under the null in the GARCH loadings matrix $B$. To address this, we proceed as in Pedersen and Rahbek (2019), and test the observationally equivalent hypothesis $\mathsf{H}_2^*$ which is given by

$$\mathsf{H}_2^* : \ \alpha_{3i} = 0 \text{ for } i = 1, 2, 3 \text{ and } \beta_{3j} = 0 \text{ for } j = 1, 2. \tag{19}$$

The idea is to apply a sup likelihood ratio (supLR) test, where the supremum is taken over the non-identified parameters $\beta_{13}, \beta_{23}$ and $\beta_{33}$.

To distinguish the non-identfied parameters from the identified, partition the parameters as $(\theta', \delta')'$, with $\theta = \left((\omega_i)_{i=1}^3, (\alpha_{ij})_{i,j=1}^3, \left(\beta_{ij}\right)_{i=1,2,3\ j=1,2}, (\phi_i)_{i=1}^3\right)'$ and $\delta = (\beta_{i3})_{i=1}^3$. The parameter space is given by the product $\Theta \times \Theta_{\text{sup}}$, where $\Theta$ and $\Theta_{\text{sup}}$ are compact. Lastly, consider the parameter space for $\theta$ as restricted by $\mathsf{H}_2^*$, i.e.

$$\Theta^* = \{\theta \in \Theta : \alpha_{3i} = 0 \text{ for } i = 1, 2, 3 \text{ and } \beta_{3j} = 0 \text{ for } j = 1, 2\}.$$

The test relies on estimating $\theta$ restricted and unrestricted for a given $\delta \in \Theta_{\text{sup}}$, i.e. let

$$\tilde{\theta}_{T,\delta} = \arg\max_{\theta \in \Theta^*} L_T(\theta, \delta) \quad \text{and} \quad \hat{\theta}_{T,\delta} = \arg\max_{\theta \in \Theta} L_T(\theta, \delta), \quad \text{for } \delta \in \Theta_{\text{sup}}. \tag{20}$$

The supLR statistic is given by

$$\sup \mathrm{LR}_T(\mathsf{H}_2^*) = \sup_{\delta \in \Theta_{\text{sup}}} L_T\left(\hat{\theta}_{T,\delta}, \delta\right) - \sup_{\delta \in \Theta_{\text{sup}}} L_T\left(\tilde{\theta}_{T,\delta}, \delta\right). \tag{21}$$

Under regularity conditions given in Appendix B the statistic converges in distribution to a limiting distribution $\mathcal{L}$,

$$\sup \mathrm{LR}_T(\mathsf{H}_2^*) \xrightarrow{d} \mathcal{L}, \tag{22}$$

with $\mathcal{L}$ given by (B.41). Also in Appendix B the implementation of the asymptotic test is discussed which is applied in Section 5.

**Remark 4.** *The key conditions for (22) as given in Appendix B are: (i) that $\tilde{\theta}_{T,\delta}$ and $\hat{\theta}_{T,\delta}$ are consistent for $\theta_0$ for any $\delta \in \Theta_{\text{sup}}$, (ii) that the score as a process indexed by $\delta$ converges weakly to a Gaussian process, and (iii) that the Hessian matrix is invertible uniformly on $\Theta_{\text{sup}}$. The conditions (i) and (iii) rely on finding conditions such that $\theta_0$ is identified, whereas (ii) typically relies on showing that the score obeys a functional CLT. The latter may be shown to hold if the score process converges in finite-dimensional distribution to a Gaussian vector, and that the score process is tight, see e.g. Pedersen and Rahbek (2019, proof of Lemma A.3). In line with Pedersen and Rahbek (2019), one may need stronger moment conditions than the ones in Assumption 3.7 in order to prove tightness. Likewise, due to the fact that $\theta_0$ is a boundary point of $\Theta$, it may require higher-order moments of $X_t$ in order so show that ratios of the type (16) have finite expectation, similar to Francq and Zakoïan (2009) and Pedersen (2017) where finite sixth-order moments are imposed.*

## 4.2 The general case of reduced rank $A$ and $B$ matrices

Next consider the general case $\mathsf{H}_q$ of reduced rank $q$ in the $p$-dimensional $\lambda$-GARCH model with general $\gamma, \alpha$ and $\beta$ matrices.

Observe initially that with the "ARCH" part of the restrictions in $\mathsf{H}_q$ imposed, $A = \gamma\alpha'$, and with $\bar{\gamma} = \gamma\left(\gamma'\gamma\right)^{-1}$ it holds by definition that

$$\bar{\gamma}'\lambda_t = \bar{\gamma}'W + \alpha'(V'X_{t-1})^{\odot 2} + \bar{\gamma}'B\gamma\bar{\gamma}'\lambda_{t-1} + \bar{\gamma}'B\gamma_c\bar{\gamma}_c'\lambda_{t-1},$$

$$\bar{\gamma}_c'\lambda_t = \bar{\gamma}_c'W + \bar{\gamma}_c'B\gamma\bar{\gamma}'\lambda_{t-1} + \bar{\gamma}_c'B\gamma_c\bar{\gamma}_c'\lambda_{t-1}.$$

Next, for $\bar{\gamma}_c'\lambda_t$ to be constant, $\bar{\gamma}_c'B\gamma = 0$ is needed, in which case the second equation reduces to

$$\bar{\gamma}_c'\lambda_t = \bar{\gamma}_c'W + \bar{\gamma}_c'B\gamma_c\bar{\gamma}_c'\lambda_{t-1},$$

which, similar to the $\mathsf{H}_2^*$ example, implies that the $(p-q)^2$ parameters $\bar{\gamma}_c'B\gamma_c$ are not identified. Moreover, as $\bar{\gamma}_c'\lambda_t$ are constant, also $\bar{\gamma}'B\gamma_c$ are not identified in the equation for $\bar{\gamma}'\lambda_t$. Collecting terms, as $(\gamma, \gamma_c)$ is of full rank $p$ by definition, it holds that the parameters in $\delta$ given by

$$\delta = B\gamma_c \quad (p \times (p-q))$$

are not identified under the null. One may therefore consider a sup-based testing approach keeping $\delta$ fixed, and, in principle, a supLR test statistic similar to (21) can be computed. However, the fact that $\gamma$ is unknown means that a reparametrization is needed to ensure identification as well as variational independence of the remaining parameters of the model. In addition, the regularity conditions for convergence in distribution of supLR statistic are beyond the scope of this paper, and we instead propose to apply a bootstrap based test. The details of the bootstrap are given in Appendix C and is illustrated in the next Section 5.

# 5 An Empirical Illustration

In this section we provide an empirical illustration of the $\lambda$-GARCH model, and test nullity of rows, as well as reduced rank of the $\lambda$-GARCH loading matrices. We use daily return data for three financial equities[2] from the S&P 500 Index with sample period January 3rd 2006 to January 2nd 2018. The log-returns are shown in Figure 5. Initial inspection of the data reveals that the unconditional densities are heavy-tailed and the data is characterized by ARCH effects. The log-returns appear to exhibit volatility clustering during the same periods, and hence may share a common factor (or eigenvalue) driving their volatility.

[Figures 4 and 5 here]

Table 1 contains the parameter estimates of the trivariate $\lambda$-GARCH model. A few of the parameter estimates in the unrestricted model are on the boundary of the parameter space, indicating that the distribution of the associated estimators may not be Gaussian, but rather follow a half-normal type distribution. The residuals of the unrestricted model and their densities are given in Figure 6. The densities of the residuals are slightly heavy-tailed, and unreported misspecification test indicate no ARCH-effects and no residual autocorrelation.

The estimated conditional eigenvalues process $\lambda_t$ is highly persistent as $\rho(\hat{A}_T + \hat{B}_T) \approx 0.997$. Thus $\lambda_t$, and hence $X_t$, exhibit near-IGARCH-type behavior, similar to standard univariate and multivariate GARCH models. The high degree of persistence is likely to be caused by the almost explosive spike in volatility during the financial crisis of 2008-2009, as can be seen in Figure 7. We note that $\lambda_{3t}$ on average explains 85% of the variation in the dataset, and inspecting the corresponding eigenvector reveals that this can be interpreted as a "market factor", with each asset have a (normalized) weight of roughly 30% in the rotated return. The two remaining eigenvalues individually explain $6-8\%$ of the variation on average, and their corresponding rotated returns are long-short portfolios of the data. Importantly, while the two smaller eigenvalues are of lesser importance compared to the "market eigenvalue", they are not constant, and all rotated returns have inherited ARCH effects, as can be seen from Figure 8.

[2]Bank of America corp. (BAC), JPMorgan Chase & co. (JPM), and Wells Fargo (WFC).

[Table 1 here]

Consider next hypothesis of a zero row in $A$ and $B$, that is $\mathsf{H}_2^*$ in (19). As mentioned in the previous section, the hypothesis is tested using a supLR test and critical values are obtained by simulations, see Appendix B.1 for details.[3] Based on the supLR statistic of 1963.16 and the associated critical value of 738.35 (both given in Table 1), we reject the hypothesis of a zero row. Intuitively, this seems sensible: under the hypothesis one of the eigenvalues is constant, and the associated rotated return homoskedastic. As already noted, this is not the case as all three rotated returns in the unrestricted model clearly have ARCH-effects (see Figure 8).

[Figure 6 here]

The second hypothesis that we test the less restrictive assumption of reduced rank $r = 2$ of the matrices $A$ and $B$, that is $\mathsf{H}_2$ in (18). Under $\mathsf{H}_2$ all eigenvalues are allowed to remain time-varying, while $p - r = 1$ linear combination of these is constant. To ensure identification of $\gamma, \alpha$ and $\beta$ under $\mathsf{H}_2$, the upper $(2 \times 2)$ block of $\gamma$ is set to $I_2$, while the last row of $\gamma$ is freely varying. We obtain critical values by the bootstrap algorithm in Appendix C, see also Cavaliere *et al.* (2019). The critical value is obtained from $B = 399$ bootstrap replications. The LR statistic is 3.0 and the associated bootstrapped 95% critical value is 18.56 such that $\mathsf{H}_2$ is not rejected.[4] From the estimated parameters for the reduced rank model (reported in tables 2 and 3), the estimated parameters, eigenvalues, and conditional covariances for unrestricted model and reduced rank model are non-distinguishable, and based on the AIC and BIC, the reduced rank model is in fact preferable to the unrestricted model.

[Table 2 here]
[Table 3 here]

From this empirical illustration we make the following notes: First, the $\lambda$–GARCH model performs well for the series studied, and the estimated time-varying eigenvalues and eigenvectors are easy to interpret, reflecting market conditions at a given time. Second, despite

[3] For each entry of the non-identified parameter vector $\delta = (B_{13}, B_{23}, B_{33})'$ we use $k = 20$ equi-distant points between 0 and 0.99 (both points included), leading to a grid of $20^3 = 8000$ points for $\delta$. Steps 1-3 of the algorithm for the asymptotic distribution of the test only draws from grid points in which: *i)* the Hessian matrix is invertible, as determined by the reciprocal condition number, and *ii)* the log-likelihood value is close to the maximum likelihood value. That is, for $i = 1, \ldots, \dim\Delta$, we only use a given grid point if rcond $(\hat{J}_{\delta_i}) > 10^{-12}$ and $L_T(\theta, \delta_i) + 5 \geq \sup_{\delta_i \in \Delta} L_T(\theta, \delta_i)$ both hold. We use $M = 10000$ Monte Carlo draws to determine the critical value.

[4] We also test the hypothesis that the rank of $A$ and $B$ is $q = 1$. This test is strongly rejected, with a LR test of 140.89 and a bootstrapped critical value of 30.57.

the fact that the three equities are all in the same sector and have a shared source of the majority of variation in a "market" eigenvalue, we cannot restrict one of the lesser important eigenvalues to be constant without a significant loss of explanatory power. Third, we note the usefulness of the reduced rank structure in conditional covariance matrices. The finding that the parameter matrices $A$ and $B$ are reduced rank is novel, and it may have interesting implications for the applications of models for the conditional covariance matrices, as it is a coherent way of imposing a structure and reduce the dimensionality of the model without losing explanatory power.

# References


Andrews, D.W. (2001): "Testing when a parameter is on the boundary of the maintained hypothesis", *Econometrica*, 69, 683–734.

Avarucci, M., E. Beutner, and P. Zaffaroni (2013): "On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models", *Econometric Theory*, 29, 545–566.

Bardet, J.M. and O. Wintenberger (2009): "Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes", *The Annals of Statistics*, 37, 2730–2759.

Bollerslev, T. (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model.", *Review of Economics and Statistics*, 72, 498–505.

Boswijk, H.P. and R. van der Weide (2011): "Method of moments estimation of GO-GARCH models", *Journal of Econometrics*, 163, 118–126.

Brown, B. M. (1971): "Martingale Central Limit Theorems", *The Annals of Mathematical Statistics*, 42, 59–66.

Cavaliere, G., H.B. Nielsen, R.S. Pedersen and A. Rahbek (2019): "Bootstrap inference on the boundary of the parameter space with application to conditional volatility models", *Journal of Econometrics*, forthcoming.

Cavaliere, G., H.B. Nielsen and A. Rahbek (2017): "On the Consistency of Bootstrap Testing for a Parameter on the Boundary of the Parameter Space", *Journal of Time Series Analysis*, 38, 513-534.

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2012): “Bootstrap determination of the co-integration rank in vector autoregressive models", *Econometrica*, 80, 1721-1740.

Comte, F. and O. Lieberman (2003): “Asymptotic theory for multivariate GARCH processes”, *Journal of Multivariate Analysis*, 84, 61–84.

Creal, D., S.J. Koopman and A. Lucas (2011): “A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations”, *Journal of Business & Economic Statistics*, 29, 552–563.

—— (2013): “Generalized autoregressive score models with applications”, *Journal of Applied Econometrics*, 28, 777–795.

Dovonon, P. and E. Renault (2013): “Testing for common conditionally heteroskedastic factors”, *Econometrica*, 81, 2561–2586.

Engle, R.F. (2002): “Dynamic conditional correlation: A simple class of multivariate GARCH models”, *Journal of Business & Economics Statistics*, 20, 339–350.

Fan, J., M. Wang and Q. Yao (2008): “Modelling multivariate volatilities via conditionally uncorrelated components”, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70, 679–702.

Francq C., L. Hovath and J.M. Zakoïan (2014): “Variance targeting estimation of multivariate GARCH models”, *Journal of Financial Econometrics*, 14, 353–382.

Francq C. and J.M. Zakoïan (2009): “Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons", *Journal of the American Statistical Association*, 104, 313–324.

—— (2012): “QML estimation of a class of multivariate asymmetric GARCH models”, *Econometric Theory*, 28, 179–206.

—— (2019): “GARCH Models: Structure, Statistical Inference and Financial Applications”, John Wiley and sons.

Hafner, C.M. and A. Preminger (2009a): “On asymptotic theory for multivariate GARCH models”, *Journal of Multivariate Analysis*, 100, 2044–2054.

Hafner, C.M. and A. Preminger (2009b): “Asymptotic theory for a factor GARCH model”, *Econometric Theory*, 25, 336–363.

HARBO, I., S. JOHANSEN, B. NIELSEN AND A. RAHBEK (1998): "Asymptotic inference on cointegration rank in partial systems", *Journal of Business & Economic Statistics*, 16, 388-99.

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

HARVEY, A.C. AND T. CHAKRAVARTY (2008): "Beta-t-(e) GARCH", working paper.

HIDALGO, J. AND P. ZAFFARONI (2007): "A goodness-of-fit test for ARCH($\infty$) models", *Journal of Econometrics*, 141, 835-875.

JEANTHEAU, T. (1998): "Strong consistency of estimators for multivariate ARCH models", *Econometric Theory*, 14(1), 70–86.

JENSEN, S.T AND A. RAHBEK (2004): "Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case", *Econometrica*, 72, 641–646.

JEONG, M. (2017): "Residual-based GARCH bootstrap and second order asymptotic refinement", *Econometric Theory*, 33(3), 779-790.

LANNE, M. AND P. SAIKKONEN (2007): "A multivariate generalized orthogonal factor GARCH model", *Journal of Business & Economic Statistics*, 25, 61–75.

NOURELDIN, D, N. SHEPHARD AND K. SHEPPARD (2014): "Multivariate rotated ARCH models", *Journal of Econometrics*, 179, 16–30.

PEDERSEN, R.S. (2017): "Inference and testing on the boundary in extended constant conditional correlation GARCH models", *Journal of Econometrics*, 196, 25–36.

PEDERSEN, R.S. AND A. RAHBEK (2014): "Multivariate variance targeting in the BEKK-GARCH model", *The Econometrics Journal*, 17, 24–55.

PEDERSEN, R.S. AND A. RAHBEK (2019): "Testing GARCH-X type models", *Econometric Theory*, 35, 1012–1047.

VAN DER WEIDE, R. (2002): "GO-GARCH: A multivariate generalized orthogonal GARCH model", *Journal of Applied Econometrics*, 17, 549–564.

# APPENDIX

# A Mathematical Proofs

## A.1 Notation and definitions

Throughout, we let $\varrho \in (0,1)$ denote a generic constant, and $K$ is a generic positive constant or positive $\mathcal{F}_{-1}$-measurable random variable. Moreover, let $Y_t(\theta) := V(\theta)'X_t$ denote the orthogonalized returns. In light of Assumption 3.2, we consider the ergodic version of the log-likelihood contributions. That is, for any $t \in \mathbb{Z}$ and $\theta \in \Theta$,

$$l_t^\star(\theta) = \log\det(\Omega_t^\star(\theta)) + X_t'\Omega_t^{\star-1}(\theta)X_t, \tag{A.1}$$

$$\Omega_t^\star(\theta) = V(\theta)\Lambda_t^\star(\theta)V(\theta)', \quad \Lambda_t^\star(\theta) = \mathrm{diag}(\lambda_t^\star(\theta)), \tag{A.2}$$

$$\lambda_t^\star(\theta) = W + A(V(\theta)'X_{t-1})^{\odot 2} + B\lambda_{t-1}^\star(\theta). \tag{A.3}$$

For derivatives,

$$\dot{B}_i = \frac{\partial B(\theta)}{\partial \theta_i}, \quad \ddot{B}_{i,j} = \frac{\partial^2 B_t(\theta)}{\partial \theta_i \partial \theta_j}, \quad \dddot{B}_{i,j,k} = \frac{\partial^3 B(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad i,j,k \in \{1,\ldots,d_\theta\},$$

denote the partial derivatives of some scalar, vector, or matrix $B(\theta)$ as a function of $\theta \in \Theta$ with $d_\theta$ the dimension of $\theta$.

Furthermore we let $\Omega_t^\star = \Omega_t^\star(\theta_0)$, that is $\Omega_t^\star$ evaluated at the true parameter values, $\theta_0$. The same holds for other quantities which depending on $\theta_0$, e.g. $Y_t = Y_t(\theta_0)$, $\Lambda_t^\star = \Lambda_t^\star(\theta_0)$, and $\lambda_t^\star = \lambda_t^\star(\theta_0)$

## A.2 Proof of Theorem 3.2

It suffices to verify conditions A1-A5 of Francq and Zakoïan (2019, Theorem 10.7). With $\Omega_t^\star(\theta)$ defined in (A.2), we immediately notice that Assumption 3.2 implies that $E[\|\Omega_t^\star(\theta_0)\|]^s < \infty$ for some $s > 0$ (condition A3). Moreover, recall that $\rho(B) < 1$ on $\Theta$, and define the function $\lambda : (\mathbb{R}^p)^\infty \times \Theta \to \mathbb{R}^p$, with $(x_0, x_{-1}, \ldots)$ a sequence of vectors in $\mathbb{R}^p$ and $\theta \in \Theta$, given by

$$\lambda(x_0, x_{-1}, \ldots; \theta) = \sum_{i=0}^{\infty} B^i \left[W + A(V(\theta)'x_{-i})^{\odot 2}\right].$$

We note that for any sequence $(x_0, x_{-1}, \ldots)$, $\lambda(x_0, x_{-1}, \ldots; \cdot)$ is continuous on $\Theta$ (condition A5). It remains to show the following three points.

**(i)** With $\Omega_t(\theta)$ and $\Omega_t^\star(\theta)$ defined in (12) and (A.2), respectively, $\sup_{\theta\in\Theta}\|\Omega_t^{-1}(\theta)\| \leq K$ and $\sup_{\theta\in\Theta}\|\Omega_t^{\star-1}(\theta)\| \leq K$ almost surely.

**(ii)** $\sup_{\theta\in\Theta}\|\Omega_t(\theta) - \Omega_t^\star(\theta)\| \leq K\varrho^t$ almost surely.

**(iii)** For $\theta \in \Theta$, $\Omega_t^\star(\theta) = \Omega_t^\star(\theta_0)$ almost surely $\Rightarrow \theta = \theta_0$.

*Proof of (i)*: Note that $\sup_{\theta\in\Theta}\|\Omega_t^{-1}(\theta)\| \leq \sup_{\theta\in\Theta}\|V\|^2\|\Lambda_t^{-1}(\theta)\| \leq K\sqrt{p\omega_L^{-2}} \leq K$. Likewise, $\sup_{\theta\in\Theta}\|\Omega_t^{\star-1}(\theta)\| \leq K$.
*Proof of (ii)*: With $\lambda_t(\theta)$ and $\lambda_t^\star(\theta)$ defined in (13) and (A.3), respectively, using that $\sup_{\theta\in\Theta}\rho(B) < 1$, we have that

$$\sup_{\theta\in\Theta}\|\Omega_t(\theta) - \Omega_t^\star(\theta)\| = \sup_{\theta\in\Theta}\|\lambda_t(\theta) - \lambda_t^\star(\theta)\| = \sup_{\theta\in\Theta}\|B^t(\bar{\lambda}_0 - \lambda_0^\star(\theta))\| \leq \varrho^t K.$$

*Proof of (iii)*: For $\theta \in \Theta$, suppose that $\Omega_t^\star(\theta) = \Omega_t^\star(\theta_0)$ a.s., such that with $\tilde{V} := V(\phi_0)'V(\phi)$ it holds that

$$\tilde{V}\Lambda_t^\star(\theta) = \Lambda_t^\star(\theta_0)\tilde{V} \text{ a.s.}$$

Suppressing dependency on $t$ this is,

$$\begin{bmatrix} \tilde{V}_{11}\Lambda_{11}^\star(\theta) & \tilde{V}_{12}\Lambda_{22}^\star(\theta) & \dots & \tilde{V}_{1p}\Lambda_{pp}^\star(\theta) \\ \tilde{V}_{21}\Lambda_{11}^\star(\theta) & \tilde{V}_{22}\Lambda_{22}^\star(\theta) & & \\ \vdots & & \ddots & \\ \tilde{V}_{p1}\Lambda_{11}^\star(\theta) & \tilde{V}_{p2}\Lambda_{22}^\star(\theta) & & \tilde{V}_{pp}\Lambda_{pp}^\star(\theta) \end{bmatrix} = \begin{bmatrix} \tilde{V}_{11}\Lambda_{11}^\star(\theta_0) & \tilde{V}_{12}\Lambda_{11}^\star(\theta_0) & \dots & \tilde{V}_{1p}\Lambda_{11}^\star(\theta_0) \\ \tilde{V}_{21}\Lambda_{22}^\star(\theta_0) & \tilde{V}_{22}\Lambda_{22}^\star(\theta_0) & & \\ \vdots & & \ddots & \\ \tilde{V}_{p1}\Lambda_{pp}^\star(\theta_0) & \tilde{V}_{p2}\Lambda_{pp}^\star(\theta_0) & & \tilde{V}_{pp}\Lambda_{pp}^\star(\theta_0) \end{bmatrix} \text{ a.s.}$$

By Assumption 3.5, there exists a $j$ such that $\tilde{V}_{jj} \neq 0$. Hence, for this $j$,

$$\Lambda_{jj}^\star(\theta) = \Lambda_{jj}^\star(\theta_0) \text{ a.s.}$$

if and only if,

$$\omega_j + A_j(V'X_{t-1})^{\odot 2} + B_j\lambda_{t-1}^\star(\theta) = \omega_{0,j} + A_{0,j}(V_0'X_{t-1})^{\odot 2} + B_{0,j}\lambda_{t-1}^\star(\theta_0) \text{ a.s.}$$

if and only if,

$$\omega_1 - \omega_{0,j} + B_j\lambda_{t-1}^\star(\theta) - B_{0,j}\lambda_{t-1}^\star(\theta_0) = A_{0,j}(V_0'X_{t-1})^{\odot 2} - A_j(V'X_{t-1})^{\odot 2} \text{ a.s.} \tag{A.4}$$

Noting that the left-hand-side of (A.4) is $\mathcal{F}_{t-2}^\eta$-measurable, we have that $A_{0,j}(V_0'X_{t-j})^{\odot 2} - A_j(V'X_{t-j})^{\odot 2}|\,\mathcal{F}_{t-2}^\eta$ is degenerate, which by Assumption 3.5 implies that $V = V_0$, and moreover, by Assumption 3.5, $\phi = \phi_0$. Since $V = V_0$, we have, by Assumptions 3.3 and 3.4

and arguments given by Francq and Zakoïan (2019, p.308), that $(W', \text{vec}(A)', \text{vec}(B)') = (W_0', \text{vec}(A_0)', \text{vec}(B_0)')$. We conclude that point 3 holds.

## A.3 Proof of Theorem 3.3

Using that $\theta_0 \in \text{int}\Theta$, with $\Theta$ compact, and $l_t^\star(\theta)$ defined in (A.1) is three times continuously differentiable (almost surely), it suffices to verify the following conditions (see e.g. Francq and Zakoïan, 2012):

**(Asymptotic normality of the score)** With $l_t^\star(\theta)$ defined in (A.1),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial l_t^\star(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Sigma), \tag{A.5}$$

with

$$\Sigma := E\left[\frac{\partial l_t^\star(\theta_0)}{\partial \theta} \frac{\partial l_t^\star(\theta_0)}{\partial \theta'}\right] \quad \text{nonnegative definite.} \tag{A.6}$$

**(Hessian)** With $l_t^\star(\theta)$ defined in (A.1),

$$\frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 l_t^\star(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} E\left[\frac{\partial^2 l_t^\star(\theta_0)}{\partial \theta \partial \theta'}\right] =: J, \tag{A.7}$$

with $J$ invertible.

**(Expectation of Third Order Derivative)** With $l_t^\star(\theta)$ defined in (A.1) for some neighborhood $N(\theta_0) \subset \Theta$ around $\theta_0$,

$$E\left[\max_{i,j,k=1,\ldots,d_\theta} \sup_{\theta \in N(\theta_0)} \left|\frac{\partial^3 l_t^\star(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_j}\right|\right] < \infty.$$

**(Initial Values)** With $l_t(\theta)$ defined in (11) and $l_t^\star(\theta)$ defined in (A.1), for some neighborhood $N(\theta_0)$ around $\theta_0$,

$$\left\|\sum_{t=1}^{T} \left(\frac{\partial l_t(\theta_0)}{\partial \theta} - \frac{\partial l_t^\star(\theta_0)}{\partial \theta}\right)\right\| = o_p(T^{1/2}),$$

and

$$\sup_{\theta \in N(\theta_0)} \left\|\sum_{t=1}^{T} \left(\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 l_t^\star(\theta)}{\partial \theta \partial \theta'}\right)\right\| = o_p(T).$$

*Proof of Asymptotic Normality*: From Lemma A.1 we have that $E[\partial l_t^\star(\theta_0)/\partial \theta | \mathcal{F}_{t-1}] = 0$ and $E[\|\partial l_t^\star(\theta_0)/\partial \theta\|] < \infty$. By a CLT for stationary and ergodic martingale differences (e.g.

Brown, 1971), we conclude that (A.5) holds. The matrix $\Sigma$ in (A.6) is nonnegative definite by construction.

*Proof of Hessian*: From Lemma A.4, we have that $E[\|\partial^2 l_t^\star(\theta_0)/\partial\theta\partial\theta'\|] < \infty$. By the Ergodic Theorem, we conclude that (A.7) holds. Moreover, Lemma A.4 states that the matrix $J$ is invertible.

*Proof of Third Derivative*: This property holds by Lemma A.5.

*Proof of Initial Value*: This holds by arguments similar to the ones given in Francq and Zakoïan (2019, pp.308).

## A.4 Proof of Lemma 1

We start out by proving that (15) holds. Recall that $\phi$ contains $p(p-1)/2$ parameters. Due to the structure of $V$ and since $\phi_0 \in \text{int}\Theta_\phi = (0, \pi/2)^{p(p-1)/2}$, it suffices to consider the $p(p-1)/2$ entries of $V$ and $V_0$ below their diagonals. Consider initially the case for $p = 3$, where $(\phi_1, \phi_2, \phi_3)' \equiv (\phi_{1,2}, \phi_{1,3}, \phi_{2,3})'$ and

$$V = \begin{pmatrix} \cos\phi_1\cos\phi_2 & \cos\phi_3\sin\phi_1 - \cos\phi_1\sin\phi_2\sin\phi_3 & \sin\phi_1\sin\phi_3 + \cos\phi_1\cos\phi_3\sin\phi_2 \\ -\cos\phi_2\sin\phi_1 & \cos\phi_1\cos\phi_3 + \sin\phi_1\sin\phi_2\sin\phi_3 & \cos\phi_1\sin\phi_3 - \cos\phi_3\sin\phi_1\sin\phi_2 \\ -\sin\phi_2 & -\cos\phi_2\sin\phi_3 & \cos\phi_2\cos\phi_3 \end{pmatrix}$$

Starting from the last element of the first column of $V$,

$$V_{3,1}(\phi_0) = V_{3,1}(\phi) \implies -\sin\phi_{0,2} = -\sin\phi_2 \implies \phi_{0,2} = \phi_2.$$

Next, notice that

$$V_{2,1}(\phi_0) = V_{2,1}(\phi) \implies -\sin\phi_{0,1}\cos(\phi_{0,2}) = -\sin\phi_1\cos(\phi_2) \implies \phi_{0,1} = \phi_1$$

$$V_{3,2}(\phi_0) = V_{3,2}(\phi) \implies -\sin\phi_{0,3}\cos(\phi_{0,2}) = -\sin\phi_3\cos(\phi_2) \implies \phi_{0,3} = \phi_3,$$

and hence $V = V_0 \Rightarrow \phi = \phi_0$. This argument can be extended to arbitrary $p \geq 2$. In particular, with $(\phi_1, \phi_2, \ldots, \phi_{p-1}, \phi_p, \ldots, \phi_{p(p-1)/2})' \equiv (\phi_{1,2}, \phi_{1,3}, \ldots, \phi_{1,p}, \phi_{2,3}, \ldots, \phi_{(p-1),p})'$ we have that $V_{p,1}(\theta_0) = V_{p,1}(\theta) \implies -\sin\phi_{0,p-1} = -\sin\phi_{p-1} \implies \phi_{0,p-1} = \phi_{p-1}$ for $\phi \in \text{int}\Theta_\phi$.

We also note that for any $p \geq 2$, the first column and last row of $V$ are given respectively by

$$V_{j,1} = -\sin\phi_{j-1}\prod_{i=1}^{p-j}\cos\phi_{p-i}, \quad j = 2, \ldots, p-1,$$

$$V_{p,j} = -\sin\phi_{p(p-1)/2-(p-j)(p-j-1)/2} \prod_{i=p-j+1}^{p-1} \cos\phi_{p(p-1)/2-i(i-1)/2}, \quad j = 2, \ldots, p-1.$$

This gives the (recursive) identification of $2p-3$ parameters ($\phi_{p-1}, \phi_{j-1}, \phi_{p(p-1)/2-(p-j)(p-j-1)/2}$ : $j = 2, \ldots, p-1$). The remaining parameters can be shown to be identified by considering next $V_{p-1,2}$ and moving row-wise up and column-wise right through elements $V_{j,2}$ for $j = p-2, \ldots, 3$ and $V_{p-1,j}$ for $j = 3, \ldots, p-2$, which yields identification of another $2(p-3)-1$ parameters. Similar arguments may be repeated for $V_{p-2,3}$ and so forth, until all elements below the diagonal have been covered. Hence, we have that $V = V_0 \Rightarrow \phi = \phi_0$.

Next, we show that (14) holds. Suppose that for some $j \in \{1, \ldots, p\}$,

$$A_j(V'X_t)^{\odot 2} - A_{j,0}(V_0'X_t)^{\odot 2}|\mathcal{F}_{t-1}^{\eta} \text{ is degenerate.} \tag{A.8}$$

Note that $V_0'X_t = V_0'V_0\Lambda_t^{1/2}(\theta_0)\eta_t = \Lambda_t^{1/2}(\theta_0)\eta_t =: M_0\eta_t$ and $V'X_t = V'V_0\Lambda_t^{1/2}(\theta_0)\eta_t =: M\eta_t$ with $M_0$ and $M$ $\mathcal{F}_{t-1}^{\eta}$-measurable matrices. If $\eta_t \sim N(0, I_p)$, then conditional on $\mathcal{F}_{t-1}^{\eta}$,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} := \begin{bmatrix} M_0\eta_t \\ M\eta_t \end{bmatrix} \sim N(0, \Sigma)$$

with

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} M_0M_0' & M_0M' \\ MM_0' & MM' \end{bmatrix}.$$

Let $\kappa = \Sigma_{12}\Sigma_{22}^{-1}$. Then $Y_{1|2} := Y_1 - \kappa Y_2$ and $Y_2$ are conditionally independent, since (conditional on $\mathcal{F}_{t-1}^{\eta}$)

$$\begin{bmatrix} Y_{1|2} \\ Y_2 \end{bmatrix} \sim N(0, \tilde{\Sigma})$$

with

$$\begin{aligned} \tilde{\Sigma} &= \begin{bmatrix} I_p & -\kappa \\ 0 & I_p \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_p & -\kappa \\ 0 & I_p \end{bmatrix}' \\ &= \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}. \end{aligned}$$

Note that $Y_{1|2}|\mathcal{F}_{t-1}^{\eta}$ (and hence $(Y_{1|2})^{\odot 2}|\mathcal{F}_{t-1}^{\eta}$) is non-degenerate if and only if $M \neq M_0$ if and only if $V \neq V_0$. Hence (A.8) is equivalent to

$$A_{0,j}(Y_1 - Y_{1|2} + Y_{1|2})^{\odot 2} - A_j(Y_2)^{\odot 2}|\mathcal{F}_{t-1}^{\eta} \text{ is degenerate}$$

if and only if

$$A_{0,j}(\kappa Y_2 + Y_{1|2})^{\odot 2} - A_j(Y_2)^{\odot 2} \,|\mathcal{F}^{\eta}_{t-1} \text{ is degenerate}$$

if and only if

$$A_{0,j}(Y_{1|2})^{\odot 2} + A_{0,j}(\kappa Y_2)^{\odot 2} + 2A_{0,j}(\kappa Y_2 \odot Y_{1|2}) - A_j(Y_2)^{\odot 2}|\mathcal{F}^{\eta}_{t-1} \text{ is degenerate}$$

Since $Y_{1|2}$ and $Y_2$ are $\mathcal{F}^{\eta}_{t-1}$-conditionally independent and all entries of $A_{0,j}$ are strictly positive, as $\theta_0$ is an interior point of $\Theta$, we can only have that $A_{0,j}(Y_{1|2})^{\odot 2} + A_{0,j}(\kappa Y_2)^{\odot 2} + 2A_{0,j}(\kappa Y_2 \odot Y_{1|2}) - A_j(Y_2)^{\odot 2}|\mathcal{F}_{t-1}$ is degenerate if $Y_{1|2}|\mathcal{F}^{\eta}_{t-1}$ is degenerate, which contradicts $V \neq V_0$.

## A.5 Auxiliary Lemmas

**Lemma A.1.** *With $l^{\star}_t(\theta)$ defined in (A.1), under Assumptions 3.1-3.8, it holds that*

$$E\left[\left.\frac{\partial l^{\star}_t(\theta_0)}{\partial \theta}\right| \mathcal{F}_{t-1}\right] = 0 \text{ almost surely,} \tag{A.9}$$

$$E\left[\left\|\frac{\partial l^{\star}_t(\theta_0)}{\partial \theta}\right\|^2\right] < \infty, \quad \text{and} \tag{A.10}$$

$$\Sigma = E\left[\frac{\partial l^{\star}_t(\theta_0)}{\partial \theta}\frac{\partial l^{\star}_t(\theta_0)}{\partial \theta'}\right]. \tag{A.11}$$

*Proof of Lemma A.1:* With $Y_t(\theta) = V(\phi)'X_t$, for $i = 1, \ldots, d_\theta$, we have from Lemma A.2 that

$$\frac{\partial l^{\star}_t(\theta)}{\partial \theta_i} = \operatorname{tr}\left\{\Lambda^{\star -1}_t(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\left[I_p - \Lambda^{\star -1}_t(\theta)Y_t(\theta)Y'_t(\theta)\right]\right\} + 2\dot{Y}'_{t,i}(\theta)\Lambda^{\star -1}_t(\theta)Y_t(\theta),$$

with $\dot{\lambda}_{t,i}(\theta) := \partial\lambda^{\star}_t(\theta)/\partial\theta_i$ and $\dot{Y}_{t,i}(\theta) := \partial Y_t(\theta)/\partial\theta_i$. Evaluating at $\theta_0$, we have

$$\begin{aligned}\frac{\partial l^{\star}_t(\theta_0)}{\partial \theta_i} &= \operatorname{tr}\left\{\Lambda^{\star -1}_t\dot{\Lambda}^{\star}_{t,i}\left[I_p - \eta_t\eta'_t\right]\right\} + 2\dot{Y}'_{t,i}\Lambda^{\star -1}_t Y_t \\ &=: M_{1,t,i} + M_{2,t,i}.\end{aligned} \tag{A.12}$$

Suppose initially that $M_{1,t,i}$ and $M_{2,t,i}$ are integrable such that their conditional expectations exist - this will indeed be verified below. We have immediately that

$$E[M_{1,t,i}|\mathcal{F}_{t-1}] = 0 \text{ almost surely,} \tag{A.13}$$

since $\Lambda_t^{\star-1}(\theta_0)\dot{\Lambda}_{t,i}^{\star}$ is $\mathcal{F}_{t-1}$ measurable and $E[\eta_t\eta_t'|\mathcal{F}_{t-1}] = E[\eta_t\eta_t'] = I_p$. Turning to $M_{2,t}$ note that

$$V(\theta)V'(\theta) = I_p,$$

which implies that

$$\frac{\partial V(\theta)}{\partial \theta_i}V'(\theta) + V(\theta)\frac{\partial V'(\theta)}{\partial \theta_i} = 0.$$

With $S_i(\theta) := (\partial V(\theta)/\partial\theta_i)V'(\theta)$, we have that

$$\frac{\partial V(\theta)}{\partial \theta_i} = S_i(\theta)V(\theta),$$

where $S_i'(\theta) = -S_i(\theta)$, and, hence, $S_i(\theta)$ is a skew-symmetric matrix satisfying

$$\mathrm{tr}(S_i(\theta)) = 0. \tag{A.14}$$

For $\theta = \theta_0$ we then have

$$M_{2,t,i} = 2\dot{Y}_{t,i}'\Lambda_t^{\star-1}Y_t = 2X_t'S_iV\Lambda_t^{\star-1}V'X_t = 2\mathrm{tr}\{S_i\Omega_t^{\star-1}X_tX_t'\},$$

using $E[X_tX_t'|\mathcal{F}_{t-1}] = \Omega_t^{\star}$ and (A.14),

$$E[M_{t,2,i}|\mathcal{F}_{t-1}] = 2\mathrm{tr}\{S_i\} = 0 \text{ almost surely.} \tag{A.15}$$

Combining (A.12), (A.13), and (A.15), we conclude that (A.9) holds. Turning to (A.10), we note that it suffices to show that $E[(\partial l_t^{\star}(\theta_0)/\partial\theta_i)^2] < \infty$ for all $i$, which in light of (A.12) and the Cauchy-Schwarz inequality holds if $E[M_{1,t,i}^2] < \infty$ and $E[M_{2,t,i}^2] < \infty$. We have that, almost surely,

$$E\left[M_{1,t,i}^2|\mathcal{F}_{t-1}\right] = E\left[\mathrm{tr}^2\left\{\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\left[I_p - \eta_t\eta_t'\right]\right\}|\mathcal{F}_{t-1}\right] = \sum_{q=1}^{p}\left(E[\eta_{q,t}^4]-1\right)[\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}]_{qq}^2,$$

where we note that $E[\eta_{q,t}^4] < \infty$, $q = 1, \ldots, p$, by Assumption 3.6. Hence,

$$E[M_{1,t,i}^2] = \sum_{q=1}^{p}\left(E[\eta_{q,t}^4]-1\right)E\left[[\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}]_{qq}^2\right],$$

and by Lemma A.6, we have that $E[M_{1,t,i}^2] < \infty$, $i = 1, \ldots, d_\theta$. Turning to the variance of

$M_{2,t,i}$, note that with $\tilde{S}_i = VS_i'V'$,

$$
\begin{aligned}
M_{2,t,i}^2 &= 4X_t'S_iV\Lambda_t^{\star-1}V'X_tX_t'S_iV\Lambda_t^{\star-1}V'X_t \\
&= 4\mathrm{tr}\left(S_i\Omega_t^{\star-1}X_tX_t'S_i\Omega_t^{\star-1}X_tX_t'\right) = 4\mathrm{tr}(\tilde{S}_i'\Lambda_t^{\star-1}Y_tY_t'\tilde{S}_i'\Lambda_t^{\star-1}Y_tY_t') \\
&\leq K\|\Lambda_t^{\star-1}Y_tY_t'\|^2 = K\left(Y_t'Y_tY_t'\Lambda_t^{\star-2}Y_t\right) = K\left(\sum_{i=1}^p y_{it}^2\right)\left(\sum_{i=1}^p \frac{y_{it}^2}{\lambda_{it}^{\star 2}}\right). \qquad \text{(A.16)}
\end{aligned}
$$

We note that (A.16) consists of terms of the form

$$
\frac{y_{it}^2 y_{jt}^2}{\lambda_{it}^{\star 2}} = \eta_{it}^2\eta_{jt}^2\frac{\lambda_{jt}^\star}{\lambda_{it}^\star}.
$$

Using Assumption 3.7 and that for $\theta_0 \in \mathrm{int}\Theta$,

$$
\frac{\lambda_{k,t}^\star}{\lambda_{l,t}^\star} = \frac{\omega_{0,k} + \sum_{i=1}^p \alpha_{0,ki}y_{i,t-1}^2 + \sum_{i=1}^p \beta_{0,ki}\lambda_{i,t-1}^\star}{\omega_{0,l} + \sum_{i=1}^p \alpha_{0,li}y_{i,t-1}^2 + \sum_{i=1}^p \beta_{0,li}\lambda_{i,t-1}^\star} \leq \frac{\omega_{0,k}}{\omega_{0,l}} + \sum_{i=1}^p \frac{\alpha_{0,ki}}{\alpha_{0,li}} + \sum_{i=1}^p \frac{\beta_{0,ki}}{\beta_{0,li}} \leq K, \quad \text{(A.17)}
$$

we have that $\eta_{it}^2\eta_{jt}^2\lambda_{jt}^\star/\lambda_{it}^\star$ is integrable for any $i, j$, and we conclude that $E[M_{2,t,i}^2] < \infty$ for any $i$.

**Lemma A.2.** *With $l_t^\star(\theta)$ defined in (A.1),*

$$
\frac{\partial l_t^\star(\theta)}{\partial \theta_i} = tr\left\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\left[I_p - \Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right]\right\} + 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta), \quad i = 1, \ldots, d_\theta,
$$

*with*

$$
\dot{\lambda}_{t,i}^\star := \frac{\partial \lambda_t^\star(\theta)}{\partial \theta_i} \quad \text{and} \quad \dot{Y}_{t,i} := \frac{\partial Y_t(\theta)}{\partial \theta_i}.
$$

*Proof of Lemma A.2:* We have that,

$$
\frac{\partial l_t^\star(\theta)}{\partial \theta_i} = \frac{\partial \log|\Lambda_t^\star(\theta)|}{\partial \theta_i} + \frac{\partial Y_t'(\theta)\Lambda_t^\star(\theta)^{-1}Y_t(\theta)}{\partial \theta_i}.
$$

Consider now,

$$
\frac{\partial \log|\Lambda_t^\star(\theta)|}{\partial \theta_i} = \mathrm{tr}\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\}.
$$

Next, consider $Y_t'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta) = \mathrm{tr}\{Y_t(\theta)Y_t'(\theta)\Lambda_t^{\star-1}(\theta)\}$. Since $Y_t(\theta)Y_t(\theta)'$ is symmetric and $\Lambda_t^{\star-1}(\theta)$ is diagonal we find

$$
\frac{\partial \mathrm{tr}\{Y_t(\theta)Y_t'(\theta)\Lambda_t^{\star-1}(\theta)\}}{\partial \theta_i} = 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta) - \mathrm{tr}\left\{Y_t(\theta)Y_t'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\right\}.
$$

Hence, the score with respect to $\theta_i$ is

$$\frac{\partial l_t^\star(\theta)}{\partial\theta_i} = \text{tr}\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\} - \text{tr}\left\{Y_t(\theta)Y_t'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\right\} + 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)$$

$$= \text{tr}\left\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\left[I_p - \Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right]\right\} + 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta).$$

**Lemma A.3.** *With $l_t^\star(\theta)$ defined in (A.1), for $i, j = 1, \ldots, d_\theta$,*

$$\begin{aligned}
\frac{\partial^2 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j} &= -tr\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}(\theta)\right) + tr\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\right) \\
&+ tr\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&- tr\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + tr\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&- 2tr\left(\tilde{S}_j'(\theta)(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + 2tr\left((\dot{S}_{i,j}(\theta) + S_i(\theta)S_j(\theta))\Omega_t^{\star-1}(\theta)X_tX_t'\right) \\
&+ 2tr\left(V'(\theta)\left(\dot{S}_{i,j}(\theta) + S_i(\theta)S_j(\theta)\right)V(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&- 2tr\left(\tilde{S}_i(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + 2tr\left(\tilde{S}_i'(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right),
\end{aligned} \tag{A.18}$$

*where $S_i(\theta)$ and $\tilde{S}_i(\theta)$ are skew-symmetric matrices given by,*

$$S_i(\theta) = \frac{\partial V(\theta)}{\partial\theta_i}V'(\theta), \tag{A.19}$$

$$\tilde{S}_i'(\theta) = V'(\theta)S_i(\theta)V(\theta) = -V'(\theta)S_i'(\theta)V(\theta) = -\tilde{S}_i(\theta). \tag{A.20}$$

*Proof of Lemma A.3:* Throughout the proof, we suppress the dependence on $\theta$. From the proof of Lemma A.2 we have that

$$\frac{\partial^2 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j} = \frac{\partial\text{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star)}{\partial\theta_j} - \frac{\partial\text{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\Lambda_t^{\star-1}Y_tY_t')}{\partial\theta_j} + 2\frac{\partial\dot{Y}_{t,i}'\Lambda_t^{\star-1}Y_t}{\partial\theta_j}$$

$$= N_{1,t} - N_{2,t} + 2N_{3,t}.$$

Where the first term, $N_{1,t}$, is

$$\frac{\partial\text{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star)}{\partial\theta_j} = -\text{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\right) + \text{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^\star\right). \tag{A.21}$$

The second term, $N_{2,t}$, is

$$\frac{\partial \mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t')}{\partial\theta_j} = \mathrm{tr}\left(\frac{\partial \Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}}{\partial\theta_j}\Lambda_t^{\star-1}Y_tY_t' + \Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\frac{\partial\Lambda_t^{\star-1}Y_tY_t'}{\partial\theta_j}\right)$$

$$= -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right)$$

$$+ \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\left(\dot{Y}_{t,j}Y_t' + Y_t\dot{Y}_{t,j}'\right)\right) - \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right)$$

$$= -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right)$$

$$- \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\dot{Y}_{t,j}Y_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_t\dot{Y}_{t,j}'\right)$$

Noting that $D_{t,i} := \Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}$is symmetric and that $\dot{Y}_{t,i} = V'S_i'X_t$ with $S_i$ defined in (A.19),

$$\mathrm{tr}\left(D_{t,i}\dot{Y}_{t,j}Y_t'\right) = \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}V'S_j'X_tX_t'V\right)$$

$$= \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}V'S_j'VY_tY_t'\right)$$

$$= \mathrm{tr}\left(\tilde{S}_j'\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right),$$

with $\tilde{S}_j = V'S_jV$ defined in (A.20). Hence, the second term of the Hessian, $N_{2,t}$, is

$$\frac{\partial \mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t')}{\partial\theta_j} = -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right)$$

$$-\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + 2\mathrm{tr}\left(\tilde{S}_j'\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right). \tag{A.22}$$

The third term, $N_{3,t}$, is

$$\frac{\partial \dot{Y}_{t,i}'\Lambda_t^{\star-1}Y_t}{\partial\theta_j} = \frac{\partial\dot{Y}_{t,i}'}{\partial\theta_j}\Lambda_t^{\star-1}Y_t + \dot{Y}_{t,i}'\frac{\partial\Lambda_t^{\star-1}}{\partial\theta_j}Y_t + \dot{Y}_{t,i}'\Lambda_t^{\star-1}\frac{\partial Y_t}{\partial\theta_j}$$

$$= \ddot{Y}_{t,i,j}'\Lambda_t^{\star-1}Y_t - \dot{Y}_{t,i}'\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}Y_t + \dot{Y}_{t,i}'\Lambda_t^{\star-1}\dot{Y}_{t,j},$$

where $\ddot{Y}_{t,i,j}'$ is,

$$\ddot{Y}_{t,i,j}' = X_t'\frac{\partial S_iV}{\partial\theta_j} = X_t'\left(\dot{S}_{i,j} + S_iS_j\right)V,$$

where $\dot{S}_{i,j} = \partial S_i/\partial\theta_j$. Hence, the first term of $N_{3,t}$ is,

$$\begin{aligned}
\ddot{Y}'_{t,i,j}\Lambda_t^{\star-1}Y_t &= X'_t\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}V'X_t \\
&= X'_tVV'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}V'X_t \\
&= \mathrm{tr}\left(V'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}Y_tY'_t\right)
\end{aligned}$$

The second term of $N_{3,t}$ is

$$\dot{Y}'_{t,i}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_t = X'_tVV'S_iV\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}V'X_t = \mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_tY'_t\right)$$

And the final term is

$$\dot{Y}'_{t,i}\Lambda_t^{\star-1}\dot{Y}_{t,j} = X'_tS_iV\Lambda_t^{\star-1}V'S'_jX_t = X'_tVV'S_iV\Lambda_t^{\star-1}V'S'_jVV'X_t = \mathrm{tr}\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_jY_tY'_t\right).$$

That is, $N_{3,t}$ is

$$\frac{\partial\dot{Y}'_{t,i}\Lambda_t^{\star-1}Y_t}{\partial\theta_j} = \mathrm{tr}\left(V'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}Y_tY'_t\right) - \mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_tY'_t\right) + \mathrm{tr}\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_jY_tY'_t\right) \tag{A.23}$$

Using (A.21)-(A.23),

$$\begin{aligned}
\frac{\partial^2 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j} &= -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}^\star_{t,i,j}\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}Y_tY'_t\right) \\
&\quad - \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}^\star_{t,i,j}\Lambda_t^{\star-1}Y_tY'_t\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_tY'_t\right) - 2\mathrm{tr}\left(\tilde{S}'_j\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}Y_tY'_t\right) \\
&\quad + 2\mathrm{tr}\left((\dot{S}_{i,j} + S_iS_j)\Omega_t^{\star-1}X_tX'_t\right) + 2\mathrm{tr}\left(V'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}Y_tY'_t\right) - 2\mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_tY'_t\right) \\
&\quad + 2\mathrm{tr}\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_jY_tY'_t\right)
\end{aligned}$$

**Lemma A.4.** *With $l_t^\star(\theta)$ defined in (A.1), under Assumptions 3.1-3.8, for $i, j = 1, \ldots, d_\theta$,*

$$E\left[\left.\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta_i\partial\theta_j}\right|\mathcal{F}_{t-1}\right] = tr\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\right) + 2tr\left(S_iS_j\right) + 2tr\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star\right), \tag{A.24}$$

$$E\left[\left\|\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'}\right\|\right] < \infty, \tag{A.25}$$

*and*

$$J = E\left[\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'}\right] \text{ is invertible.} \tag{A.26}$$

*Proof of Lemma A.4:* Using the expression for $\partial^2 l_t^\star(\theta)/\partial\theta_i\partial\theta_j$ from Lemma A.3, we immediately have that

$$\begin{aligned} E\left[\left.\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta_i\partial\theta_j}\right|\mathcal{F}_{t-1}\right] &= \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\right) + 2\mathrm{tr}\left(\dot{S}_{i,j}\right) + 2\mathrm{tr}\left(S_iS_j\right) \\ &\quad + 2\mathrm{tr}\left(\tilde{S}_j\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\right) - 2\mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\right) + 2\mathrm{tr}\left(\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star\right). \end{aligned}$$

This expression can be simplified further as both $\dot{S}_{i,j}$, $\mathrm{tr}\left(\tilde{S}_j\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\right)$ and $\mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\right)$ are skew-symmetric, and hence $\mathrm{tr}(\dot{S}_{i,j}) = \mathrm{tr}\left(\tilde{S}_j\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\right) = \mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\right) = 0$, and we obtain (A.24).

Turning to (A.25), we consider each term in (A.24). Notice that $E\left[\left|\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\right)\right|\right] < \infty$ by Lemma A.6. Trivially, $\mathrm{tr}(S_iS_j)$ is bounded, since $\Theta$ is compact and $S_i$ is continuous in $\phi$. Lastly, consider $\mathrm{tr}\left(\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star\right)$,

$$\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star =$$

$$\begin{pmatrix} 0 & -\tilde{s}_{i,12} & \dots & -\tilde{s}_{i,1p} \\ \tilde{s}_{i,12} & 0 & \dots & -\tilde{s}_{i,2p} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{s}_{i,1p} & \tilde{s}_{i,2p} & \dots & 0 \end{pmatrix} \begin{pmatrix} \lambda_{1,t} & 0 & \dots & 0 \\ 0 & \lambda_{2,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{p,t} \end{pmatrix} \begin{pmatrix} 0 & \tilde{s}_{j,12} & \dots & \tilde{s}_{j,1p} \\ -\tilde{s}_{j,12} & 0 & \dots & \tilde{s}_{j,2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{s}_{j,1p} & -\tilde{s}_{j,2p} & \dots & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_{1,t}} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_{2,t}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\lambda_{p,t}} \end{pmatrix},$$

which has the trace,

$$\mathrm{tr}\left(\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star\right) = \sum_{k=1}^{p-1}\sum_{l=k+1}^{p}\tilde{s}_{i,kl}\tilde{s}_{j,kl}\left(\frac{\lambda_{k,t}^\star}{\lambda_{l,t}^\star} + \frac{\lambda_{l,t}^\star}{\lambda_{k,t}^\star}\right),$$

which is bounded in light of (A.17). We conclude that (A.25) holds.

By standard arguments, see e.g. Comte and Lieberman (2003) or Bardet and Wintenberger (2009), it suffices to show that there exists no $\gamma = (\gamma_1, ..., \gamma_{d_\theta})' \in \mathbb{R}^{d_\theta} \setminus \{0_{d_\theta\times 0}\}$, such that

$$\sum_{i=1}^{d_\theta}\gamma_i\mathrm{vec}\left(\frac{\partial\Omega_t^\star}{\partial\theta_i}\right) = 0_{p^2\times 1} \quad \text{a.s.}, \tag{A.27}$$

where we have suppressed the dependence on $\theta_0$. For simplicity, we consider the case $p = 2$ and emphasize that the arguments can, tediously, be extended to arbitrary dimension $p$. For the case $p=2$, $d_\theta = 11$ such that $\theta = (\omega_1, \omega_2, \alpha_{11}, \alpha_{21}, \alpha_{12}, \alpha_{22}, \beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \phi)'$, and we seek to show that there exists no $\gamma = (\gamma_1, ..., \gamma_{11})' \in \mathbb{R}^{11} \setminus \{0_{11}\}$, such that

$$\sum_{i=1}^{11} \gamma_i \text{vec}\left(\frac{\partial \Omega_t^\star}{\partial \theta_i}\right) = 0_4 \quad \text{a.s.} \tag{A.28}$$

We have that

$$\Omega_t^\star = \begin{pmatrix} \lambda_{1,t}^\star \cos^2\phi + \lambda_{2,t}^\star \sin^2\phi & \left(\lambda_{2,t}^\star - \lambda_{1,t}^\star\right)\cos\phi\sin\phi \\ \left(\lambda_{2,t}^\star - \lambda_{1,t}^\star\right)\cos\phi\sin\phi & \lambda_{2,t}^\star\cos^2\phi + \lambda_{1,t}^\star\sin^2\phi \end{pmatrix},$$

such that for $i = 1, \dots, 10$,

$$\begin{aligned} \frac{\partial \Omega_t^\star}{\partial \theta_i} &= V \frac{\partial \Lambda_t^\star}{\partial \theta_i} V' = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i} & 0 \\ 0 & \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}\cos^2\phi + \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i}\sin^2\phi & \left(\frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} - \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}\right)\cos\phi\sin\phi \\ \left(\frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} - \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}\right)\cos\phi\sin\phi & \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i}\cos^2\phi + \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}\sin^2\phi \end{pmatrix}, \end{aligned} \tag{A.29}$$

and for $i = 11$

$$\begin{aligned} \frac{\partial \Omega_t^\star}{\partial \theta_i} &= \frac{\partial \Omega_t^\star}{\partial \phi} = \begin{pmatrix} \frac{\partial \Omega_{t,11}^\star}{\partial \phi} & \frac{\partial \Omega_{t,12}^\star}{\partial \phi} \\ \frac{\partial \Omega_{t,12}^\star}{\partial \phi} & \frac{\partial \Omega_{t,22}^\star}{\partial \phi} \end{pmatrix}, \\ \frac{\partial \Omega_{t,11}^\star}{\partial \phi} &= \cos^2\phi \frac{\partial \lambda_{1,t}^\star}{\partial \phi} + \sin^2\phi \frac{\partial \lambda_{2,t}^\star}{\partial \phi} + (\lambda_{2,t}^\star - \lambda_{1,t}^\star)\sin 2\phi \\ \frac{\partial \Omega_{t,12}^\star}{\partial \phi} &= (\lambda_{2,t}^\star - \lambda_{1,t}^\star)\cos 2\phi + \left(\frac{\partial \lambda_{2,t}^\star}{\partial \phi} - \frac{\partial \lambda_{1,t}^\star}{\partial \phi}\right)\cos\phi\sin\phi \\ \frac{\partial \Omega_{t,22}^\star}{\partial \phi} &= \sin^2\phi \frac{\partial \lambda_{1,t}^\star}{\partial \phi} + \cos^2\phi \frac{\partial \lambda_{2,t}^\star}{\partial \phi} + (\lambda_{1,t}^\star - \lambda_{2,t}^\star)\sin 2\phi, \end{aligned}$$

where

$$\begin{aligned} &\frac{\partial \lambda_t^\star}{\partial w_1} = \sum_{j=0}^\infty B^j \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \frac{\partial \lambda_t^\star}{\partial w_2} = \sum_{j=0}^\infty B^j \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \frac{\partial \lambda_t^\star}{\partial \alpha_{11}} = \sum_{j=0}^\infty B^j \begin{pmatrix} y_{1,t-j-1}^2 \\ 0 \end{pmatrix}, \\ &\frac{\partial \lambda_t^\star}{\partial \alpha_{12}} = \sum_{j=0}^\infty B^j \begin{pmatrix} y_{2,t-j-1}^2 \\ 0 \end{pmatrix}, \frac{\partial \lambda_t^\star}{\partial \alpha_{21}} = \sum_{j=0}^\infty B^j \begin{pmatrix} 0 \\ y_{1,t-j-1}^2 \end{pmatrix}, \frac{\partial \lambda_t^\star}{\partial \alpha_{22}} = \sum_{j=0}^\infty B^j \begin{pmatrix} 0 \\ y_{2,t-j-1}^2 \end{pmatrix}, \\ &\frac{\partial \lambda_t^\star}{\partial \beta_{nm}} = \sum_{j=0}^\infty \left(\frac{\partial B^j}{\partial \beta_{nm}}\right)\left(\begin{pmatrix} w_1 \\ w_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}\begin{pmatrix} y_{1,t-j-1}^2 \\ y_{2,t-j-1}^2 \end{pmatrix}\right). \end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial \lambda_t^\star}{\partial \phi} &= \sum_{j=0}^{\infty} B^j A \begin{pmatrix} \frac{\partial}{\partial \phi} y_{1,t-j-1}^2 \\ \frac{\partial}{\partial \phi} y_{2,t-j-1}^2 \end{pmatrix} = 2 \sum_{j=0}^{\infty} B^j A \begin{pmatrix} -y_{1,t-j-1} y_{2,t-j-1} \\ y_{1,t-j-1} y_{2,t-j-1} \end{pmatrix} \\
&= 2 \sum_{j=0}^{\infty} B^j \begin{pmatrix} (\alpha_{12} - \alpha_{11}) \\ (\alpha_{22} - \alpha_{21}) \end{pmatrix} y_{1,t-j-1} y_{2,t-j-1} \\
&= 2 \sum_{j=0}^{\infty} \begin{pmatrix} (B^j)_{11}(\alpha_{12} - \alpha_{11}) + (B^j)_{12}(\alpha_{22} - \alpha_{21}) \\ (B^j)_{21}(\alpha_{12} - \alpha_{11}) + (B^j)_{22}(\alpha_{22} - \alpha_{21}) \end{pmatrix} y_{1,t-j-1} y_{2,t-j-1}
\end{aligned}
$$

Hence, the first row of (A.28) has the form

$$
\begin{aligned}
& C_0 + \sum_{j=0}^{\infty} \left( C_{1,j} y_{1,t-j-1}^2 + C_{2,j} y_{2,t-j-1}^2 \right) \\
& + \gamma_{11} 2 \cos^2 \phi \left( \sum_{j=0}^{\infty} \left( (B^j)_{11}(\alpha_{12} - \alpha_{11}) + (B^j)_{12}(\alpha_{22} - \alpha_{21}) \right) y_{1,t-j-1} y_{2,t-j-1} \right) \\
& + \gamma_{11} 2 \sin^2 \phi \left( \sum_{j=0}^{\infty} \left( (B^j)_{21}(\alpha_{12} - \alpha_{11}) + (B^j)_{22}(\alpha_{22} - \alpha_{21}) \right) y_{1,t-j-1} y_{2,t-j-1} \right) \\
& = 0 \quad \text{almost surely,}
\end{aligned}
$$

where the constants $C_1$, $C_{2,j}$, $C_{3,j}$ may depend on $\gamma_1, \ldots, \gamma_{10}$. Suppose that $\gamma_{11} \neq 0$. By Assumption 3.3 we have that $y_{1,t-j-1} y_{2,t-j-1}$ is non-degenerate and linearly independent of $y_{1,t-j-1}^2$ and $y_{2,t-j-1}^2$, so it must hold that

$$
\begin{aligned}
& \gamma_{11} 2 \cos^2 \phi \left( \sum_{j=0}^{\infty} \left( (B^j)_{11}(\alpha_{12} - \alpha_{11}) + (B^j)_{12}(\alpha_{22} - \alpha_{21}) \right) y_{1,t-j-1} y_{2,t-j-1} \right) \\
& + \gamma_{11} 2 \sin^2 \phi \left( \sum_{j=0}^{\infty} \left( (B^j)_{21}(\alpha_{12} - \alpha_{11}) + (B^j)_{22}(\alpha_{22} - \alpha_{21}) \right) y_{1,t-j-1} y_{2,t-j-1} \right) \\
= \;& 0 \quad \text{almost surely.}
\end{aligned}
$$

This implies that

$$
\gamma_{11} 2 \left( \cos^2 \phi (\alpha_{12} - \alpha_{11}) + \sin^2 \phi (\alpha_{22} - \alpha_{21}) \right) y_{1,t-1} y_{2,t-1} | \mathcal{F}_{t-2}^{\eta} \text{ is degenerate}
$$

which is the case if and only if

$$\cos^2\phi(\alpha_{12}-\alpha_{11})+\sin^2\phi(\alpha_{22}-\alpha_{21})=0. \tag{A.30}$$

The same reasoning applied to the second and third rows of (A.28) yields that

$$\gamma_{11}2\cos\phi\sin\phi\left((\alpha_{22}-\alpha_{21})-(\alpha_{12}-\alpha_{11})\right)y_{1,t-j-1}y_{2,t-j-1}|\mathcal{F}^{\eta}_{t-2} \text{ is degenerate}$$

and hence, using that $\cos\phi$ and $\sin\phi$ are non-zero on int$\Theta$, that

$$(\alpha_{22}-\alpha_{21})-(\alpha_{12}-\alpha_{11})=0 \Leftrightarrow (\alpha_{22}-\alpha_{21})=(\alpha_{12}-\alpha_{11}). \tag{A.31}$$

Combining (A.30) and (A.31), we have that $\alpha_{12}=\alpha_{11}$ and $\alpha_{22}=\alpha_{21}$, which is ruled out by Assumption 3.8, and we conclude that (A.28) only holds whenever $\gamma_{11}=0$. Hence (A.28) has the form

$$\sum_{i=1}^{10}\gamma_i\text{vec}\left(\frac{\partial\Omega_t^\star}{\partial\theta_i}\right)=(V\otimes V)\sum_{i=1}^{10}\gamma_i\text{vec}\left(\frac{\partial\Lambda_t^\star}{\partial\theta_i}\right)=0_4 \quad \text{a.s.},$$

which, using that $V$ has full rank, implies that

$$\sum_{i=1}^{10}\gamma_i\text{vec}\left(\frac{\partial\Lambda_t^\star}{\partial\theta_i}\right)=0_4 \quad \text{a.s.}$$

The non-zero rows of $\text{vec}(\partial\Lambda_t^\star/\partial\theta)_i$, $i=1,\dots,10$, are

$$\frac{\partial\lambda_t^\star}{\partial\theta_i}=\sum_{j=0}^{\infty}\frac{\partial B^j}{\partial\theta_i}\left(\frac{\partial W}{\partial\theta_i}+\frac{\partial A}{\partial\theta_i}(V'X_{t-1-j})^{\odot 2}\right),$$

and by arguments similar to the ones given in Francq and Zakoïan (2019, pp. 311-312), it follows that there exist no non-zero $\gamma$ such that (A.28) holds. We conclude that $J$ is invertible.

**Lemma A.5.** *With $l_t^\star(\theta)$ defined in (A.1), suppose that Assumptions 3.1-3.8 hold. Then there exists a neighborhood around $\theta_0$, $N(\theta_0)\subset\Theta$, such that*

$$\max_{h,i,j=1,\dots,d_\theta} E\left[\sup_{\theta\in N(\theta_0)}\left|\frac{\partial^3 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|\right]<\infty.$$

*Proof of Lemma A.5:* Throughout, we exploit that $\theta_0\in\text{int}\Theta$ such that $N(\theta_0)$ satisfies that all entries of $A$ and $B$ are bounded away from zero on $N(\theta_0)$. In the following, for some real-valued random variable $f_t(\theta)$ depending on $\theta\in N(\theta_0)$, we write $f_t(\theta)\in\mathcal{L}_{N(\theta_0)}$ if

$E[\sup_{\theta \in N(\theta_0)} |f_t(\theta)|] < \infty$ and we say that $f_t(\theta)$ belongs to $\mathcal{L}_{N(\theta_0)}$.

Consider the $(i, j, k)$'th element of the array of third derivatives of the log-likelihood function, which is obtained by taking the derivative of the $(i, j)$th element of the Hessian in (A.18) with respect to some parameter $\theta_k$:

$$
\begin{aligned}
\frac{\partial^3 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} = & -\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\right) & (\#1)\\
& + \frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\right) & (\#2)\\
& + \frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#3)\\
& - \frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#4)\\
& + \frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#5)\\
& - 2\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#6)\\
& + 2\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(V'(\theta)\left(\dot{S}_{i,j}(\theta)+S_i(\theta)S_j(\theta)\right)V(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#7)\\
& - 2\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\tilde{S}_i(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) & (\#8)\\
& + 2\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\tilde{S}_i'(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right). & (\#9)
\end{aligned}
$$

In the following, we consider each partial derivative in turn, and show that all terms belong to $\mathcal{L}_{N(\theta_0)}$.

**Term #1** The partial derivative is,

$$
\begin{aligned}
& \frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}\right)\\
& = -2\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\right)\\
& + \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,k}^\star(\theta)\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,j,k}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\right). \qquad (A.32)
\end{aligned}
$$

Noting that $\mathrm{tr}\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,k}^\star(\theta)\} = \sum_{s=1}^p \dot{\lambda}_{s,t,i}^\star(\theta)\ddot{\lambda}_{s,t,i,j}^\star(\theta)/\lambda_{s,t}^{\star 2}(\theta)$, we conclude that the second term in (A.32) belongs to $\mathcal{L}_{N(\theta_0)}$. The same argument applies to the other terms in (A.32).

**Term #2** The second term is,

$$\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}(\theta)\right) = -\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}(\theta)\right)+\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dddot{\Lambda}_{t,i,j,k}(\theta)\right),$$

and we apply arguments similar to the ones given with respect to Term # 1 in order to conclude that Term #2 belongs to $\mathcal{L}_{N(\theta_0)}$.

**Terms #3 and #5** Terms #3 and #5 are the same up to indexing, and we here show that #3 has finite expectation uniformly on $N(\theta_0)$.

$$\begin{aligned}
&\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) = \\
&-3\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,k}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&+\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}^{\star}_{t,j,k}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&+\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}^{\star}_{t,i,k}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) \\
&+2\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)\dot{Y}_{t,k}'(\theta)\right).
\end{aligned} \tag{A.33}$$

Note that the first term in (A.33) we may use that $Y_t(\theta) = V'(\theta)X_t$, where $X_t = V\Lambda_t^{\star 1/2}\eta_t$ (with $V = V(\theta_0)$ and $\Lambda_t^{\star 1/2} = \Lambda_t^{\star 1/2}(\theta_0)$), such that

$$\begin{aligned}
&\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,k}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)V'(\theta)V\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}V'V(\theta)\right) = \\
&\mathrm{vec}(V'(\theta)V)'(\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}\otimes\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,k}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}(\theta)\Lambda_t^{\star-1}(\theta)) \\
&\times\mathrm{vec}(V'V(\theta)).
\end{aligned} \tag{A.34}$$

Since $\mathrm{vec}(V'(\theta)V)$ consist of rotations based on trigonometric functions, it is bounded on $N(\theta_0)$. Next, note that the quantity $\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}\otimes\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,k}\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,j}\Lambda_t^{\star-1}(\theta)\dot{\Lambda}^{\star}_{t,i}\Lambda_t^{\star-1}$ entering (A.34) is a symmetric $p^2\times p^2$ matrix, with $p\times p$ blocks, $Q_{g,h}$, $g,h = 1,\ldots,p$, each of which are diagonal,

$$Q_{g,h} = \mathrm{diag}\left(\lambda_{g,t}^{\star 1/2}\eta_{g,t}\lambda_{h,t}^{\star 1/2}\eta_{h,t}\frac{\dot{\lambda}^{\star}_{s,t,i}(\theta)\dot{\lambda}^{\star}_{s,t,j}(\theta)\dot{\lambda}^{\star}_{s,t,k}(\theta)}{\lambda_{s,t}^{\star 4}(\theta)}\right),$$

for $s = 1,\ldots,p$, where $\dot{\lambda}^{\star}_{s,t,i}(\theta)\dot{\lambda}^{\star}_{s,t,j}(\theta)\dot{\lambda}^{\star}_{s,t,k}(\theta)/\lambda_{s,t}^{\star 3}(\theta)$ has finite $r$th moment for any $r > 0$ by Lemma A.6. Notice however that such property does not appear to apply to

$\lambda_{g,t}^{\star 1/2}\eta_{g,t}\lambda_{h,t}^{\star 1/2}\eta_{h,t}/\lambda_{s,t}^{\star}(\theta)$ for $g = h \neq s$ as the numerator and denominator are evaluated in $\theta_0$ and $\theta$ respectively. Instead we note that $\sup_{\theta \in N(\theta_0)} |\lambda_{g,t}^{\star 1/2}\eta_{g,t}\lambda_{h,t}^{\star 1/2}\eta_{h,t}/\lambda_{s,t}^{\star}(\theta)| \leq K\|\eta_t\|^2\|\lambda_t(\theta_0)\|$, and use Assumption 3.7, Lemma A.6, and Hölder's inequality in order to ensure that any entry of $Q_{g,h}$ belongs to $\mathcal{L}_{N(\theta_0)}$. The three other parts of Term #3 can be shown to belong to $\mathcal{L}_{N(\theta_0)}$ using similar arguments. To illustrate, consider

$$\operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)\dot{Y}_{t,k}'(\theta)\right) =$$

$$\operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)V'(\theta)V\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}V'S_k(\theta)V(\theta)\right) =$$

$$\operatorname{vec}(V'(\theta)S_k'(\theta)V)(\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2} \otimes \Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta))\operatorname{vec}(V'(\theta)V),$$

which belongs to $\mathcal{L}_{N(\theta_0)}$, applying the same arguments as for (A.34).

**Term #4** The derivative is,

$$\frac{\partial}{\partial\theta_k}\operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$= -2\operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+ \operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\dddot{\Lambda}_{t,i,j,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+ 2\operatorname{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)\dot{Y}_{t,k}'\right),$$

and it belongs to $\mathcal{L}_{N(\theta_0)}$, applying the same arguments as used for Terms #1 and #3.

**Terms #6 and #8** These terms are the same up to indexing. The partial derivative in Term #6 is,

$$\frac{\partial}{\partial\theta_k}\operatorname{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$= \operatorname{tr}\left(\dot{\tilde{S}}_{j,k}'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$-\operatorname{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+\operatorname{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$-\operatorname{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+\operatorname{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\left(\dot{Y}_{t,k}Y_t(\theta)' + Y_t(\theta)\dot{Y}_{t,k}'\right)\right),$$

and, again, this can be shown to belong to $\mathcal{L}_{N(\theta_0)}$ as Terms # 1, # 3 and # 4.

**Term #7** For simplicity, define $\bar{S}_{i,j}(\theta) := V'(\theta)\left(\dot{S}_{i,j}(\theta) + S_i(\theta)S_j(\theta)\right)V(\theta)$

$$\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\bar{S}_{ij}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) = \mathrm{tr}\left(\dot{\bar{S}}_{i,j,k}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$-\,\mathrm{tr}\left(\bar{S}_{i,j}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+\,\mathrm{tr}\left(\bar{S}_{i,j}(\theta)\Lambda_t^{\star-1}(\theta)\left(\dot{Y}_{t,k}(\theta)Y_t(\theta) + Y_t(\theta)\dot{Y}_{t,k}'(\theta)\right)\right),$$

which belongs to $\mathcal{L}_{N(\theta_0)}$ by the same arguments as for Terms #1, #3, #4 and #6.

**Term #9** Note that

$$\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\tilde{S}_i'(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$= \mathrm{tr}\left(\dot{\tilde{S}}_{i,k}'(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$-\,\mathrm{tr}\left(\tilde{S}_i'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+\,\mathrm{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\tilde{S}}_{j,k}(\theta)Y_t(\theta)Y_t'(\theta)\right)$$
$$+\,\mathrm{tr}\left(\tilde{S}_i'(\theta)\Lambda_t^{\star}(\theta)\tilde{S}_j(\theta)\left(\dot{Y}_{t,k}(\theta)Y_t'(\theta) + Y_t(\theta)\dot{Y}_{t,k}'(\theta)\right)\right).$$

This term also belong to $\mathcal{L}_{N(\theta_0)}$ per the arguments used above.

**Lemma A.6.** *With $\lambda_t^\star(\theta)$ defined in (A.3), let $\lambda_{h,t}^\star(\theta)$ denote its hth entry. For $i,j,k = 1,\ldots,d_\theta$, let*

$$\dot{\lambda}_{h,t,i}^{\star}(\theta) = \frac{\partial\lambda_{h,t}^\star}{\partial\theta_i}, \quad \ddot{\lambda}_{h,t,i,j}^{\star}(\theta) = \frac{\partial^2\lambda_{h,t}^\star}{\partial\theta_i\partial\theta_j}, \quad \textit{and} \quad \dddot{\lambda}_{h,t,i,j}^{\star}(\theta) = \frac{\partial^3\lambda_{h,t}^\star}{\partial\theta_i\partial\theta_j\partial\theta_k}.$$

*Under Assumptions 3.1-3.8, for any $r>0$, $i,j,k = 1,\ldots,d_\theta$, and $h = 1,\ldots,p$ there exists a neighborhood $N(\theta_0)\subset\Theta$ of $\theta_0$ such that*

$$E\left[\sup_{\theta\in N(\theta_0)}\left|\frac{\dot{\lambda}_{h,t,i}^{\star}(\theta)}{\lambda_{h,t}^\star(\theta)}\right|^r\right] < \infty, \quad E\left[\sup_{\theta\in N(\theta_0)}\left|\frac{\ddot{\lambda}_{h,t,i,j}^{\star}(\theta)}{\lambda_{h,t}^\star(\theta)}\right|^r\right] < \infty, \quad \textit{and} \quad E\left[\sup_{\theta\in N(\theta_0)}\left|\frac{\dddot{\lambda}_{h,t,i,j,k}^{\star}(\theta)}{\lambda_{h,t}^\star(\theta)}\right|^r\right] < \infty.$$

*Proof of Lemma A.6:* Throughout, we exploit that $\theta_0 \in \mathrm{int}\Theta$ such that $N(\theta_0)$ satisfies that all entries of $A$ and $B$ are bounded away from zero on $N(\theta_0)$. We start out by considering the first-order derivatives $\dot{\lambda}^{\star}_{h,t,i}(\theta)/\lambda^{\star}_{h,t}(\theta)$. With $Y_t = V(\theta)'X_t$, and suppressing the dependence on $\theta$,

$$\lambda^{\star}_t = \sum_{j=1}^{\infty}\left(\underbrace{B^{j-1}W}_{:=C_1^{(j-1)}} + \underbrace{B^{j-1}A}_{:=C_2^{(j-1)}} Y^{\odot 2}_{t-j}\right) = \sum_{j=1}^{\infty}\left(C_1^{(j-1)} + C_2^{(j-1)}Y^{\odot 2}_{t-j}\right),$$

which has derivatives

$$\begin{aligned}
\frac{\partial \lambda^{\star}_t}{\partial \omega_i} &= \sum_{j=1}^{\infty}\dot{C}^{(j-1)}_{1,i}, & \dot{C}^{(j-1)}_{1,i} &= B^{j-1}\frac{\partial W}{\partial \omega_i},\\
\frac{\partial \lambda^{\star}_t}{\partial \alpha_i} &= \sum_{j=1}^{\infty}\dot{C}^{(j-1)}_{2,i}Y^{\odot 2}_{t-j}, & \dot{C}^{(j-1)}_{2,i} &= B^{j-1}\frac{\partial A}{\partial \alpha_i},\\
\frac{\partial \lambda^{\star}_t}{\partial \beta_i} &= \sum_{j=1}^{\infty}\dot{C}^{(j-1)}_{3,i}(W + AY^{\odot 2}_{t-j}), & \dot{C}^{(j-1)}_{3,i} &= \frac{\partial B^{j-1}}{\partial \beta_i} = \sum_{k=1}^{j-1}B^{k-1}\frac{\partial B}{\partial \beta_i}B^{j-1-k},\\
\frac{\partial \lambda^{\star}_t}{\partial \phi_i} &= 2\sum_{j=1}^{\infty}C_2^{(j-1)}(Y_{t-j}\odot \tilde{S}_i Y_{t-j}),
\end{aligned} \tag{A.35}$$

where $\omega_i$, $\alpha_i$, $\beta_i$, $\phi_i$ denote arbitrary entries of, respectively, $W$, $A$, $B$, $\phi$, and where $\tilde{S}_i$ is defined in (A.20).

We now verify that $\sup_{\theta\in N(\theta_0)}|\dot{\lambda}^{\star}_{s,t}/\lambda^{\star}_{s,t}|^r < \infty$ has finite expectation by considering $\omega_i, \alpha_i, \beta_i$ and $\phi_i$ in (i)–(iv) below.

**(i)** Consider first $\theta_i = \omega_i$. Here

$$\frac{\partial \lambda^{\star}_{s,t}/\partial \omega_i}{\lambda^{\star}_{s,t}} = \frac{\sum_{j=1}^{\infty}[\dot{C}^{(j-1)}_{1,i}]_s}{\sum_{j=1}^{\infty}\left([C_1^{(j-1)}]_s + \sum_{h=1}^{p}[C_2^{(j-1)}]_{s,h}y^2_{h,t-j}\right)} \leq \sum_{j=1}^{\infty}\frac{[\dot{C}^{(j-1)}_{1,i}]_s}{\omega_L} \leq \sum_{j=1}^{\infty}\frac{\varrho^{j-1}}{\omega_L} \leq K,$$

where we have used that $\lambda^{\star}_{s,t} \geq \omega_L$ and $\sup_{\theta\in\Theta}\rho(B) < 1$.

**(ii)** Next, consider $\theta_i = \alpha_i$. Since $\partial\lambda^{\star}_t/\partial\alpha_i = \sum_{j=1}^{\infty}\dot{C}^{(j-1)}_{2,i}Y^{\odot 2}_{t-j}$, with $\dot{C}^{(j-1)}_{2,i} = B^{j-1}\partial A/\partial\alpha_i$. Here $\partial A/\partial \alpha_i$ is a matrix of zeros except for a 1 in the place of $\alpha_i$ in $A$. We can therefore use that, elementwise,

$$\alpha_i\frac{\partial \lambda^{\star}_t}{\partial \alpha_i} \leq \lambda^{\star}_t.$$

Hence, for $s = 1, \dots, p$,

$$\left|\frac{\partial \lambda^{\star}_{s,t}/\partial \alpha_i}{\lambda^{\star}_{s,t}}\right| \leq K.$$

**(iii)** Next, consider $\theta_i = \beta_i$. Let $\bar{C}_{t-j} = W + AY_{t-j}^{\odot 2}$, and notice that

$$\frac{\partial \lambda_t^\star}{\partial \beta_i} = \sum_{j=1}^{\infty} \left( \sum_{k=1}^{j} B^{k-1} \frac{\partial B}{\partial \beta_i} B^{j-k} \bar{C}_{t-j} \right),$$

where $\partial B / \partial \beta_i$ is a matrix of zeros, apart a one in the same place as $\beta_i$ in $B$. We can therefore apply the inequality, (with $\beta_i > 0$ uniformly on $N(\theta_0)$),

$$\beta_i \frac{\partial \lambda_t^\star}{\partial \beta_i} \leq \sum_{j=1}^{\infty} j B^j \bar{C}_{t-j},$$

which elementwise corresponds to,

$$\beta_i \frac{\partial \lambda_{s,t}^\star}{\partial \beta_i} \leq \sum_{j=1}^{\infty} j \sum_{h=1}^{p} [B^j]_{s,h} [\bar{C}_{t-j}]_h.$$

Recall furthermore that,

$$\lambda_{s,t}^\star \geq \omega_L + \sum_{h=1}^{p} [B^j]_{s,h} [\bar{C}_{t-j}]_h,$$

with $\omega_L = \min_{s=1,\dots,p} \inf_{\theta \in N(\theta_0)} w_s > 0$. Lastly, we use the inequality $x/(1+x) \leq x^k$ for all $x \geq 0$ and $k \in (0,1)$, such that,

$$\begin{aligned} \beta_i \tfrac{\partial \lambda_{s,t}^\star / \partial \beta_i}{\lambda_{s,t}^\star} &\leq \tfrac{\sum_{j=1}^{\infty} j \sum_{h=1}^{p} [B^j]_{s,h} [\bar{C}_{t-j}]_s}{\omega_L + \sum_{h=1}^{p} [B^j]_{s,h} [\bar{C}_{t-j}]_s} \leq \sum_{j=1}^{\infty} \sum_{h=1}^{p} j \left( \tfrac{[B^j]_{sh} [\bar{C}_{t-j}]_h}{\omega_L} \right)^k \\ &= \sum_{j=1}^{\infty} \sum_{h=1}^{p} j [B^j]_{sh}^k \left( \tfrac{[\bar{C}_{t-j}]_h}{\omega_L} \right)^k \leq K \sum_{j=1}^{\infty} j \varrho^j \sum_{h=1}^{p} \left( \tfrac{[\bar{C}_{t-j}]_h}{\omega_L} \right)^k \end{aligned}$$

Using that $\sup_{\theta \in \Theta} \rho(B) < 1$, for any $r > 0$, we can choose $k > 0$ sufficiently small, such that $E[\sup_{\theta \in N(\theta_0)} |(\partial \lambda_{s,t}^\star / \partial \beta_i) / \lambda_{s,t}|^r] < \infty$, where we have used Assumption 3.2 that $\|X_t\|$ has some finite (potentially fractional) moment.

**(iv)** Finally, consider $\theta_i = \phi_i$. The partial derivative $\partial \lambda_t^\star / \partial \phi_i$ in (A.35) contains the matrix product $\tilde{S}_i Y_{t-n}$, where the $j$th row of $\tilde{S}_i Y_{t-n}$ is

$$\left[ \tilde{S}_i Y_{t-n} \right]_j = -\sum_{k=1}^{j-1} \tilde{s}_{i,kj} y_{k,t-n} + \sum_{k=j+1}^{p} \tilde{s}_{i,jk} y_{k,t-n}.$$

Hence,

$$\left[Y_{t-n} \odot \tilde{S}_i Y_{t-n}\right]_j = y_{j,t-1}\left(-\sum_{k=1}^{j-1} \tilde{s}_{i,kj} y_{k,t-n} + \sum_{k=j+1}^{p} \tilde{s}_{i,jk} y_{k,t-n}\right),$$

and we have that

$$\begin{aligned} |[Y_{t-n} \odot \tilde{S}_i Y_{t-n}]_s| &\leq K\left(\sum_{k=1}^{s-1} |y_{s,t-n}||y_{k,t-n}| + \sum_{h=s+1}^{p} |y_{s,t-n}||y_{h,t-n}|\right) \\ &\leq pK\|Y_{t-n}\|^2, \end{aligned}$$

where we have used the simple inequality that $a^2 + b^2 \geq |ab|$ for $a, b \in \mathbb{R}$. Hence, for $s = 1, \ldots, p$,

$$\frac{\partial \lambda^{\star}_{s,t}}{\partial \phi_i} \leq pK \sum_{j=1}^{\infty} \sum_{h=1}^{p} [C_2^{(j-1)}]_{s,h} \|Y_{t-j}\|^2.$$

Note that on $N(\theta_0)$, elementwise,

$$C_2^{(j-1)} = B^{j-1} A \leq \alpha_U B^{j-1}(\iota_p, \ldots, \iota_p),$$

where $\iota_p$ is a $p$-dimensional column vector of ones. Then, with $[B^{j-1}]_s$ the $s$th row of $[B^{j-1}]$, $\sum_{h=1}^{p} [C_2^{(j-1)}]_{s,h} \leq p\alpha_U [B^{j-1}]_s \iota_p$, and we have that

$$\left|\frac{\partial \lambda^{\star}_{s,t}}{\partial \phi_i}\right| \leq Kp^2 \alpha_U \sum_{j=1}^{\infty} [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2. \tag{A.36}$$

Moreover, since the entries of $A$ are bounded away from zero on $N(\theta_0)$, the entries are also bounded away from some (small) constant $\alpha_L > 0$, and we have that $[C_2^{(j-1)}]_{s,h} \geq \alpha_L [B^{j-1}]_s \iota_p$ for $h, s = 1, \ldots, p$. Hence for any $j \geq 1$, and $s = 1, \ldots, p$,

$$\begin{aligned} \lambda^{\star}_{s,t} &= \sum_{j=1}^{\infty} [B^{j-1}]_s W + \sum_{j=1}^{\infty} \sum_{h=1}^{p} [C_2^{(j-1)}]_{s,h} y^2_{h,t-j} \geq \bar{\omega} + \sum_{j=1}^{\infty} \sum_{h=1}^{p} \alpha_L [B^{j-1}]_s \iota_p y^2_{h,t-j} \\ &= \bar{\omega} + \alpha_L \sum_{j=1}^{\infty} [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2 \geq \bar{\omega} + \alpha_L [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2, \end{aligned} \tag{A.37}$$

where $\bar{\omega} = \min_{s=1,\ldots,p} \inf_{\theta \in N(\theta_0)} \sum_{j=1}^{\infty} [B^{j-1}]_s W > 0$. Combining (A.36) and (A.37), we have that for $s = 1, \ldots, p$ and $k \in (0,1)$

$$\begin{aligned}
\left|\frac{\partial \lambda^\star_{s,t}/\partial \phi_i}{\lambda^\star_{s,t}}\right| &\leq Kp^2 \alpha_U \sum_{j=1}^{\infty} \frac{[B^{j-1}]_s \iota_p \|Y_{t-j}\|^2}{\bar{\omega} + \alpha_L [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2} \\
&= Kp^2 \frac{\alpha_U}{\alpha_L} \sum_{j=1}^{\infty} \frac{[B^{j-1}]_s \iota_p \|Y_{t-j}\|^2}{\bar{\omega}/\alpha_L + [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2} \\
&\leq Kp^2 \frac{\alpha_U}{\alpha_L} \sum_{j=1}^{\infty} \left(\frac{[B^{j-1}]_s \iota_p \|Y_{t-j}\|^2}{\bar{\omega}/\alpha_L}\right)^k \\
&\leq Kp^2 \frac{\alpha_U}{\alpha_L} \sum_{j=1}^{\infty} \varrho^{j-1} \left(\frac{\|Y_{t-j}\|^2}{\bar{\omega}/\alpha_L}\right)^k,
\end{aligned}$$

and we may again choose $k > 0$ sufficiently small such that $E[\sup_{\theta \in N(\theta_0)} |(\partial \lambda^\star_{s,t}/\partial \phi_i)/\lambda^\star_{s,t}|^r] < \infty$. The integrability of $\sup_{\theta \in N(\theta_0)} |\ddot{\lambda}^\star_{h,t,i,j}(\theta)/\lambda^\star_{h,t}(\theta)|^r$ and $\sup_{\theta \in N(\theta_0)} |\dddot{\lambda}^\star_{h,t,i,j,k}(\theta)/\lambda^\star_{h,t}(\theta)|^r$ are shown to hold by similar arguments.

# B Testing for Nullity of Rows

In this Section we first consider sufficient regularity conditions under which the asymptotic distribution of the (sup) likelihood ratio statistic for the hypothesis $\mathsf{H}_2^*$ in (19) can be derived. In Section B.2, the implementation of the test is discussed.

## B.1 Zero-rows in $A$ and $B$

Recall from Section 4.1 that when testing the hypothesis $\mathsf{H}_2^*$ in (19) that $(\theta, \delta) \in \Theta \times \Theta_{\text{sup}}$, where $\delta = (B_{13}, B_{23}, B_{33})'$ denotes the unidentified parameters, while $\theta \in \Theta$ denotes the remaining $d_\theta = 21$ parameters. As in Appendix A.1, consider the stationary and ergodic version of the log-quasi-likelihood contributions given by,

$$\begin{aligned}
l^\star_t(\theta, \delta) &= \log \det(\Omega^\star_t(\theta, \delta)) + X_t' \Omega^{\star -1}_t(\theta, \delta) X_t, \\
\Omega^\star_t(\theta, \delta) &= V(\theta) \Lambda^\star_t(\theta, \delta) V(\theta)', \quad \Lambda^\star_t(\theta, \delta) = \operatorname{diag}(\lambda^\star_t(\theta, \delta)), \\
\lambda^\star_t(\theta, \delta) &= W + A(V(\theta)' X_{t-1})^{\odot 2} + B \lambda^\star_{t-1}(\theta, \delta).
\end{aligned}$$

The limiting distribution of the supLR statistic in (21) can be derived under the following conditions, see Andrews (2001) for details and Pedersen and Rahbek (2019) for an application to GARCH-X models.

**(i)** With $\tilde{\theta}_{T,\delta}$ and $\hat{\theta}_{T,\delta}$ defined in (20), assume that $\tilde{\theta}_{T,\delta}, \hat{\theta}_{T,\delta} \xrightarrow{p} \theta_0$.

**(ii)** Assume that $T^{-1/2} \sum_{t=1}^{T} \partial l_t^\star(\theta, \cdot)/\partial\theta \xrightarrow{w} G.$, where $G.$ is a mean zero $d_\theta$ dimensional Gaussian process with kernel

$$\Sigma_{\delta_1 \delta_2} = E(\tfrac{\partial l_t^\star(\theta_0,\delta_1)}{\partial\theta} \tfrac{\partial l_t^\star(\theta_0,\delta_1)}{\partial\theta'}), \quad \text{for } \delta_1, \delta_2 \in \Theta_{\text{sup}}. \tag{B.38}$$

**(iii)** For any $\delta \in \Theta_{\text{sup}}$, $T^{-1}\partial^2 l_t^\star(\theta_0, \delta)/\partial\theta\partial\theta' \xrightarrow{p} J_\delta$, where

$$J_\delta = E(\tfrac{\partial^2 l_t^\star(\theta_0,\delta)}{\partial\theta\partial\theta'}), \tag{B.39}$$

with $J_\delta$ invertible uniformly on $\Theta_{\text{sup}}$.

**(iv)** The sets $\Theta - \theta_0$ and $\Theta^* - \theta_0$ are locally equal to some convex cones $C$ and $C^*$, respectively.[1]

**(v)** There exists a neighborhood $N(\theta_0)$ of $\theta_0$ such that

$$\sup_{\delta\in\Theta_{\text{sup}}} \left\| T^{-1/2} \sum_{t=1}^{T} \left( \tfrac{\partial l_t(\theta_0,\delta)}{\partial\theta} - \tfrac{\partial l_t^\star(\theta_0,\delta)}{\partial\theta} \right) \right\| \xrightarrow{p} 0,$$

and

$$\sup_{\delta\in\Theta_{\text{sup}}, \theta\in N(\theta_0)\cap\Theta} \left\| T^{-1} \sum_{t=1}^{T} \left( \tfrac{\partial^2 l_t(\theta,\delta)}{\partial\theta\partial\theta'} - \tfrac{\partial^2 l_t^\star(\theta,\delta)}{\partial\theta\partial\theta'} \right) \right\| \xrightarrow{p} 0.$$

**(vi)** For any fixed $\delta \in \Theta_{\text{sup}}$, and any deterministic scalar sequence $(\epsilon_T : T = 1, 2, ...)$ with $\epsilon_T \to 0$,

$$\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\epsilon_T} \left\| T^{-1} \sum_{t=1}^{T} \left( \tfrac{\partial^2 l_t^\star(\theta,\delta)}{\partial\theta\partial\theta'} - \tfrac{\partial^2 l_t^\star(\theta_0,\delta)}{\partial\theta\partial\theta'} \right) \right\| \xrightarrow{p} 0.$$

By Andrews (2001, Theorem 4), under conditions (i)-(vi) and $\mathsf{H}_2^*$,

$$\sup \text{LR}_T(\mathsf{H}_2^*) \xrightarrow{d} \sup_{\delta\in\Theta_{\text{sup}}} \{\lambda_\delta' J_\delta \lambda_\delta\} - \sup_{\delta\in\Theta_{\text{sup}}} \{\lambda_\delta^{*\prime} J_\delta \lambda_\delta^*\}, \tag{B.40}$$

[1] The set $\Theta - \theta_0$ is locally equal to $C$ if there exists a $\varepsilon > 0$ such that $\{\Theta - \theta_0\} \cap H(0, \varepsilon) = C \cap H(0, \varepsilon)$ where $H(0, \varepsilon) \subset \mathbb{R}^{\dim\theta}$ is an open cube centered at zero and with side length $2\varepsilon$.

where

$$\lambda_\delta = \arg\inf_{\eta\in C}\left\{(\eta - Z_\delta)' J_\delta (\eta - Z_\delta)\right\},$$

$$\lambda^*_\delta = \arg\inf_{\eta\in C^*}\left\{(\eta - Z_\delta)' J_\delta (\eta - Z_\delta)\right\}$$

and $Z_\delta = J_\delta^{-1}G_\delta$ which is $N_{d_\theta}\left(0, J_\delta^{-1}\Sigma_{\delta\delta}J_\delta^{-1}\right)$ distributed. By definition, the limiting distribution in (B.40) depends on the cones $C$ and $C^*$, and hence implicitly on the location of the nuisance parameters, see e.g. Cavaliere *et al.* (2019) for a general discussion. In line with Francq and Zakoïan (2009) and Pedersen (2017) we make the additional assumption that the nuisance parameters are in the interior. To do so, without loss of generality, order the parameters in $\theta$ as

$$\theta = (\theta_1', \theta_2')',$$

with $\theta_1 = (\alpha_{31}, \alpha_{32}, \alpha_{33}, \beta_{31}, \beta_{32})'$ of dimension $d_{\theta_1} = 5$, and with $\theta_2$ containing the remaining $d_{\theta_2} = 16$ (nuisance) parameters in $W, A$ and $B$.

**(vii)** Assume that $\theta_{2,0} \in \text{int}\Theta_2$ and $\Theta = \Theta_1 \times \Theta_2$, with $\theta_1 \in \Theta_1$ and $\theta_2 \in \Theta_2$.

Under the additional assumptions in (vii), $C = \mathbb{R}_+^{d_{\theta_1}} \times \mathbb{R}^{d_{\theta_2}}$ and $C^* = \{0_{d_{\theta_1}}\} \times \mathbb{R}^{d_{\theta_2}}$, which implies that

$$\sup \mathrm{LR}_T(\mathsf{H}_2^*) \xrightarrow{d} \sup_{\delta\in\Theta_{\sup}}\left\{\lambda_\delta' \left(KJ_\delta^{-1}K'\right)^{-1}\lambda_\delta\right\}, \tag{B.41}$$

where $K$ is given by $K\theta = \theta_1$ and

$$\lambda_\delta = \arg\inf_{\eta\in\mathbb{R}_+^{d_{\theta_1}}}\left\{(\eta - Z_\delta)'\left(KJ_\delta^{-1}K'\right)^{-1}(\eta - Z_\delta)\right\}. \tag{B.42}$$

and with $Z_\delta = KJ_\delta^{-1}G_\delta$ such that $Z_\delta$ is a $d_{\theta_1}$ dimensional Gaussian process.

## B.2 Implementation:

One may obtain a critical value for the supLR test by relying on the following steps, see also Andrews (2001) and Pedersen (2017). By definition, $\delta$ is $d_\delta = 3$ dimensional and we choose $k$ different values for each entry of $\delta$, such that we have a discrete grid $\Delta$ with $d_\Delta = k^{d_\delta}$ different values of $\delta$.

**Initialization** For given $\delta, \delta_1, \delta_2 \in \Delta$ estimate $J_\delta$ and $\Sigma_{\delta_1\delta_2}$ as

$$\hat{J}_\delta = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\hat\theta_{T,\delta},\delta)}{\partial\theta\partial\theta'}, \text{ and } \hat\Sigma_{\delta_1\delta_2} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial l_t(\hat\theta_{T,\delta_1},\delta_1)}{\partial\theta}\frac{\partial l_t(\hat\theta_{T,\delta_2},\delta_2)}{\partial\theta'}.$$

**Step 1** Draw a realization of $(Z_\delta : \delta \in \Delta)$ as

$$(Z_{\delta_1}, ..., Z_{\delta_{d_\Delta}}) = N_{d_{\theta_1} \times d_\Delta}(0, \begin{pmatrix} \hat{\Sigma}^Z_{\delta_1\delta_1} & \hat{\Sigma}^Z_{\delta_1\delta_2} & \dots & \hat{\Sigma}^Z_{\delta_1\delta_{d_\Delta}} \\ \hat{\Sigma}^Z_{\delta_2\delta_1} & \hat{\Sigma}^Z_{\delta_2\delta_2} & & \vdots \\ & & \ddots & \vdots \\ \hat{\Sigma}^Z_{\delta_{d_\Delta}\delta_1} & \dots & \dots & \hat{\Sigma}^Z_{\delta_{d_\Delta}\delta_{d_\Delta}} \end{pmatrix}),$$

where $\hat{\Sigma}^Z_{\delta_i\delta_j} = K\hat{J}^{-1}_{\delta_i}\hat{\Sigma}_{\delta_i\delta_j}\hat{J}^{-1}_{\delta_j}K'$ for $i, j = 1, 2, ..., d_\Delta$.

**Step 2** For $i = 1, 2, .., , d_\Delta$, compute the $d_{\theta_1}$ dimensional $\lambda_{\delta_i}$ by solving the constrained minimization problem in (B.42), with $Z_\delta$ and $J_\delta$ replaced with $Z_{\delta_i}$ and $\hat{J}_{\delta_i}$, respectively. Next, compute

$$\mu = \max_{\delta \in \Delta} \left\{ \lambda'_\delta \left( K\hat{J}^{-1}_\delta K' \right)^{-1} \lambda_\delta \right\}.$$

**Step 3** A critical value for a test with nominal size $a$ is found by repeating Steps 1 and 2 $M$ times and computing the empirical $(1-a)$-percentile of $(\mu_i)_{i=1,2,...,M}$.

# C Bootstrap Algorithm for Testing Reduced Rank

Following Cavaliere *et al.* (2017) and Cavaliere *et al.* (2019), we apply a restricted recursive bootstrap to obtain critical values for the likelihood ratio statistic, $LR_T(\mathsf{H}_2)$, where the null hypothesis of reduced rank is imposed on the bootstrap data generating process. The recursive bootstrap scheme applied is standard in the context of GARCH models, see e.g. Hidalgo and Zaffaroni (2007) or Jeong (2017). The bootstrap algorithm is as follows:

**Initialization** Estimate the model parameters with $\mathsf{H}_2$. That is, the likelihood function in (11) is maximized with $A = \gamma\alpha'$ and $B = \gamma\beta'$ where the $(3 \times 2)$ matrices $\gamma, \alpha$ and $\beta$ have non-negative entries. With $\tilde{\theta}_T$ denoting the obtained restricted estimator, for $t = 1, ..., T$ compute the centered and standardized residuals,

$$\hat{\eta}^c_t = \hat{\Sigma}^{-1/2}_\eta \left( \hat{\eta}_t - \frac{1}{T} \sum_{t=1}^T \hat{\eta}_t \right),$$

where $\hat{\Sigma}_\eta$ is the sample covariance matrix of $\hat{\eta}_t$, and

$$\hat{\eta}_t = \Lambda^{-1/2}_t(\tilde{\theta}_T)V(\tilde{\theta}_T)'X_t.$$

**Step 1** Using the estimated parameter vector under the null hypothesis, $\tilde{\theta}_T$, generate the bootstrap process $X_t^*$ as follows:

$$\begin{aligned} X_t^* &= V(\tilde{\theta}_T)\Lambda_t^{*1/2}(\tilde{\theta}_T)\eta_t^*, \quad \Lambda_t^*(\tilde{\theta}_T) = \operatorname{diag}(\lambda_t^*(\tilde{\theta}_T)) \\ \lambda_t^*(\tilde{\theta}_T) &= W(\tilde{\theta}_T) + A(\tilde{\theta}_T)(V(\tilde{\theta}_T)'X_{t-1}^*)^{\odot 2} + B(\tilde{\theta}_T)\lambda_{t-1}^*(\tilde{\theta}_T), \end{aligned}$$

for $t = 1, ..., T$. Here the bootstrap innovations, $\eta_t^*$, are drawn uniformly from $\hat{\eta}_t^c$ with replacement, and the initial values are $X_0^* = X_0$ and $\lambda_0^* = W\left(\tilde{\theta}_T\right)$.

**Step 2** With the bootstrap log-likelihood function $L_T^*(\theta)$ given by,

$$L_T^*(\theta) = \sum_{t=1}^{T} l_t^*(\theta), \quad l_t^*(\theta) = \log\det(\Omega_t^*(\theta)) + X_t^{*\prime}\Omega_t^{*-1}(\theta)X_t^*,$$

$$\Omega_t^*(\theta) = V(\theta)\Lambda_t^*(\theta)V(\theta)', \quad \Lambda_t^*(\theta) = \operatorname{diag}(\lambda_t^*(\theta)),$$

$$\lambda_t^*(\theta) = W + A(V(\theta)'X_{t-1}^*)^{\odot 2} + B\lambda_{t-1}^*(\theta),$$

this is maximized unrestricted and under the hypothesis in order to obtain the bootstrap estimators $\hat{\theta}_T^*$ and $\tilde{\theta}_T^*$. Compute next the bootstrap LR statistic,

$$LR_T^*(\mathsf{H}_2) = 2(L_T^*(\hat{\theta}_T^*) - L_T^*(\tilde{\theta}_T^*)).$$

**Step 3** A critical value for a test with nominal size $a$ is found by repeating Steps 1 and 2 $B$ times and computing the empirical $(1-a)$-percentile of $(LR_T^*(b) : b = 1, \ldots, B)$.

**Remark 5.** *Note that the bootstrap distribution approximates the $LR_T(\mathsf{H}_2)$ statistic for the case where, under $\mathsf{H}_2$, nuisance parameters are assumed to be in the interior of the parameter space. To allow nuisance parameters on the boundary of the parameter space, one may alternatively apply the shrinkage-based bootstrap proposed by Cavaliere* et al. *(2019).*

# D Monte Carlo

In this section, we investigate the finite sample properties of the QMLE discussed in Section 3.2. The asymptotic distribution theory for the QMLE is presented in Theorem 3.3 for the general model with $A$ and $B$ general $(p \times p)$ dimensional matrices. For the simulations in Cases (i)-(iii) below, we consider the case of $B$ diagonal (or even zero) as detailed in order to keep the discussion simple. The emphasis of the simulations is on the sufficient regularity condition of finite second order moments of $X_t$ in Theorem 3.3, which we conjecture is not necessary. In addition, we investigate the necessity of the rotation parameters in $\phi$ being restricted to the interval $[0, \pi/2]$, which is sufficient for identification. The simulations indeed indicate that the conditions of finite second order moments and the restrictions on $\phi$ are not necessary.

## D.1 Case (i): Asymptotic Conditions Satisfied

[Figure 1 here]

In Case (i), the bivariate $\lambda$-GARCH model is considered, where

$$X_t = V\Lambda_t^{1/2}\eta_t, \quad \eta_t \; i.i.d.N(0, I_2), \quad \Lambda_t = \operatorname{diag}(\lambda_t), \quad \lambda_t = W + AY_{t-1}^{\odot 2} + B\lambda_{t-1}, \qquad \text{(D.43)}$$

and $B$ is assumed to be diagonal[2]. For the data-generating process (dgp), set $\phi_0 = 0.70 \in [0, \pi/2]$, $W_0 = (0.50, 0.75)'$ and

$$A_0 = \begin{pmatrix} 0.10 & 0.06 \\ 0.05 & 0.01 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 0.85 & 0.00 \\ 0.00 & 0.77 \end{pmatrix},$$

such that $\rho(A_0 + B_0) = 0.98 < 1$. By Theorem 3.1 (setting $k = 1$), the stationary solution of the process has finite second order moments, and the conditions of Theorem 3.3 are satisfied.

We simulate $N = 1000$ realizations the process with $T = 10000$ observations, and estimate $\phi, \; W, \; A, \; B$ by QMLE. Figure 1 contains kernel density estimates of the centered and scaled estimates of $\phi$, $W_1$, $A_{11}$, and $B_{11}$. The solid line is the estimated density, and the dashed line is the normal density. As expected Figure 1 confirms asymptotic normality.

## D.2 Case (ii): Lack of Second Order Moments

[Figure 2 here]

Consider again the model in (D.43) with $A$ and $B$ diagonal. For the dgp $\phi_0$ is as before, $W_0 = (0.1, 0.1)'$

$$A_0 = \begin{pmatrix} 0.12 & 0.00 \\ 0.00 & 0.10 \end{pmatrix}, \quad B_0 = \begin{pmatrix} 0.88 & 0.00 \\ 0.00 & 0.84 \end{pmatrix},$$

[2] The theory in Theorem 3.3 is straightforward to modify to the case of $A$ and $B$ diagonal.

such that $\rho(A_0 + B_0) = 1$. Hence, by definition, the stationary solution does not have finite second-order moments which violates the suffient condition in Theorem 3.3. Figure 2 contains kernel density estimates of the centered and scaled estimates of $\phi$, $W_1$, $A_{11}$, and $B_{11}$. Despite the fact that the sufficient condition for asymptotic normality is violated, the estimates seem to fit a normal distribution, indicating that the requirement of finite second order moments in Theorem 3.3 is not a necessary condition.

## D.3 Case (iii): The Rotation Parameter $\phi$

[Figure 3 here]

Consider here the trivariate $\lambda$-GARCH,

$$X_t = V'\Lambda_t^{1/2}\eta_t, \quad \eta_t \ i.i.d.\ N(0, I_3), \quad \Lambda_t = \operatorname{diag}(\lambda_t), \quad \lambda_t = W + AY_{t-1}^{\odot 2} + B\lambda_{t-1},$$

with $B = 0_{3\times 3}$ and with the parameter space for $\phi = (\phi_1, \phi_2, \phi_3)'$ is extended such that $\phi_i \in [-\pi/2, \pi/2]$. For the dgp set

$$\phi_0 = \begin{pmatrix} 0.47 \\ 1.45 \\ -1.30 \end{pmatrix}, \quad W_0 = \begin{pmatrix} 0.45 \\ 1.50 \\ 0.95 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0.25 & 0.05 & 0.09 \\ 0.03 & 0.35 & 0.06 \\ 0.07 & 0.12 & 0.3 \end{pmatrix}, \quad B_0 = 0_{3\times 3},$$

such that $\phi_{0,3} \notin [0, \pi/2]$. Figure 3 contains standardized densities of $\hat{\phi}_1$, $\hat{\phi}_2$, and $\hat{\phi}_3$. Lemma 1 restricts $\phi_i$ to be in the interval $[0, \pi/2]$, as both the sine and cosine functions are monotonic in this interval. If the parameters $\phi$ are not uniquely identified, we expect to see one of two things: One, if the parameters are not uniquely identified, we would expect the estimated densities to be multi-modal or not centered around zero. Figure 3 again indicates that the condition can be relaxed.

# Tables

## Table 1: Estimation results - unrestricted model

| Rank | $W$ | $A$ | | | $B$ | | | $\phi$ | $V$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.108 | 0.127 | 0.140 | 0.010 | $9.4 \times 10^{-6}$ | 0.088 | 0.047 | 0.321 | 0.711 | $-0.240$ | 0.661 |
| | (0.318) | (0.533) | (1.008) | (0.013) | (2.942) | (1.558) | (0.212) | (1.732) | (1.337) | (1.008) | (1.085) |
| $q = 3$ | 0.094 | 0.137 | 0.110 | 0.005 | $5.8 \times 10^{-6}$ | 0.027 | 0.031 | 0.723 | $-0.236$ | 0.804 | 0.546 |
| | (0.041) | (0.174) | (0.163) | (0.035) | (2.303) | (1.488) | (0.139) | (1.525) | (0.957) | (0.633) | (0.530) |
| | 0.033 | 0.081 | 0.164 | 0.068 | 0.018 | $5.4 \times 10^{-5}$ | 0.912 | 0.813 | $-0.662$ | $-0.545$ | 0.515 |
| | (0.454) | (0.884) | (1.164) | (0.024) | (0.719) | (3.224) | (0.044) | (2.362) | (1.143) | (0.585) | (1.952) |

| | | Factor model | | Reduced rank model | |
|---|---|---|---|---|---|
| Log-likelihood | -14939.63 | | | | |
| AIC | 29927.26 | $LR$ test | 1963.16 | $LR$ test | 3.00 |
| BIC | 30071.58 | 95%-CV | 738.35 | 95%-CV | 18.56 |

The model with rank $q = 3$ is the unrestricted model. Standard errors are reported below the point estimates. We use the delta-method to obtain standard errors for the eigenvectors, $V$. The acronyms AIC and BIC are shorthand for the Aikike Information Criterium and the Bayesian Information Criterium. The LR test for the factor model is the supLR test where the critical value is approximated in a simulation. The LR test for the reduced rank model is a standard LR test, where the critical value is approximated using a restricted bootstrap.

## Table 2: Estimation results - reduced rank model

| Rank | $W$ | $A$ | | | $B$ | | | $\phi$ | $V$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.109 | 0.158 | 0.009 | 0.155 | $3.6 \times 10^{-6}$ | 0.043 | 0.115 | 0.328 | 0.715 | 0.660 | 0.233 |
| | (0.069) | (0.199) | (0.002) | (0.236) | (0.182) | (0.066) | (0.333) | (0.191) | (0.073) | (0.026) | (0.153) |
| q=2 | 0.034 | 0.080 | 0.068 | 0.155 | $3.4 \times 10^{-4}$ | 0.914 | $5.5 \times 10^{-5}$ | 0.715 | $-0.243$ | 0.547 | $-0.801$ |
| | (0.041) | (0.408) | (0.012) | (0.595) | (0.151) | (0.074) | (0.287) | (0.192) | (0.177) | (0.028) | (0.072) |
| | 0.089 | 0.108 | 0.006 | 0.106 | $2.5 \times 10^{-6}$ | 0.030 | 0.079 | $-0.752$ | $-0.656$ | 0.516 | 0.551 |
| | (0.033) | (0.052) | (0.005) | (0.078) | (0.124) | (0.013) | (0.289) | (0.151) | (0.145) | (0.007) | (0.170) |
| Log-likelihood | -14941.13 | | | | | | | | | | |
| AIC | 29922.26 | | | | | | | | | | |
| BIC | 30042.53 | | | | | | | | | | |

The reduced rank model is denoted $q = 2$. Standard errors are reported below the point estimates. We use the delta-method to obtain standard errors for $A$ and $B$ and for the eigenvectors, $V$. The acronyms AIC and BIC are shorthand for the Aikike Information Criterium and the Bayesian Information Criterium.

## Table 3: Estimated parameters - reduced rank matrices

| Rank | $\alpha'$ | | | $\beta'$ | | | $\gamma'$ | | |
|---|---|---|---|---|---|---|---|---|---|
| q=2 | 0.158 | 0.080 | 0.009 | $3.6 \times 10^{-6}$ | $3.4 \times 10^{-4}$ | 0.043 | 1 | 0 | 0.685 |
| | (0.199) | (0.408) | (0.002) | (0.182) | (0.151) | (0.066) | | | (0.538) |
| | 0.068 | 0.155 | 0.155 | 0.914 | 0.115 | $5.5 \times 10^{-5}$ | 0 | 1 | $1.4 \times 10^{-6}$ |
| | (0.012) | (0.236) | (0.595) | (0.074) | (0.333) | (0.287) | | | (0.014) |

Parameter estimated for the matrices $\alpha, \beta$ and $\gamma$. Recall that $A = \gamma\alpha'$ and $B = \gamma\beta'$ for $q < p$. Standard errors are reported below the point estimates.

# Figures

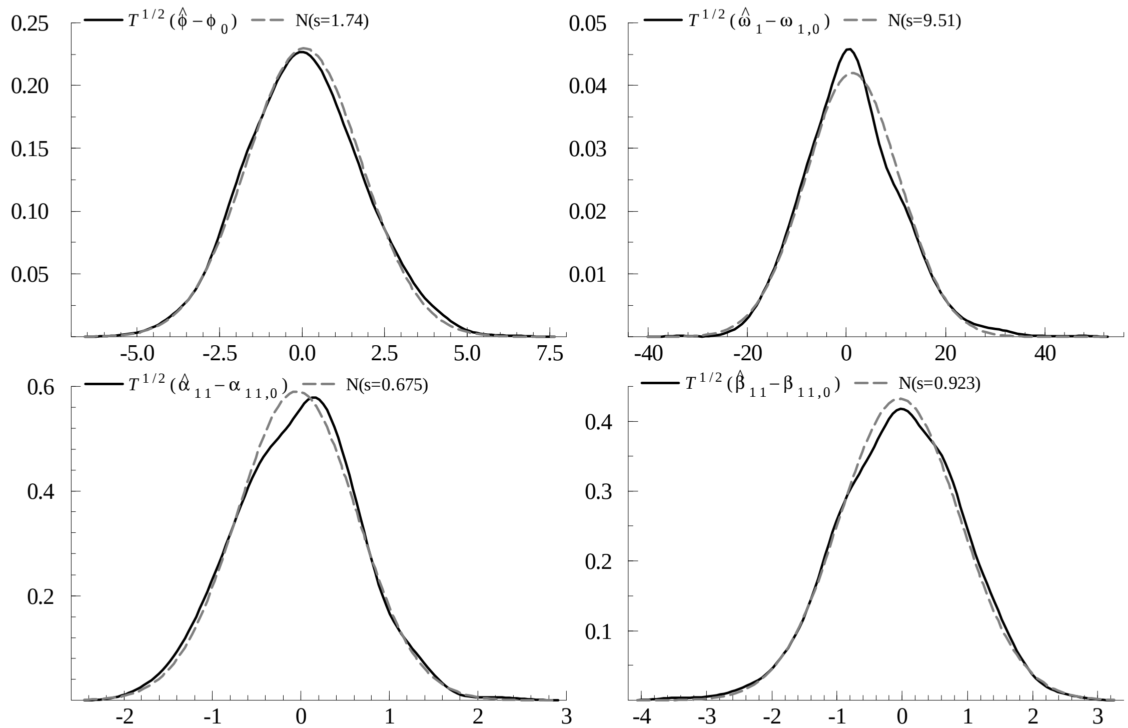


Figure 1: Densities of estimated parameters when the DGP has finite second order moments.

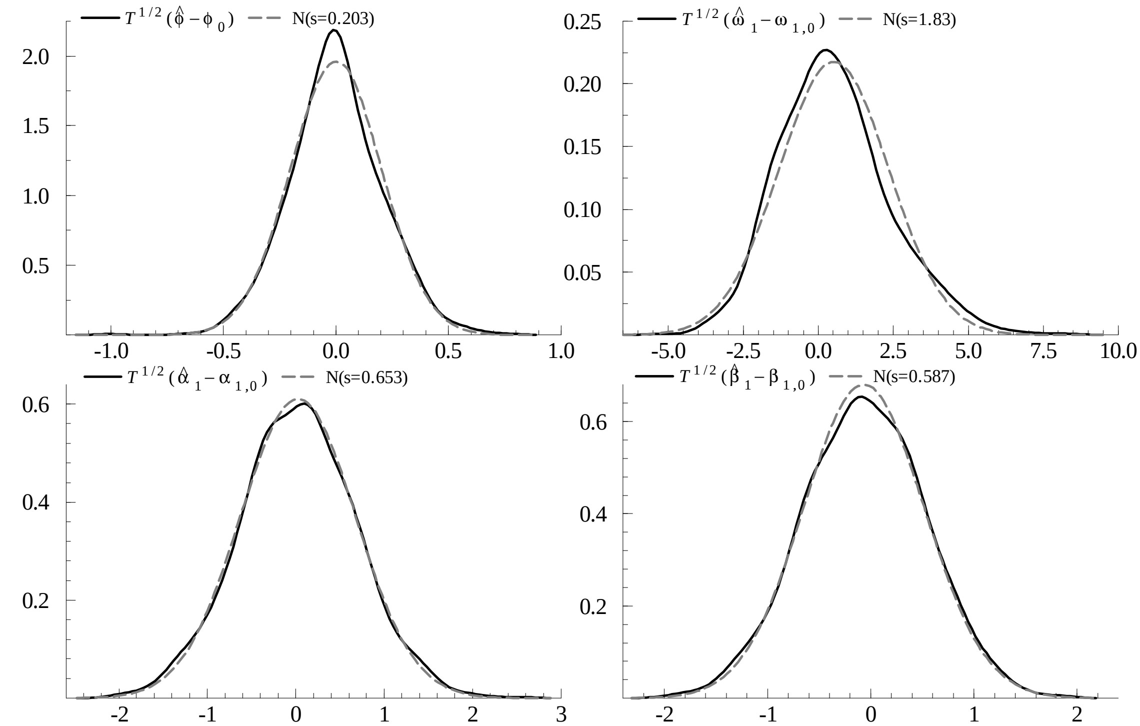


Figure 2: Densities of estimated parameters when the DGP does not have finite second order moments.

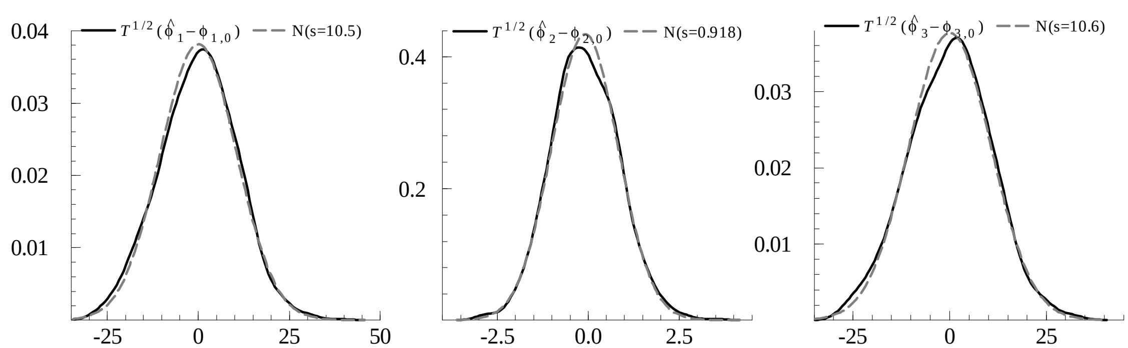


Figure 3: Densities of estimated parameters when we extend the parameter space of the rotation parameters.

Figure 4:

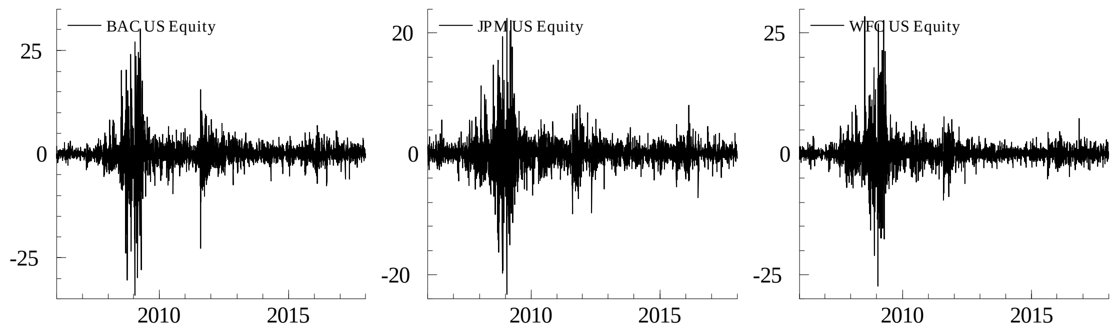


Figure 5: Log-returns

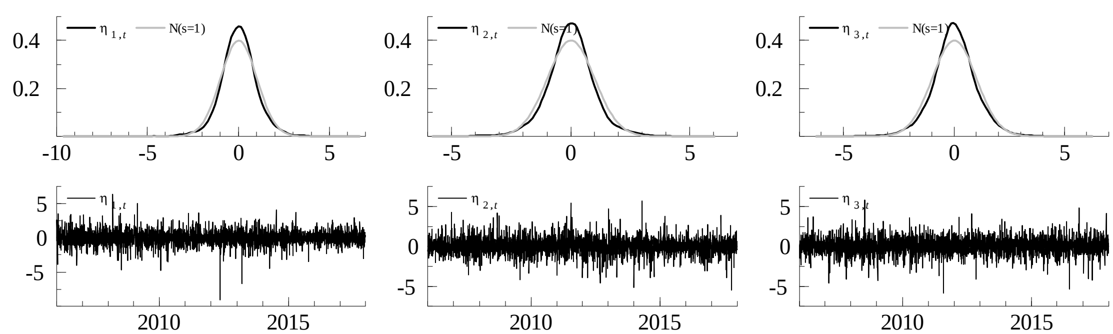


Figure 6: Estimated residuals

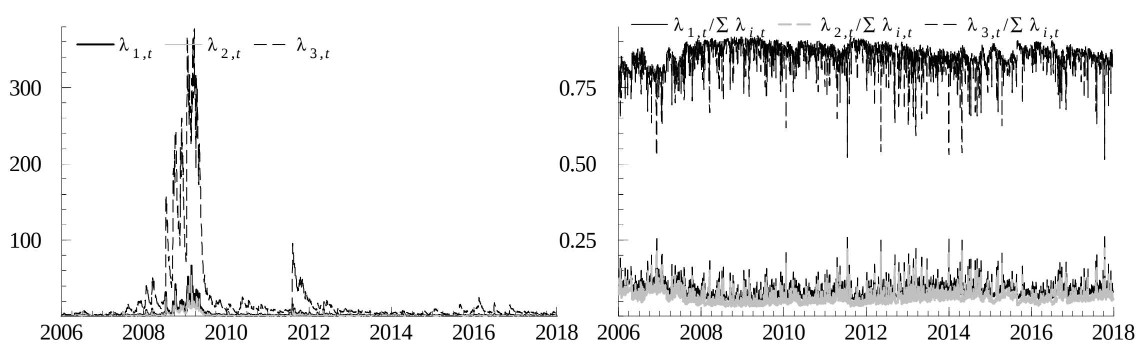


Figure 7: Estimated conditional eigenvalues

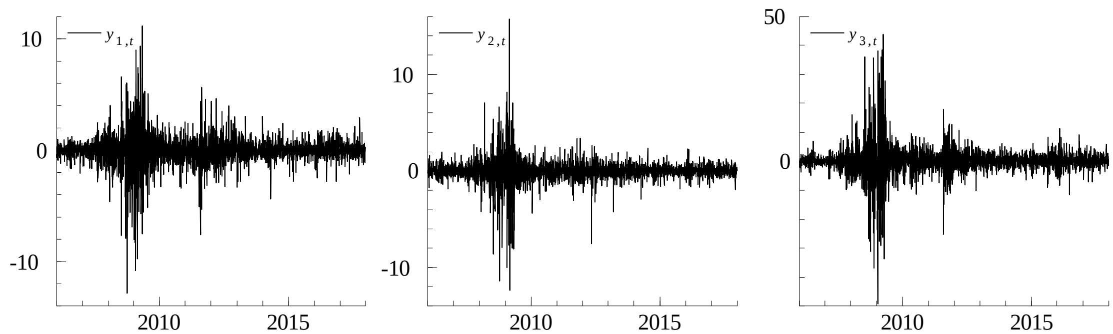


Figure 8: Estimated rotated returns