DEPARTMENT OF ECONOMICS
FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD Thesis

Simon Thinggaard Hetland

Dynamic Conditional Eigenvalues

Inference and Testing in the λ-GARCH Model

Supervisors: Anders Rahbek and Rasmus Søndergaard Pedersen
Submitted on: March 31, 2021

Contents

Acknowledgments

This dissertation was written as a part of my PhD-studies at the Department of Economics at the University of Copenhagen. I am grateful for having been given this opportunity and for the financial support over the past three years. My deepest gratitude to my doctoral advisors, Anders Rahbek and Rasmus Søndergaard Pedersen, for all their support, supervision and mentorship. Their guidance has been a constant source of inspiration, and their critical comments have improved much of the work in this thesis. I am also grateful to Heino Bohn Nielsen for valuable feedback and support since the final year of my bachelor's degree.

Writing this dissertation has been a mountain to climb, and while it has been a challenging and sometimes lonely task, I would not have been without. Luckily, I have had excellent colleagues, with whom I thoroughly have enjoyed every single coffee and lunch break. In particular, I would like to thank Philipp Kless, Anne Lundgaard Hansen, Marcus Mølbak Ingholt, Nick Fabrin Nielsen, Patrick Thöni, Sigurd Nellemann Thorsen and Stefan Voigt, for all the good times in and out of the office.

I would also like to thank Paolo Zaffaroni for hosting me during my research visit at Imperial College Business School in the fall of 2019. I enjoyed being a visitor at the Department of Finance, and the faculty, staff and graduate students were welcoming during my stay. My visit was financially supported by Etly og Jørgen Stjerngrens Fond, Augustinus Fonden, Rudolph Als Fondet, Knud Højgaards Fond, and William Demant Fonden, for which I am deeply appreciative.

I am immensely grateful to my family and friends for their support. To my brother, Andreas, thank you for not only introducing me to the wonderful world of time series econometrics and GARCH models, but also inspiring me to do a PhD. I look forward to keeping on pestering our family with our many more discussions on econometrics in the coming years. To my parents, thank you for supporting me from day one, and for your everlasting encouragement. Finally, to my wonderful wife Helena, thank you for listening to my rambling monologues about GARCH models, for always believing in me, and for your endless love and support.

Summary

In this thesis, we study a class of multivariate generalized autoregressive heteroskedasticity (GARCH) models, denoted the Dynamic Conditional Eigenvalue GARCH (or λ-GARCH) model. Multivariate GARCH models are useful for estimating and filtering time varying (co-)variances, which are used e.g. in empirical asset pricing, Markovitz-type portfolio optimization and value-at-risk estimation. GARCH models have long been a staple in empirical finance and financial econometrics. This thesis contains three self-contained chapters on the λ-GARCH, covering large-sample properties and bootstrap-based inference.

In the first chapter, *"The Dynamic Conditional Eigenvalue GARCH Model"*, we introduce the λ-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 provide new results on asymptotic theory for the Gaussian quasi-maximum likelihood estimator, 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.

The second chapter, *"Spectral Targeting Estimation of Dynamic Conditional Eigenvalue GARCH Models"* investigates a two-step estimator of the λ-GARCH model, combining (eigenvalue and -vector) targeting estimation with stepwise (univariate) estimation. This estimator is denoted the "spectral targeting estimator". This type of estimator has long been used in empirical modeling, and in this chapter, we present novel asymptotic theory. We find that the estimator is consistent under finite second order moments, while asymptotic normality holds under finite fourth order moments. The estimator is especially well suited for modeling larger portfolios: we compare the empirical performance of the spectral targeting estimator to that of the quasi-maximum likelihood estimator for five portfolios of 25 assets. The spectral targeting estimator dominates in terms of computational complexity, being up to 57 times faster in estimation, while both estimators produce similar out-of-sample forecasts, indicating that the spectral targeting estimator is well suited for high-dimensional empirical applications.

In the third and final chapter, *"Bootstrap-Based Inference and Testing in Multivari-*

ate Dynamic Conditional Eigenvalue GARCH Models", we study fixed-design bootstrap for quasi-maximum likelihood estimation of multivariate GARCH processes. Specifically, we extend the univariate bootstrap of Cavaliere, Pedersen, and Rahbek (2018) to the λ-GARCH model. We show, under fairly mild conditions, that the bootstrap Wald test statistic is consistent, conditional on the original data. The theoretically investigated fixed-design bootstrap is contrasted to a recursive bootstrap, and the asymptotic test statistic. Through Monte Carlo simulations, we find evidence that the fixed-design bootstrap is superior to the recursive bootstrap and the asymptotic test in small samples. In larger samples, both bootstrap designs and the asymptotic test share properties, as expected from the asymptotic theory. An empirical application illustrates the empirical merits of the bootstrap in multivariate GARCH models. The appealing theoretical properties, along with the excellent finite sample properties, suggest that the fixed-design bootstrap is an important tool for small sample inference in multivariate GARCH models.

Dansk Résume

Denne afhandling undersøger en klasse af multivariate generaliserede autoregressive betingede heteroskedastiske (GARCH) modeller, kendt som *Dynamic Conditional Eigenvalue GARCH* (eller λ-GARCH) modeller. Multivariate GARCH modeller er særdeles anvendelige til at estimere og filtrere tidsvarierende (ko-)varianser, som anvendes bl.a. til prisfastsætning af aktiver, Markovitz porteføljeallokering og value-at-risk estimering. GARCH modeller har længe været en fast bestanddel i finansiering og finansiel økonometri. Denne afhandling indeholder tre kapitler om λ-GARCH modellen, og vi undersøger de asymptotiske egenskaber af to forskellige estimatorer og bootstrap-baseret inferens.

Det første kapitel, *"The Dynamic Conditional Eigenvalue GARCH Model"*, introducerer λ-GARCH modellen, hvori egenværdierne af den betingede kovariansmatrix er tidsvarierende. Dynamikken for egenværdierne er udledt via teorien for dynamiske betingede score modeller, oprindeligt introduceret af Creal, Koopman, and Lucas (2013) og Harvey (2013). Vi udleder ny asymptotisk teori for den Gaussiske quasi-maximum likelihood estimator, og introducerer en ny test for reduceret rank af (G)ARCH parametermatricerne af de tidsvarierende egenværdier. Den udledte teori anvendes på amerikansk data, hvor vi finder at strukturen af egenværdierne kan reduceres på samme måde som man tester for faktorer i volatilitetsmodeller.

Det andet kapitel, *"Spectral Targeting Estimation of Dynamic Conditional Eigenvalue GARCH Models"*, undersøger en to-trins estimator af λ-GARCH modellen, hvori (egenværdi og -vektor) *targeting* estimation kombineres med trinvis (univariat) estimation. Vi kalder denne estimator "spectral targeting estimator". Denne type estimator er længe blevet brugt i empiriske sammenhæng, og i kapitlet udleder vi ny asymptotisk teori. Konkret finder vi at estimatoren er konsistent når den ubetingede kovariansmatrix er veldefineret, og at estimatoren er asymptotisk normalfordelt når fjerde moment er endeligt. Denne estimator er især velegnet til modellering af større finansielle porteføljer: Vi sammenligner den empiriske præstation af spectral targeting estimatoren med quasi-maximum likelihood estimatoren for fem forskellige porteføljer bestående af 25 aktiver. Vores resultater indikerer, at spectral targeting estimatoren dominerer med hensyn til beregningskompleksitet, hvor den er op til 57 gange hurtigere, og at begge estimatorer producerer sammenlignbare out-of-sample prognoser. Dette indikerer at spectral targeting estimatoren er særdeles velegnet til høj dimensionale anvendelser.

I det tredje og sidste kapitel, *"Bootstrap-Based Inference and Testing in Multivariate Dynamic Conditional Eigenvalue GARCH Models"*, undersøger vi en *fixed-design* bootstrap til quasi-maximum likelihood estimation af multivariate GARCH processer. Vi udvider den univariate bootstrap fra Cavaliere, Pedersen, and Rahbek (2018) til λ-GARCH modellen. Vi viser, under milde betingelser, at bootstrap Wald test statistikken er konsistent, betinget på det observerede data. Vi sammenligner fixed-design bootstrappen med en rekursiv bootstrap og den asymptotiske test statistik. Vores Monte Carlo simulationer viser, at fixed-design bootstrappen er overlegen i små datasæt. I større datasæt deler bootstap algoritmerne og den asymptotiske test egenskaber, som forventet fra den asymptotiske teori. I en lille empirisk anvendelse illustrerer vi de empiriske meritter af vores fixed-design bootstrap i multivariate GARCH modeller. De tiltalende teoretiske egenskaber, sammen med den gode præstation i små datasæt, indikerer at fixed-design bootstrappen er et vigtigt redskab til inferens i multivariate GARCH modeller i små datasæt.

Chapter 1

Dynamic Conditional Eigenvalue GARCH

This chapter is joint research with Anders Rahbek (University of Copenhagen) and Rasmus Søndergaard Pedersen (University of Copenhagen).[1]

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 λ-GARCH model. We provide new results on asymptotic theory for the Gaussian quasi-maximum likelihood estimator (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.

[1]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 λ-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 λ-GARCH model is an extended version of the generalized orthogonal GARCH (GO-GARCH) model of van der Weide (2002). Here the λ-GARCH model extend the GO-GARCH model as the spill-over effects allow for more 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_v-distributed case, the dynamics of the λ-GARCH model generalizes and extends the univariate β-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 λ-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 time-invariance of all the conditional 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 λ-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 the sufficient condition of finite second-order moments may not be necessary, which is 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, novel results on identification are derived, as needed in particular for establishing consistency of the QMLE.

Moreover, testing reduced rank in the context of multivariate GARCH models is non-standard 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 (2020).

Existing theory for the classic (non-extended) multivariate GO-GARCH model typically relies on two (or, three) step estimators. For 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, see also Noureldin, Shephard and Sheppard (2014). Lanne and Saikkonen (2007) consider one-step estimation of a factor GO-GARCH model, where, using a BEKK-type representation, it is argued that asymptotic theory in Comte and Lieberman (2003) applies. This in turn relies on the assumption of finite eighth-order moments of X_t, see also Hafner and Preminger (2009a), Avarucci, Beutner and Zaffaroni (2013), and Pedersen and Rahbek (2014) for discussion of moment requirements. In contrast, we show that the QMLE for the λ-GARCH model is asymptotically normal under the mild sufficient condition of finite second-order moments.

The paper is structured as follows. Section 2 defines the λ-GARCH model for the case of conditional Gaussian and conditional Student's t distributed returns. In Section 3, the stochastic properties of the λ-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. Let $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$. 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, $\text{diag}\,(x) = \text{diag}\,((x_i)_{i=1}^p)$ 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 $\overset{p}{\to}$, $\overset{d}{\to}$ and $\overset{w}{\to}$ 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 | λ-GARCH

We consider a class of multivariate conditionally heteroskedastic models with time-varying eigenvalues of the conditional covariance matrix. As detailed below, dynamic specifications of the time-varying eigenvalues are derived under different distributional assumptions on the innovations, using the score-based approach in Creal, Koopman and Lucas (2011) and Harvey (2013).

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 σ-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 $\mathcal{F}_{t-1}$. We assume for simplicity that the conditional mean $E\,(X_t|\mathcal{F}_{t-1})$ is zero, $E\,(X_t|\mathcal{F}_{t-1}) = 0$, and hence that the conditional distribution of X_t can be characterized in terms of the time-varying conditional covariance matrix $\Omega_t = E(X_t X_t'|\mathcal{F}_{t-1})$ in addition to distributional shape parameters.

The conditional covariance matrix Ω_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,

$$\Omega_t = V\Lambda_t V', \tag{2.1}$$

with $V = (v_1, ..., v_p)$ and $\Lambda_t = \text{diag}\left((\lambda_{i,t})_{i=1}^p\right)$. By definition the eigenvectors $(v_i)_{i=1}^p$ are orthogonal, such that V is an orthogonal matrix, $V'V = VV' = I_p$. The eigenvalues $\lambda_{i,t}$ are strictly positive, $\lambda_{i,t} > 0$ (almost surely) for $i = 1, ..., p$ and for all t. With

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

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

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

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

$$\lambda_t = W + \mathcal{A}s_{t-1} + \mathcal{B}\lambda_{t-1}, \tag{2.2}$$

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 λ_t, up to an appropriate scaling. That is, the score contribution in the dynamics is given by,

$$s_t = S_t \frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t}, \tag{2.3}$$

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.,

$$S_t = \left(E\left[\frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t} \frac{\partial \log f(X_t|\lambda_t)}{\partial \lambda_t'} \middle| \mathcal{F}_{t-1} \right] \right)^{-1}. \tag{2.4}$$

Below we consider the implied λ-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.

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,

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

with Ω_t defined in (2.1). Using the definitions in (2.2)–(2.4) yields that the implied dynamics of λ_t can be represented in the multivariate GARCH-type form as given by,

$$\lambda_t = W + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1}. \tag{2.5}$$

Here 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 $\text{Cov}\left(v_j' X_{t-1}, v_k' X_{t-1} | \mathcal{F}_{t-1}\right) = 0$ (and hence $\text{Cov}\left(v_j' X_{t-1}, v_k' X_{t-1}\right) = 0$ as $\Omega_t = E\left(X_t X_t' | \mathcal{F}_{t-1}\right)$ by assumption) for all $j \neq k$. In addition, our proposed λ-GARCH model allows $\lambda_{i,t}$ to depend on all entries of λ_{t-1}, and hence the Gaussian score-driven eigenvalue model is a generalization of the GO-GARCH models considered by Fan *et al.* (2008) and Boswijk and van der Weide (2011). 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 $\Lambda_t = \text{diag}\left((\lambda_{i,t})_{i=1}^p\right)$ satisfying[2], with B_d a $(p \times p)$ diagonal matrix,

$$\lambda_t = (I - A - B_d) + A \left(V' X_{t-1}\right)^{\odot 2} + B_d \lambda_{t-1}. \tag{2.6}$$

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 λ-GARCH model, for which the zero row 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 Λ_t defined by (2.6) such that the condition that $E\left(X_t X_t'\right) = I_p$ (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. The conditional density is given by

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

[2]Boswijk and van der Weide (2011) state that λ_{it} is "assumed to follow a GARCH-type structure" [p.119], and the parametrization in (2.6) is the one considered in the empirical application therein.

where $\Gamma(\cdot)$ is the Gamma function, and it follows that

$$\partial \log f(X_t|\lambda_t)/\partial \lambda_t' = \Psi' \left(\Lambda_t^{-1}\right)^{\otimes 2} \left(\delta_t \left(V'X_t\right)^{\otimes 2} - \operatorname{vec}(\Lambda_t)\right), \tag{2.7}$$

where $\Psi = \partial \operatorname{vec}(\Lambda_t)/\partial \lambda_t'$ and $\delta_t = (v+p)/\left(v-2+\sum_{i=1}^{p}\left(V_i'X_t\right)^2/\lambda_{it}\right)$, see also Creal et al. (2011, Theorem 1). Stated differently, $\partial \log f(X_t|\lambda_t)/\partial \lambda_{it} = \frac{1}{2\lambda_{it}^2}\left[\delta_t\left(V_i'X_t\right)^2 - \lambda_{it}\right]$ for $i = 1,...,p$. Moreover, S_t in (2.4) is given by,

$$S_t^{-1} = \frac{1}{4}\Psi' \left(\Lambda_t^{-1/2}\right)^{\otimes 2} \left[gG - \operatorname{vec}(I)\operatorname{vec}(I)'\right] \left(\Lambda_t^{-1/2}\right)^{\otimes 2} \Psi. \tag{2.8}$$

Using (2.7) and (2.8) it follows that the Student's t λ-GARCH conditional eigenvalue dynamics can be stated as

$$\lambda_{t+1} = W + A_t \left(V'X_t\right)^{\odot 2} + B\lambda_t. \tag{2.9}$$

It differs from the Gaussian case in (2.5) as the "ARCH-loadings" A_t are time varying in the Student's t case, while the vector W and the matrix B are constant as in the Gaussian case. Specifically, and as applied in the empirical Section 5, the time varying "ARCH-loadings" A_t and the constant "GARCH-loadings" B in (2.9) are for the $p = 3$ dimensional case given by,

$$A_t = w_t \mathcal{A} \begin{pmatrix} v+1 & \frac{\lambda_{1t}}{\lambda_{2t}} & \frac{\lambda_{1t}}{\lambda_{3t}} \\ \frac{\lambda_{2t}}{\lambda_{1t}} & v+1 & \frac{\lambda_{2t}}{\lambda_{3t}} \\ \frac{\lambda_{3t}}{\lambda_{1t}} & \frac{\lambda_{3t}}{\lambda_{2t}} & v+1 \end{pmatrix} \quad \text{and} \quad B = \mathcal{B} - \frac{v+5}{v}\mathcal{A}, \tag{2.10}$$

where $w_t = \left(\frac{v+5}{v(v+3)}\right)\delta_t$ for $p = 3$, i.e., $w_t = (1+5/v)/\left(v-2+\sum_{i=1}^{3}\left(V_i'X_t\right)^2/\lambda_{it}\right)$. As expected for $p = 1$, and setting $w_t = \frac{v+3}{v(v+1)}\delta_t$ for $p = 1$, we obtain the univariate Beta-t-GARCH in Harvey (2013, Ch.4.7). Hence, and as observed in Creal et al. (2011, p.555), the "*weight w_t in (8) (here: (2.10)) automatically accounts for extreme values because it decreases if $y_t'\Sigma_t^{-1}y_t$ (here: $\sum_{i=1}^{3}\left(V_i'X_t\right)^2/\lambda_{it}$) is large*", see also the discussion in Section 5.

3 Properties and Estimation of the λ-GARCH Model

For the estimation theory 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 QMLE for all parameters.

The (Gaussian) λ-GARCH model may be summarized as,

$$X_t = V\Lambda_t^{1/2}\eta_t, \quad \Lambda_t = \text{diag}\left((\lambda_{i,t})_{i=1}^p\right), \quad V'V = VV' = I_p, \tag{3.1}$$

$$\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{p,t})' = W + A(V'X_{t-1})^{\odot 2} + B\lambda_{t-1}, \tag{3.2}$$

with η_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 $B = (\beta_{ij})_{i,j=1,..,p}$ with non-negative entries, $\alpha_{ij},\beta_{ij} \geq 0$.

As in van der Weide (2002), the orthogonal matrix $V = (v_1, ..., v_p)$ is defined in terms of rotation matrices $R(i, j; \phi)$ specified below and thus parametrized by $p(p-1)/2$ rotation angles ϕ_{ij} collected in $\phi = (\phi_{12}, \ \ldots, \ \phi_{(p-1)p})' \in \mathbb{R}^{p(p-1)/2}$, with $j > i$ and $i = 1, ..., p-1$, $j = 2, ..., p$. Specifically, the $(p \times p)$ dimensional $V = V(\phi)$ is given by,

$$V(\phi) = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} R(i, j; \phi),$$

with $R(i, j; \phi)$ $(p \times p)$-dimensional rotation matrices as defined by,

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

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

Remark 1. *Note that by definition* $\det(V) = 1$*, that is,* V *is a rotation matrix. By definition this excludes orthogonal matrices with* $\det(V) = -1$*, that is, the class of reflection matrices. For further details on this, and on the parametrization of* $\det(V) = 1$ *in terms of the rotation angles, or the so-called Givens parametrization, see Pinheiro and Bates (1996, Section 2.5) and the references therein.*

Remark 2. *It follows that for the case of* $p = 3$*,* $V(\phi)$ *is given by*

$$\begin{aligned} V(\phi) &= R(1, 2; \phi) R(1, 3; \phi) R(2, 3; \phi) \\ &= \begin{pmatrix} \cos(\phi_{12}) & \sin(\phi_{12}) & 0 \\ -\sin(\phi_{12}) & \cos(\phi_{12}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\phi_{13}) & 0 & \sin(\phi_{13}) \\ 0 & 1 & 0 \\ -\sin(\phi_{13}) & 0 & \cos(\phi_{13}) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi_{23}) & \sin(\phi_{23}) \\ 0 & -\sin(\phi_{23}) & \cos(\phi_{23}) \end{pmatrix}, \end{aligned}$$

while for $p = 2$*,*

$$V(\phi) = R(1, 2; \phi) = \begin{pmatrix} \cos(\phi_{12}) & \sin(\phi_{12}) \\ -\sin(\phi_{12}) & \cos(\phi_{12}) \end{pmatrix}.$$

3.1 Stochastic properties

For the stochastic properties of X_t satisfying equations (3.1)-(3.2), we note that $V'X_t$ satisfies the stochastic recursion,

$$V'X_t = \Lambda_t^{1/2}\eta_t, \text{ with } \lambda_t = W + A\left(V'X_{t-1}\right)^{\odot 2} + B\lambda_{t-1}, \tag{3.3}$$

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 λ_t in (3.3) as the stochastic recurrence equation (SRE),

$$\lambda_t = W + \Phi_{t-1}\lambda_{t-1} \tag{3.4}$$

where Φ_t are i.i.d. random matrices,

$$\Phi_t = A \text{ diag}\left((\eta_{i,t}^2)_{i=1}^p\right) + B, \tag{3.5}$$

with Φ_t and λ_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 $(X_t : t \in \mathbb{Z})$ given by (3.1)-(3.2) is strictly stationary and ergodic if and only if $\xi < 0$, where ξ is the top Lyapunov coefficient of $(\Phi_t : t \in \mathbb{Z})$ defined by*

$$\xi = \lim_{n\to\infty} n^{-1} E(\log || \prod_{t=1}^{n} \Phi_t ||), \tag{3.6}$$

with Φ_t defined in (3.5). 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 $\{\rho(E(\Phi_t^{\otimes k})) < 1$ and $E||\eta_t||^{2k} < \infty\}$.

Remark 3. *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*

$$E\left(X_t X_t'\right) = E\left(\Omega_t\right) = E(V\Lambda_t V') = V\left(\text{diag}\left\{(I_p - A - B)^{-1} W\right\}\right)V'.$$

3.2 Quasi-Maximum Likelihood Estimation

The parameters of the λ-GARCH model in (3.1)-(3.2) are given by,

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

with the vector $\theta \in \Theta$, where the parameter space Θ is defined by,

$$\Theta = \Theta_W \times \Theta_A \times \Theta_B \times \Theta_\phi, \tag{3.7}$$

with $\Theta_W \subset \mathbb{R}^p_{++}$, $\Theta_A \subset \mathbb{R}^{p^2}_+$, $\Theta_B \subset \mathbb{R}^{p^2}_+$, and $\Theta_\phi \subset \mathbb{R}^{p(p-1)/2}$.

Given a realization $(X_t : t = 0, 1, \ldots, T)$ of the λ-GARCH process in (3.1)-(3.2), the Gaussian quasi-maximum likelihood estimator (QMLE), $\hat{\theta}_T$, for θ is defined as

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

where the log-Gaussian likelihood function is given by,

$$L_T(\theta) = \sum_{t=1}^{T} l_t(\theta), \quad l_t(\theta) = \log\det(\Omega_t(\theta)) + X_t'\Omega_t^{-1}(\theta)X_t, \tag{3.8}$$

$$\Omega_t(\theta) = V(\phi)\Lambda_t(\theta)V(\phi)', \quad \Lambda_t(\theta) = \mathrm{diag}(\lambda_t(\theta)), \tag{3.9}$$

$$\lambda_t(\theta) = W(\theta) + A(\theta)\left(V(\phi)'X_{t-1}\right)^{\odot 2} + B(\theta)\,\lambda_{t-1}(\theta), \quad t = 1, \ldots, T, \tag{3.10}$$

with $\lambda_0(\theta) = \bar{\lambda}_0$ fixed and with strictly positive entries.

In order to investigate the stochastic properties of the QMLE we make the following simple assumptions about the parameter space Θ in (3.7):

Assumption 3.1. *Assume that the true value $\theta_0 = (W_0', \mathrm{vec}(A_0)', \mathrm{vec}(B_0)', \phi_0')' \in \Theta$. Moreover, assume that $\Theta_W = [\omega_L, \omega_U]^p$ for some $0 < \omega_L < \omega_U < \infty$, $\Theta_A = [0, \alpha_U]^{p^2}$ for some $0 < \alpha_U < \infty$, $\Theta_\phi = [\phi_L, \phi_U]^{p(p-1)/2}$ with $-\infty < \phi_L < \phi_U < \infty$, and, finally, that Θ_B is compact with $\Theta_B \subset \mathbb{R}^{p^2}_+$ satisfying $\sup_{\mathrm{vec}(B) \in \Theta_B} \rho(B) < 1$.*

Note that Assumption 3.1 implies in particular that Θ is compact. To establish consistency, we make the following assumption about the data generating process $(X_t : t \in \mathbb{Z})$:

Assumption 3.2. *Assume that $\xi < 0$, with ξ defined in (3.6), 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 state an identifying high-level assumption in terms of the ergodic version $\Omega_t^\star(\theta)$ of $\Omega_t(\theta)$, as defined in (A.2) in the appendix. Note that $\Omega_t^\star(\theta)$ is well-defined for any $\theta \in \Theta$ by Assumptions 3.1-3.2.

Assumption 3.3. *For $\theta \in \Theta$, if $\Omega_t^\star(\theta) = \Omega_t^\star(\theta_0)$ almost surely, then $\theta = \theta_0$.*

We have the following strong consistency result for the QMLE.

Theorem 3.2 (Consistency)**.** *Under Assumptions 3.1-3.3, $\hat{\theta}_T \to \theta_0$ almost surely.*

The identifying condition in Assumption 3.3 is high-level, and next we state in Assumptions 3.4–3.6 sufficient and primitive conditions under which it holds.

Assumption 3.4. *Assume for the i.i.d.*$(0, I_p)$ *sequence* $(\eta_t : t \in \mathbb{Z})$ *that* η_{it} *and* η_{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 3.5. *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 3.6. *With* Θ_ϕ *defined in Assumption 3.1, assume that the true value* ϕ_0 *of* ϕ *belongs to the interior of* Θ_ϕ, *i.e.,* $\phi_0 \in \text{int}\Theta_\phi$, *and assume that* $\phi_L = 0$ *and* $\phi_U = \pi/2$.

Remark 4. *Assumptions 3.4-3.5 are classic for standard multivariate GARCH models, see e.g., Francq and Zakoïan (2012). In contrast, Assumption 3.6, which addresses the rotation parameters* ϕ, *is non-standard. As demonstrated in Lemma 3.1 this condition is indeed sufficient for identification. However, the condition may not be necessary as also investigated in Appendix D, where we consider practical implications of extending the parameter space* Θ_ϕ *. Note also in this respect, that the results in the empirical illustration in Section 5 do not change by extending the parameter space by setting* $\phi_L = -\pi/2 < 0$.

Remark 5. *Note that for the static model (where, say,* $A = B = 0$*) identification of* W *and the rotation parameters in* ϕ *can be obtained by ordering the (assumed) distinct eigenvalues* $(\omega_1, \ldots, \omega_p)$, *and moreover restricting the rotation angles* ϕ_{ij} *to lie in the interval* $[\phi_L, \phi_U] = [0, \pi]$, *see e.g., Pinheiro and Bates (1996, Section 2.5).*

Remark 6. *In the proof of Lemma 3.1 properties of the matrix* $V(\phi)'V(\phi_0)$ *for* $\phi \neq \phi_0$ *are exploited. The proof applies an auxiliary Lemma A.1, which relies on the assumption that* ϕ_0 *lies in the interior of* $\Theta_\phi = [\phi_L, \phi_U]^{p(p-1)/2}$, *ruling out permutations in* $\lambda_t(\theta)$. *To illustrate, consider the simple case of dimension* $p = 2$ *and with* $\phi_0 = 0$, $\phi = \pi/2 \notin \text{int}(\Theta_\phi)$, *such that* $\phi \neq \phi_0$. *In this case, with* $(W, A, B) = K(W_0, A_0, B_0)$ *where*

$$K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

it holds that $\Omega_t^\star(\theta) = \Omega_t^\star(\theta_0)$ *almost surely, despite* $\theta \neq \theta_0$. *Note also that* $\phi_0 \in \text{int}\,\Theta_\phi$ *rules out the case where* $V_{\phi_0} = I_p$, *that is, when the entries of* X_t *are conditionally uncorrelated. In order to allow for this, an alternative to Assumption 3.6 is to assume* $\phi_0 \in \Theta_\phi = [\phi_L, \phi_U - \delta]^{p(p-1)/2}$ *for some small* δ *satisfying* $\phi_L = 0 < \delta < \pi/2 = \phi_U$.

We have the following result for identification:

Lemma 3.1. *Under Assumptions 3.1, 3.2, 3.4-3.6, it holds that Assumption 3.3 is satisfied.*

The main argument used in the proof of the lemma is to show that if $\Omega_t^\star(\theta) = \Omega_t^\star(\theta_0)$ almost surely and $\phi \neq \phi_0$, then at least one of the conditional eigenvalues of $\Omega_t^\star(\theta_0)$ is a linear combination of the remaining conditional eigenvalues. Such a property implies that the matrix $[A_0, B_0]$ has reduced rank, and hence contradicts Assumption 3.5. Hence, we show that under Assumption 3.5, it must hold that $\phi = \phi_0$, and identification of the remaining parameters follows by well-known arguments.

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

Assumption 3.7. *The true value of the parameter vector* $\theta_0 \in \operatorname{int}\Theta$*.*

Assumption 3.8. *The data-generating process satisfies (i) that* $E\|\eta_t\|^4 < \infty$*, and (ii)* $E\|X_t\|^{2+s} < \infty$ *for some* $s > 0$*.*

Assumption 3.9. *The matrix* A_0 *has a row with a unique entry.*

Assumptions 3.7 and 3.8*(i)* are standard. Assumption 3.8*(ii)* 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 θ_0 in the proof of Lemma A.6 in Appendix A.5.

Specifically, the third-order derivatives contain terms essentially of the form

$$\frac{\dot{\lambda}_{s,t,i}(\theta)\dot{\lambda}_{s,t,j}(\theta)\dot{\lambda}_{s,t,k}(\theta)}{\lambda_{s,t}^3(\theta)} \times \frac{\lambda_{g,t}^{1/2}(\theta_0)\lambda_{h,t}^{1/2}(\theta_0)\eta_{h,t}\eta_{g,t}}{\lambda_{s,t}(\theta)}, \tag{3.11}$$

where $\eta_{s,t}$ denotes the sth entry of the noise η_t, $\lambda_{s,t}(\theta)$ is the sth entry of $\lambda_t(\theta)$ in (3.10), 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 (3.11) has finite expectation if $E\|X_t\|^{2+s} < \infty$ for some $s > 0$. 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 $E\|X_t\|^s < \infty$ for some $s > 0$ 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 3.9 is used in the proof of Lemma A.5 in order to show that the expectation of the Hessian, i.e., the information matrix $J = \operatorname{plim}(T^{-1}\partial^2 L_T(\theta_0)/\partial\theta\partial\theta')$ as defined in

(A.6), is invertible. A sufficient condition for this to hold is the assumption that A_0 has full rank, see also Assumption 3.5. Note that often the proof of invertibility of J relies on showing that there exists no non-zero $\gamma \in \mathbb{R}^{d_\theta}$ such that for all t

$$\gamma' \frac{\partial \lambda_t(\theta_0)}{\partial \theta} = 0_{p\times 1} \quad \text{almost surely.} \tag{3.12}$$

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 (3.12), $\gamma' \partial \lambda_t(\theta_0)/\partial \theta$ is linear in $(V_0' X_{t-1})^{\odot 2}$ and $\lambda_{t-1}(\theta_0)$ and that θ_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 ϕ. This leads to additional considerations about invertibility of J, and we make the additional Assumption 3.9; see the proof of Lemma A.5 for details.

We have the following result:

Theorem 3.3 (Asymptotic normality)**.** *Under Assumptions 3.1, 3.2, 3.4-3.9,*

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

where J is an invertible matrix defined in (A.6) and Σ is a non-negative definite matrix defined in (A.5) 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. Lastly, the simulation indicate that while sufficient for identification, it may not be necessary to restrict $\phi_{ij} \in [\phi_L, \phi_U]$ with $\phi_L = 0 < \phi_U = \pi/2$.

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

4 Reduced Rank of A and B

Consider the λ-GARCH model in (3.1)-(3.2) 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 λ_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 λ_t may be parametrized as the hypothesis H_q of reduced rank $q < p$ of A and B, as given by

$$H_q : A = \gamma\alpha' \quad \text{and} \quad B = \gamma\beta'. \tag{4.1}$$

Here γ, α and β 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 γ_c is $(p \times p-q)$ dimensional and $\gamma_c'\gamma = 0$ with rank of (γ, γ_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 cointegrated vector autoregressive processes, see e.g., Cavaliere, Rahbek and Taylor (2012) – this raises the issue of non-identified parameters under 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 λ-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 γ in (4.1) known such that testing H_2 reduces to testing for a zero row in A and B. Next, in Section 4.2, we discuss general testing of H_q, that is, extend the discussion to include an unknown γ matrix (and general dimension p). In the empirical illustration in Section 5 we consider implementation of both cases.

4.1 Testing with γ known

Consider the case of a $p = 3$ dimensional system with γ known, and given by

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

that is, with a zero row in γ. This is a special case of H_2, as with the (3×2) matrices α and β given by

$$\alpha = \begin{pmatrix} \alpha_{21} & \alpha_{31} \\ \alpha_{22} & \alpha_{32} \\ \alpha_{23} & \alpha_{33} \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_{21} & \beta_{31} \\ \beta_{22} & \beta_{32} \\ \beta_{23} & \beta_{33} \end{pmatrix},$$

one can write A and B as

$$A = \gamma\alpha' = \begin{pmatrix} 0 & 0 & 0 \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \quad \text{and } B = \gamma\beta' = \begin{pmatrix} 0 & 0 & 0 \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix}.$$

We denote this hypothesis by $H_2^\dagger$. Observe, that under $H_2^\dagger$ the loading matrices A and B indeed have reduced rank (less than or equal to) $q = 2$, as induced by the zero row. Note also under $H_2^\dagger$, $\gamma_c = (1, 0, 0)'$ such that $\gamma_c'\lambda_t = \lambda_{1t}$ is constant, while the remaining two linear combinations in $\gamma'\lambda_t = (\lambda_{2t}, \lambda_{3t})'$ are time-varying.

Remark 7. *The case of testing for a zero row in A and B, or $H_2^\dagger$, is similar to testing the hypothesis of weak exogeneity known from cointegration analysis, see Harbo* et al. *(1998).*

In terms of testing $H_2^\dagger$, it follows that β_{11} 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 λ_{2t} and λ_{3t} under $H_2^\dagger$,

$$\begin{aligned} \lambda_{jt} &= \omega_j + \sum_{i=1}^{3} \alpha_{ji} \left(V_i' X_{t-1}\right)^{\odot 2} + \sum_{i=1}^{3} \beta_{ji}\lambda_{it-1} \\ &= (\omega_j + \beta_{j1}\omega_1) + \sum_{i=1}^{3} \alpha_{ji} \left(V_i' X_{t-1}\right)^{\odot 2} + \sum_{i=2}^{3} \beta_{ji}\lambda_{it-1},\ j = 2, 3. \end{aligned}$$

Hence, in addition to β_{11}, we also see that the parameters β_{21} and β_{31} 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 H_2^* which is given by

$$H_2^* :\ \alpha_{1i} = 0 \text{ for } i = 1, 2, 3 \text{ and } \beta_{1j} = 0 \text{ for } j = 2, 3. \tag{4.2}$$

The idea is to apply a sup likelihood ratio (supLR) test, where the supremum is taken over the non-identified parameters β_{11}, β_{21} and β_{31}.

To distinguish the non-identified parameters from the identified, partition the parameter vector θ as $\theta = (\tau', \delta')'$, with (the identified) $\tau = \left((\omega_i)_{i=1}^3, (\alpha_{ij})_{i,j=1}^3, (\beta_{ij})_{i=1,2,3\ j=2,3}, (\phi_i)_{i=1}^3\right)'$ and (the unidentified) $\delta = (\beta_{i1})_{i=1}^3$. Similarly, with Θ defined in (3.7), the equivalent partition of the parameter space is given by the product $\Theta = \Theta_\tau \times \Theta_\delta$, where Θ_τ and Θ_δ are compact. The parameter space for τ as restricted by H_2^* is given by

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

The test relies on estimating τ restricted and unrestricted for a given $\delta \in \Theta_\delta$, with the restricted and unrestricted estimators given respectively by,

$$\tilde{\tau}_{T,\delta} = \arg\max_{\tau\in\Theta_\tau^*} L_T(\tau,\delta) \quad \text{and} \quad \hat{\tau}_{T,\delta} = \arg\max_{\tau\in\Theta_\tau} L_T(\tau,\delta), \quad \text{for } \delta\in\Theta_\delta. \tag{4.3}$$

The supLR statistic is given by

$$\sup \mathrm{LR}_T(H_2^*) = \sup_{\delta\in\Theta_\delta} L_T(\hat{\tau}_{T,\delta},\delta) - \sup_{\delta\in\Theta_\delta} L_T(\tilde{\tau}_{T,\delta},\delta). \tag{4.4}$$

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

$$\sup \mathrm{LR}_T(H_2^*) \xrightarrow{d} \mathcal{L}, \tag{4.5}$$

with $\mathcal{L}$ given by (B.4). Also in Appendix B the implementation of the asymptotic test is discussed which is applied in Section 5.

Remark 8. *The key conditions for (4.5) as given in Appendix B are: (i) that $\tilde{\tau}_{T,\delta}$ and $\hat{\tau}_{T,\delta}$ are consistent for $\tau_0\in\Theta_\tau$ for any $\delta\in\Theta_\delta$, (ii) that the score as a process indexed by δ converges weakly to a Gaussian process, and (iii) that the Hessian matrix is invertible uniformly on Θ_δ. The conditions (i) and (iii) rely on finding conditions such that τ_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 3.8 in order to prove tightness. Likewise, due to the fact that τ_0 is a boundary point of Θ_τ, it may require higher-order moments of X_t in order so show that ratios of the type (3.11) 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 H_q of reduced rank q in the p-dimensional λ-GARCH model with general γ, α and β matrices.

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

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

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 H_2^* example, implies that the $(p-q)^2$ parameters $\bar{\gamma}_c'B\gamma_c$ are not identified. Moreover, as the linear combinations $\bar{\gamma}_c'\lambda_t$ are constant, also $\bar{\gamma}'B\gamma_c$ are not identified in the equation for $\bar{\gamma}'\lambda_t$. Collecting terms, using (γ, γ_c) is of full rank p by definition, it holds that the unidentified $(p \times (p-q))$ dimensional parameter matrix δ is given by

$$\delta = B\gamma_c.$$

As above one may consider a sup-based testing approach keeping δ fixed, and a supLR test statistic similar to (4.4) can be computed. However, the fact that γ 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 apply the λ-GARCH model to daily returns of three financial equities[3] from the S&P 500 Index with sample period January 3rd 2006 to January 2nd 2018 (with $T = 3020$ observations).

The log-returns are shown in Figure 5.1. Initial inspection reveals, as expected, heavy-tailedness and that the log-returns can be characterized by having ARCH effects, or volatility clustering. As to the observed volatility clustering it seems to occur during the same epochs of time, and hence the log-returns tentatively share a common factor (or eigenvalue) driving their conditional volatilities.

In the following we consider: *(i)* dynamics of the estimated eigenvalues, *(ii)* testing reduced rank of the loading matrices A and B, as well as for constant eigenvalues, and *(iii)* adequacy of the λ-GARCH model. Overall, we make the following notes: First, the λ-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 the fact that the three equities all are 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, see below, that the parameter matrices A and B are of reduced rank is novel, and it may have implications for the applications

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

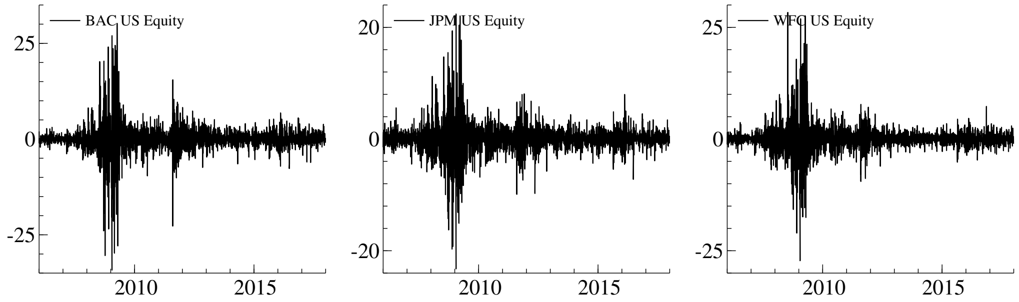


FIGURE 5.1: Log-returns of the three series analyzed: Bank of America corp. (BAC), JPMorgan Chase & co. (JPM), and Wells Fargo (WFC).

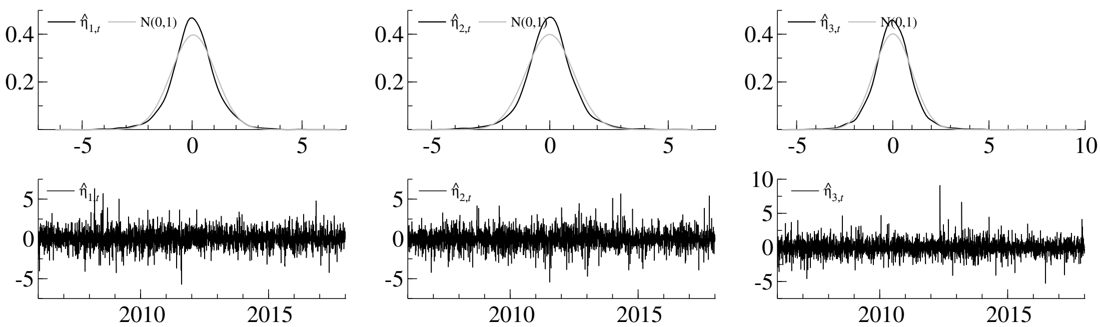


FIGURE 5.2: Estimated standardized residuals $\{\hat{\eta}_{it}\}_{t=1}^{T}$ from the Gaussian λ-GARCH model.

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.

5.1 Dynamics of the estimated eigenvalues

As to the dynamic variation of the time-varying estimated eigenvalues $(\hat{\lambda}_{it})_{i=1}^{3}$ in the first row of Figure 5.3, we note that $\hat{\lambda}_{3t}$ on average explains about 85% of the variation of the aggregated eigenvalues, $\sum_{i=1}^{3} \hat{\lambda}_{it}$. Moreover, the estimated corresponding eigenvector $\hat{V}_3$ reveals that $\hat{\lambda}_{3t}$ may be interpreted as a "market factor", with each asset having a (normalized) weight of roughly 30%. The two remaining eigenvalues $\hat{\lambda}_{1t}$ and $\hat{\lambda}_{2t}$ each explain $6 - 8\%$ of the variation on average, and the loadings of the corresponding eigenvectors correspond to long-short portfolios. Importantly, while the two smaller eigenvalues are of lesser importance compared to the "market eigenvalue", they appear non-constant, and all orthogonalized returns, $V(\hat{\phi}_T)'X_t$, have inherited ARCH effects, as can be seen from Figure 5.4.[4]

[4]Note that, in line with the Monte Carlo results, unreported estimation results show that the dynamics of the eigenvalues in $\hat{\Lambda}_t$ do not change if ϕ_{ij} is allowed to lie in $[\phi_L, \phi_U]$ with $\phi_L = -\pi/2$ (and not $\phi_L = 0$).

Table 5.1: Estimation of the λ-GARCH.

Rank	W	A			B			ϕ	V		
	0.105	0.122	0.152	0.010	4.19×10^{-5}	0.126	0.045	0.323	0.712	-0.239	0.661
	(0.040)	(0.050)	(0.076)	(0.003)	(2.64×10^{-4})	(0.240)	(0.013)	(0.071)	(0.018)	(0.054)	(0.007)
$q = 3$	0.094	0.139	0.108	0.006	0.060	2.55×10^{-8}	0.027	0.722	-0.238	0.803	0.546
	(0.027)	(0.062)	(0.054)	(0.002)	(0.166)	(4.03×10^{-7})	(0.015)	(0.054)	(0.061)	(0.018)	(0.006)
	0.039	0.081	0.168	0.071	3.98×10^{-9}	3.66×10^{-8}	0.910	0.815	-0.661	-0.546	0.515
	(0.028)	(0.067)	(0.124)	(0.017)	(1.65×10^{-7})	(4.12×10^{-7})	(0.021)	(0.047)	(0.041)	(0.050)	(0.007)

		Hypothesis H_2^*		Hypothesis H_2	
Log-likelihood	-14941.04				
AIC	29930.08	supLR test	182.27	LR test	2.11
BIC	30074.40	95%-CV	137.61	95%-CV	23.51
ξ	-0.013 (0.006)				

Here ξ denotes the top Lyapunov exponent, and standard errors are reported below the point estimates. We use the delta-method to obtain standard errors for V and ξ. The reported LR statistics are used for testing H_2^* and H_2 respectively.

5.2 Testing reduced rank

As to reduced rank and constancy of eigenvalues, consider initially the hypothesis that the (on average) smallest eigenvalue λ_{1t} is constant, i.e., H_2^* in (4.2). From Table 5.1 it follows that one cannot accept the hypothesis based on the supLR test (see Appendix B.1 for details on implementation[5]). Intuitively, this is sensible as under the hypothesis H_2^*, λ_{1t} is constant and the associated orthogonalized returns $V_1'X_t$ homoskedastic. As already noted, this is not the case as all three orthogonalized returns $V(\hat{\phi}_T)'X_t$ in the unrestricted model appear to exhibit volatility clustering (see Figure 5.4).

Table 5.2: Estimation of the λ-GARCH under H_2.

Rank	W	$A = \alpha\gamma'$			$B = \beta\gamma'$			ϕ	V		
	0.108	0.143	0.138	0.009	1.10×10^{-7}	0.087	0.047	0.338	0.717	-0.225	0.660
	(0.037)	(0.042)	(0.057)	(0.002)	(0.001)	(0.195)	(0.012)	(0.065)	(0.016)	(0.050)	(0.007)
$q = 2$	0.094	0.125	0.119	0.006	3.01×10^{-8}	0.077	0.027	0.708	-0.252	0.799	0.546
	(0.024)	(0.053)	(0.048)	(0.002)	(0.001)	(0.166)	(0.010)	(0.051)	(0.058)	(0.018)	(0.006)
	0.038	0.080	0.162	0.071	4.16×10^{-6}	6.52×10^{-10}	0.911	0.825	-0.650	-0.558	0.516
	(0.027)	(0.067)	(0.116)	(0.017)	(7.05×10^{-6})	(4.89×10^{-8})	(0.021)	(0.043)	(0.038)	(0.046)	(0.007)
Log-likelihood		-14942.09	AIC		29924.19	BIC	30044.45	ξ	-0.013 (0.006)		

Table 5.3: Estimated parameters - reduced rank matrices, under H_2.

Rank	α'			β'			γ'		
$q = 2$	0.125	0.119	0.006	3.01×10^{-8}	0.077	0.027	1.132	1	0
	(0.053)	(0.048)	(0.002)	(0.001)	(0.166)	(0.010)	(0.374)		
	0.080	0.162	0.071	4.16×10^{-6}	6.52×10^{-10}	0.911	0.018	0	1
	(0.067)	(0.116)	(0.017)	(7.05×10^{-6})	(4.89×10^{-8})	(0.021)	(0.018)		

Recall that $A = \gamma\alpha'$ and $B = \gamma\beta'$ for $q < p$, with A and B given in Table 2.

Next consider the less restrictive hypothesis of reduced rank $r = 2$ of A and B, that is H_2 in (4.1). Under H_2 all eigenvalues are allowed to remain time-varying, while $p - r = 1$

[5] For each entry of the non-identified parameter vector $\delta = (\beta_{11}, \beta_{21}, \beta_{31})'$ we use $k = 20$ equidistant points between 0 and 0.99 (both points included), leading to a grid of $20^3 = 8000$ points for the grid search over δ. 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 (rcond), and *ii)* the log-likelihood value is close to the maximum likelihood value. That is, for $i = 1, \dots, \dim\Delta$, 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 $M = 10.000$ Monte Carlo draws to determine the critical value.

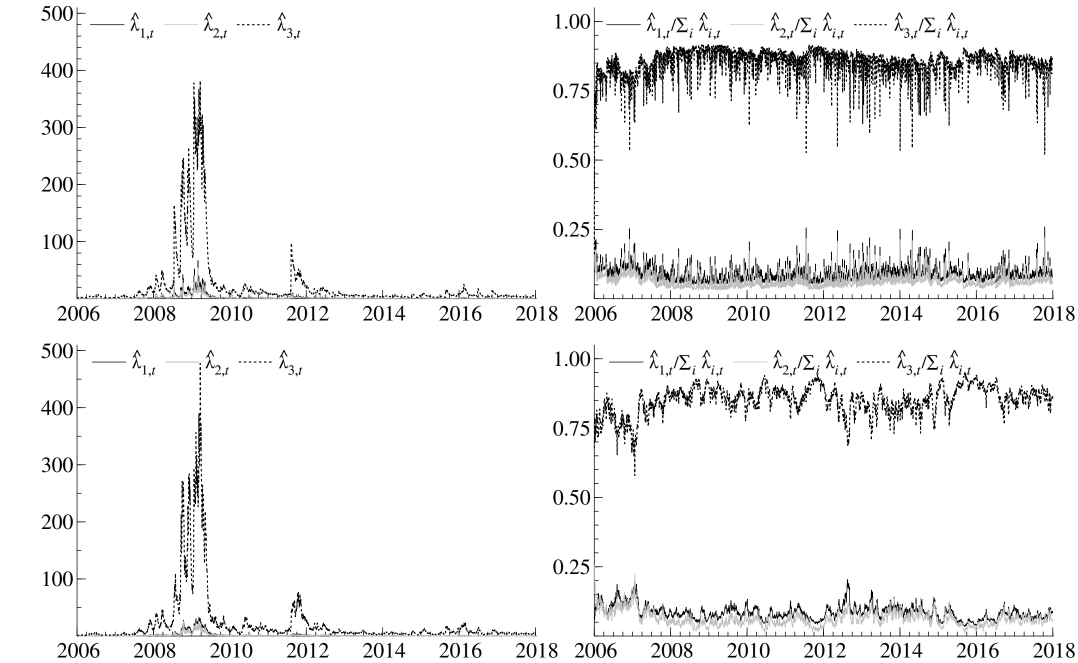


FIGURE 5.3: The first row shows the estimated conditional eigenvalues based on the Gaussian λ-GARCH model, while the second row shows estimated eigenvalues from the Student's t version. The left hand side shows the level of the eigenvalues, while the right hand side show how much variation in the data is explained by each of the estimated eigenvalues across the sample.

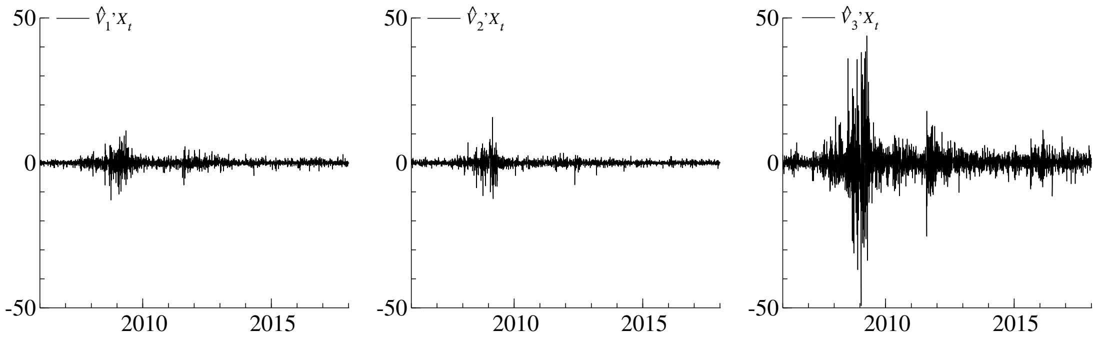


FIGURE 5.4: Estimated rotated returns, $\{V'(\hat{\phi}_T)X_t\}_{t=1}^{T}$ from the Gaussian λ-GARCH model.

linear combination of these is constant. To ensure identification of γ, α and β under H_2, the lower (2×2) block of γ is set to I_2, while the first row of γ is freely varying. We obtain critical values by the bootstrap algorithm in Appendix C, see also Cavaliere *et al.* (2020). The critical value is obtained from $B = 399$ bootstrap replications. The LR statistic is 2.1 and the associated bootstrapped 95% critical value is 17.55 such that H_2 is not rejected.[6] From the estimated parameters for the reduced rank model (reported in tables 5.2 and 5.3), the estimated parameters, eigenvalues, and conditional covariances for unrestricted model and reduced rank model are non-distinguishable, and based on the AIC and BIC information criteria, the reduced rank model seems in fact to be preferable to the unrestricted model.

Remark 9. *As kindly raised by the Editor, note that while for the considered sample of 12 years of data the λ-GARCH model appears well-specified (please see Section 5.3 for more details) and supporting the hypothesis of reduced rank $r = 2$, unreported results indicate that this seems not to be the case if the sample is extended to include the 1987-crash (as for example the period 1981-2018). Thus considering longer samples with breaks, or crashes (such as the 1987-crash), may suggest that a λ-GARCH model with structrual break(s), or some other type of switching mechanism, is more suitable.*

5.3 Model adequacy

As to model adequacy we make the following observations. Table 5.1 contains the parameter estimates of the λ-GARCH model in (3.1)–(3.2), and standardized residuals are given in Figure 5.2. In terms of the regularity conditions for the asymptotic theory, the estimate $\hat{\xi}$[7] of the Lyapunov coefficient ξ in (3.6) and its standard error reported in Table 5.1 suggest that $\hat{\xi} < 0$.[8] For the standardized residuals, while unreported misspecification tests indicate no ARCH-effects and no residual autocorrelation, they appear slightly heavy-tailed as is common in Gaussian GARCH-type models. In order to address the issue of heavy-tailedness, Table 5.4 reports estimation results from the trivariate Student's t version of the λ-GARCH model as given by (2.9) and (2.10). We observe that $\hat{v} = 4.749$, confirming the heavy-tailedness, and moreover from Table 5.5 that the empirical quantiles of the standardized residuals match the $t_{\hat{v}}(0, 1)$ quantiles. Interestingly, when considering the estimated eigenvalues $\hat{\lambda}_{it}$ for $i = 1, 2, 3$ in Figure 5.3 for the Gaussian and Student's

[6]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 140.21 and a bootstrapped critical value of 65.15.

[7]The estimate of ξ and its standard error are obtained as in Nielsen and Rahbek (2014).

[8]As kindly pointed out by a referee it appears that some of the parameter estimates may be close to the boundary of the parameter space, which again suggests that the distribution of the associated estimators may be better approximated by "half-normal" type distributions, see e.g., Pedersen and Rahbek (2019). To avoid this, an alternative could be to allow for negative entries in the A and B matrices, see also Conrad and Karanasos (2010). However, extending the asymptotic theory is beyond the scope of this paper; unreported estimation results indicates that allowing this gives (a few) negative entries in $\hat{B}$ while the estimated conditional eigenvalues remain positive (and with similar dynamics).

t cases, it follows that the normalized eigenvalues $\hat{\lambda}_{it}/\left(\sum_{i=1}^{3}\hat{\lambda}_{it}\right)$ for $i = 1, 2, 3$ appear more "smooth" for the Student's t case, possibly reflecting the dampening effect of the weight w_t in (2.10) as previously discussed.

TABLE 5.4: Estimation of the Student's t $\lambda-$GARCH model.

v	W		$\mathcal{A}$			$\mathcal{B}$		ϕ		V	
	0.008	0.025	0.047	0.003	0.973	6.08×10^{-10}	0.002	0.294	0.706	-0.257	0.660
4.749	0.033	0.034	0.083	0.003	0.165	0.707	9.59×10^{-9}	0.741	-0.214	0.811	0.545
	0.087	0.172	3.44×10^{-7}	0.075	5.48×10^{-8}	3.51×10^{-8}	0.995	0.794	-0.675	-0.526	0.517
Log-likelihood		-14401.90									

TABLE 5.5: Empirical quantiles of the standardized residuals from the Student's t $\lambda-$GARCH.

Quantile	0.010	0.025	0.050	0.250	0.500	0.750	0.950	0.975	0.990
$t_{4.749}(0,1)$	- 2.62	- 1.99	- 1.55	- 0.56	0.00	0.56	1.55	1.99	2.62
$N(0,1)$	- 2.33	- 1.96	- 1.64	- 0.67	0.00	0.67	1.64	1.96	2.33
$\hat{\eta}_{1,t}$	- 2.65	- 2.10	- 1.62	- 0.58	0.01	0.55	1.51	1.97	2.46
$\hat{\eta}_{2,t}$	- 2.89	- 2.06	- 1.55	- 0.55	0.01	0.55	1.54	1.96	2.66
$\hat{\eta}_{3,t}$	- 2.70	- 2.00	- 1.56	- 0.58	- 0.03	0.52	1.47	1.92	2.73

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, we let $Y_t(\theta) := V(\phi)'X_t$ 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 $\theta \in \Theta$,

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

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

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

For derivatives,

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

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

Furthermore we let $\Omega_t^\star = \Omega_t^\star(\theta_0)$, that is $\Omega_t^\star$ evaluated at the true parameter values, θ_0. The same holds for other quantities which depending on θ_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), and that Assumption 3.3 is equivalent to A4. Moreover, recall that $\rho(B) < 1$ on Θ, 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

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

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

(i) With $\Omega_t(\theta)$ and $\Omega_t^\star(\theta)$ defined in (3.9) 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.

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 (3.10) and (A.3), respectively, using that $\sup_{\theta\in\Theta} \rho(B) < 1$, we have that

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

A.3 Proof of Theorem 3.3

Using that $\theta_0 \in \text{int}\Theta$, with Θ 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),

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial l_t^\star(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, \Sigma), \tag{A.4}$$

with

$$\Sigma := E\left[\frac{\partial l_t^\star(\theta_0)}{\partial \theta} \frac{\partial l_t^\star(\theta_0)}{\partial \theta'}\right] \quad \text{non-negative definite.} \tag{A.5}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} E\left[\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'}\right] =: J, \tag{A.6}$$

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 θ_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 (3.8) and $l_t^\star(\theta)$ defined in (A.1), for some neighborhood $N(\theta_0)$ around θ_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.2 we have that $E[\partial l_t^\star(\theta_0)/\partial\theta|\mathcal{F}_{t-1}] = 0$, $E[\|\partial l_t^\star(\theta_0)/\partial\theta\|] < \infty$, and that Σ is non-negative definite. By a CLT for stationary and ergodic martingale differences (e.g., Brown, 1971), we conclude that (A.4) holds.
Proof of Hessian: From Lemma A.5, we have that $E[\|\partial^2 l_t^\star(\theta_0)/\partial\theta\partial\theta'\|] < \infty$. By the Ergodic Theorem, we conclude that (A.6) holds. Moreover, Lemma A.5 states that the matrix J is invertible.
Proof of Third Order Derivative: This property holds by Lemma A.6.
Proof of Initial Values: This holds by arguments similar to the ones given in Francq and Zakoïan (2012, pp.204-206).

A.4 Proof of Lemma 3.1

We make some initial considerations about the structure of $V(\phi)$. Note that

$$V(\phi) = \prod_{i=1}^{p-1}\prod_{j=i+1}^{p} R(i,j;\phi) = \prod_{i=1}^{p-1}\tilde{V}_i(\phi),$$

where for $i = 1,\ldots p-1$,

$$\tilde{V}_i(\phi) := \prod_{j=i+1}^{p} R(i,j;\phi). \tag{A.7}$$

Note in particular that, by construction, $\tilde{V}_i(\phi)$ depends only on the rotation parameters $(\phi_{i,i+1},\ldots,\phi_{i,p})$. Define, for $k=1,\ldots,p-1$,

$$U_k := \left(\prod_{i=k}^{p-1}\tilde{V}_i(\phi)\right)'\left(\prod_{i=k}^{p-1}\tilde{V}_i(\phi_0)\right). \tag{A.8}$$

In particular, we note that

$$U_1 = \left(\prod_{i=1}^{p-1}\tilde{V}_i(\phi)\right)'\left(\prod_{i=1}^{p-1}\tilde{V}_i(\phi_0)\right) = V(\phi)'V(\phi_0).$$

We will rely on some essential features of U_k stated in Lemma A.1 in the next section.

Suppose that $\Omega_t^\star(\theta)=\Omega_t^\star(\theta_0)$ almost surely. Then, almost surely,

$$U_1\Lambda_t^\star(\theta_0) = \Lambda_t^\star(\theta)U_1. \tag{A.9}$$

Hence, in light of Lemma A.1, we have that (almost surely)

$$\lambda_{1t}^\star(\theta_0) = \lambda_{1t}^\star(\theta). \tag{A.10}$$

Moreover, in light of (A.9), we also have that

$$U_1\Lambda_t^\star(\theta_0)U_1' = \Lambda_t^\star(\theta),$$

which combined with (A.10) yields that (almost surely)

$$(U_{1,11}^2-1)\lambda_{1t}^\star(\theta_0) + \sum_{j=2}^{p} U_{1,1j}^2\lambda_{jt}^\star(\theta_0) = 0.$$

Suppose that $(\phi_{1,2},\ldots,\phi_{1,p}) \neq (\phi_{1,2,0},\ldots,\phi_{1,p,0})$ such that in light of Lemma A.1, $U_{1,11}^2 < 1$, which implies that (almost surely)

$$\lambda_{1t}^\star(\theta_0) = (1-U_{1,11}^2)^{-1}\sum_{j=2}^{p} U_{1,1j}^2\lambda_{jt}^\star(\theta_0).$$

This implies that

$$\lambda_t^\star(\theta_0) = \begin{pmatrix} \lambda_{1t}^\star(\theta_0) \\ \lambda_{2t}^\star(\theta_0) \\ \vdots \\ \lambda_{pt}^\star(\theta_0) \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{1-U_{1,11}^2} \sum_{j=2}^p \omega_{j,0} U_{1,1j}^2 \\ \omega_{2,0} \\ \vdots \\ \omega_{p,0} \end{pmatrix} + \begin{pmatrix} \frac{1}{1-U_{1,11}^2} \sum_{j=2}^p A_{j,0} U_{1,1j}^2 \\ A_{2,0} \\ \vdots \\ A_{p,0} \end{pmatrix} (Y_{t-1})^{\odot 2}$$

$$+ \begin{pmatrix} \frac{1}{1-U_{1,11}^2} \sum_{j=2}^p B_{j,0} U_{1,1j}^2 \\ B_{2,0} \\ \vdots \\ B_{p,0} \end{pmatrix} \lambda_{t-1}^\star(\theta_0),$$

where $A_{i,0}$ ($B_{i,0}$) denotes the ith row of A_0 (B_0). Note that it must hold that $U_{1,1j}$ is non-zero for some $j = 2, \dots, p$ – otherwise $\lambda_{1t}^\star(\theta_0)$ is degenerate which is ruled out by Assumptions 3.4-3.5. Hence, $[A_0, B_0]$ has reduced rank, which is ruled out by Assumption 3.5. We conclude that $U_{1,11} = 1$, which in light of Lemma A.1 implies that $(\phi_{1,2}, \dots, \phi_{1,p}) = (\phi_{1,2,0}, \dots, \phi_{1,p,0})$. This in turn implies that $\tilde{V}_1(\phi) = \tilde{V}_1(\phi_0)$, so that

$$\begin{aligned} U_1 &= \left(\prod_{i=1}^{p-1} \tilde{V}_i(\phi)\right)' \left(\prod_{i=1}^{p-1} \tilde{V}_i(\phi_0)\right) \\ &= \left(\prod_{i=2}^{p-1} \tilde{V}_i(\phi)\right)' \tilde{V}_1(\phi)'\tilde{V}_1(\phi_0) \left(\prod_{i=2}^{p-1} \tilde{V}_i(\phi_0)\right) \\ &= \left(\prod_{i=2}^{p-1} \tilde{V}_i(\phi)\right)' \left(\prod_{i=2}^{p-1} \tilde{V}_i(\phi_0)\right) = U_2. \end{aligned}$$

This combined with (A.9) implies that (almost surely)

$$U_2 \Lambda_t^\star(\theta_0) = \Lambda_t^\star(\theta) U_2 \tag{A.11}$$

and

$$U_2 \Lambda_t^\star(\theta_0) U_2' = \Lambda_t^\star(\theta). \tag{A.12}$$

By arguments similar to the ones given above, in light of Lemma A.1, we have that

$$\lambda_{2t}^\star(\theta_0) = \lambda_{2t}^\star(\theta), \tag{A.13}$$

and hence using (A.12) (almost surely)

$$(U_{2,22}^2 - 1)\lambda_{2t}^{\star}(\theta_0) + \sum_{j=1,j\neq 2}^{p} U_{2,2j}^2 \lambda_{jt}^{\star}(\theta_0) = 0.$$

Now suppose that $(\phi_{2,3}, \ldots, \phi_{2,p}) \neq (\phi_{2,3,0}, \ldots, \phi_{2,p,0})$ such that in light of Lemma A.1, $U_{2,22}^2 < 1$, which implies that (almost surely)

$$\lambda_{2t}^{\star}(\theta_0) = (1 - U_{2,22}^2)^{-1} \sum_{j=1,j\neq 2}^{p} U_{2,2j}^2 \lambda_{jt}^{\star}(\theta_0).$$

By arguments similar to the ones above, we have that this violates the assumption that $[A_0, B_0]$ has full rank. By contradiction we have that $(\phi_{2,3}, \ldots, \phi_{2,p}) = (\phi_{2,3,0}, \ldots, \phi_{2,p,0})$. This combined with the fact that $(\phi_{1,2}, \ldots, \phi_{1,p}) = (\phi_{1,2,0}, \ldots, \phi_{1,p,0})$ implies (by arguments identical to the ones given above) that $U_1 = U_3$. By identical arguments, it must hold that $(\phi_{3,4}, \ldots, \phi_{3,p}) = (\phi_{3,4,0}, \ldots, \phi_{3,p,0})$, and by repeating these arguments we have that $\phi = \phi_0$. The identification of the remaining parameters (W, A, B) follows by standard arguments, see e.g., Francq and Zakoïan (2012, pp.196-197), using Assumptions 3.4-3.5.

A.5 Auxiliary Lemmas

Lemma A.1. *Let $\tilde{V}_i(\phi)$ be defined in (A.7). Let $\tilde{V}_{i,ji}(\phi)$ denote entry (j,i) of $\tilde{V}_i(\phi)$. The following holds.*

1. *The $i-$th column of $\tilde{V}_i(\phi)$ is given by $(\tilde{V}_{i,1i}(\phi), \ldots, \tilde{V}_{i,pi}(\phi))'$ where for $j = 1, \ldots, p$*

$$\tilde{V}_{i,ji}(\phi) = \begin{cases} 0 & \text{if } j < i \\ \prod_{k=i+1}^{p} \cos(\phi_{i,k}) & \text{if } j = i \\ -\prod_{k=j+1}^{p} \cos(\phi_{i,k}) \sin(\phi_{i,j}) & \text{if } j > i \end{cases},$$

   with the convention that $\prod_{k=p+1}^{p} \cos(\phi_{i,k}) = 1$.

2. *For $k = 1, \ldots, p-1$, the k-th column of $\prod_{i=k}^{p-1} \tilde{V}_i(\phi)$ equals the k-th column of $\tilde{V}_k(\phi)$.*

3. *Let* U_k *be given by (A.8). With* $U_{k,kk}$ *the* (k,k) *entry of* U_k*, it holds that*

$$\begin{aligned} U_{k,kk} &= \sum_{j=1}^{p} \tilde{V}_{k,jk}(\phi)\tilde{V}_{k,jk}(\phi_0) && \text{(A.14)} \\ &= \prod_{j=k+1}^{p} \cos(\phi_{k,j})\cos(\phi_{k,j,0}) && \text{(A.15)} \\ &+ \sum_{j=k+1}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l})\cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j})\sin(\phi_{k,j,0}) \\ &= \sum_{j=k+1}^{p} \left(\cos(\phi_{k,j} - \phi_{k,j,0}) - 1\right) \prod_{l=j+1}^{p} \cos(\phi_{k,l})\cos(\phi_{k,l,0}) + 1 && \text{(A.16)} \end{aligned}$$

4. *For* $\phi \in [0, \pi/2]^{p(p-1)/2}$ *and* $\phi_0 \in (0, \pi/2)^{p(p-1)/2}$ *it holds that* $U_{k,kk} > 0$.

5. *For* $\phi \in [0, \pi/2]^{p(p-1)/2}$ *and* $\phi_0 \in (0, \pi/2)^{p(p-1)/2}$, $U_{k,kk} \leq 1$ *with equality if and only if* $(\phi_{k,k+1}, \ldots, \phi_{k,p}) = (\phi_{k,k+1,0}, \ldots, \phi_{k,p,0})$.

Proof of Lemma A.1: Points 1-2 follow immediately by inspecting the structure of $\tilde{V}_i(\phi)$. The equality in (A.14) follows by noticing that, by definition of U_k and point 2, $U_{k,kk}$ equals the dot product of the k-th columns of $\tilde{V}_k(\phi)$ and $\tilde{V}_k(\phi_0)$. The equality in (A.15) follows by point 1. The equality in (A.16) follows by repeated use of trigonometric identities:

$$\prod_{j=k+1}^{p} \cos(\phi_{k,j}) \cos(\phi_{k,j,0}) + \sum_{j=k+1}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j}) \sin(\phi_{k,j,0})$$

$$= \prod_{j=k+1}^{p} \cos(\phi_{k,j}) \cos(\phi_{k,j,0}) + \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,k+1}) \sin(\phi_{k,k+1,0})$$

$$+ \sum_{j=k+2}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j}) \sin(\phi_{k,j,0})$$

$$= \cos(\phi_{k,k+1}) \cos(\phi_{k,k+1,0}) \sin(\phi_{k,k+1}) \sin(\phi_{k,k+1,0}) \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right)$$

$$+ \sum_{j=k+2}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j}) \sin(\phi_{k,j,0})$$

$$= [\cos(\phi_{k,k+1} - \phi_{k,k+1,0}) - 1] \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right)$$

$$+ \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) + \sum_{j=k+2}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j}) \sin(\phi_{k,j,0})$$

$$= [\cos(\phi_{k,k+1} - \phi_{k,k+1,0}) - 1] \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right)$$

$$+ [\cos(\phi_{k,k+2} - \phi_{k,k+2,0}) - 1] \left(\prod_{l=k+3}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right)$$

$$+ \left(\prod_{l=k+3}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) + \sum_{j=k+3}^{p} \left(\prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) \sin(\phi_{k,j}) \sin(\phi_{k,j,0})$$

$$= [\cos(\phi_{k,k+1} - \phi_{k,k+1,0}) - 1] \left(\prod_{l=k+2}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right)$$

$$+ [\cos(\phi_{k,k+2} - \phi_{k,k+2,0}) - 1] \left(\prod_{l=k+3}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \right) + \dots$$

$$+ [\cos(\phi_{k,p-1} - \phi_{k,p-1,0}) - 1] \cos(\phi_{k,p}) \cos(\phi_{k,p,0})$$

$$+ \cos(\phi_{k,p}) \cos(\phi_{k,p,0}) + \sin(\phi_{k,j}) \sin(\phi_{k,j,0}) + (1 - 1)$$

$$= \sum_{j=k+1}^{p} (\cos(\phi_{k,j} - \phi_{k,j,0}) - 1) \prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) + 1.$$

Turning to point 4, note that all terms in (A.15) are non-negative for $\phi_{k,k+1}, \dots, \phi_{k,p} \in [0, \pi/2]^{p-k}$ and $\phi_{k,k+1,0}, \dots, \phi_{k,p,0} \in (0, \pi/2)^{p-k}$. Hence, $U_{k,kk} \geq 0$ with equality if and

only if all terms in (A.15) are zero. Noting that $\cos(x)$ and $\sin(x)$ are strictly positive for $x \in (0, \pi/2)$, we have that for $\phi_{k,k+1,0}, \ldots, \phi_{k,p,0} \in (0, \pi/2)^{p-k}$, $U_{k,kk} = 0$ if and only if $\prod_{j=k+1}^{p} \cos(\phi_{k,j}) = \prod_{j=k+2}^{p} \cos(\phi_{k,j}) \sin(\phi_{k,k+1}) = \prod_{j=k+3}^{p} \cos(\phi_{k,j}) \sin(\phi_{k,k+2}) = \ldots = \cos(\phi_{k,p}) \sin(\phi_{k,p-1}) = \sin(\phi_{k,p}) = 0$. This implies that $\phi_{k,p} = 0$. Hence $\cos(\phi_{k,p}) = 1$, and it must hold that $\phi_{k,p-1} = 0$ in order to have that $\cos(\phi_{k,p}) \sin(\phi_{k,p-1}) = 0$. By similar arguments, we must have that $\phi_{k,k+1} = \ldots = \phi_{k,p} = 0$. But this implies that $\prod_{j=k+1}^{p} \cos(\phi_{k,j}) = 1 > 0$. Hence, it is not possible that $U_{k,kk} = 0$, and we conclude that $U_{k,kk} > 0$.
Lastly, turning to point 5, we start out by considering (A.16). For $j = k+1, \ldots, p$ we have that $\xi_j := (\cos(\phi_{k,j} - \phi_{k,j,0}) - 1) \prod_{l=j+1}^{p} \cos(\phi_{k,l}) \cos(\phi_{k,l,0}) \leq 0$. Hence $U_{k,kk} = \sum_{j=k+1}^{p} \xi_j + 1 \leq 1$, with equality if and only if $\xi_j = 0$ for all $j = k+1, \ldots, p$. Clearly $\xi_j = 0$ for all $j = k+1, \ldots, p$ if $(\phi_{k,k+1}, \ldots, \phi_{k,p}) = (\phi_{k,k+1,0}, \ldots, \phi_{k,p,0})$. Suppose now that $\xi_j = 0$ for all $j = k+1, \ldots, p$. Note that $\phi_{k,j} - \phi_{k,j,0} \in (-\pi/2, \pi/2)$ so that $\cos(\phi_{k,j} - \phi_{k,j,0}) - 1 = 0$ if and only if $\phi_{k,j} = \phi_{k,j,0}$. Hence $\xi_p = (\cos(\phi_{k,p} - \phi_{k,p,0}) - 1) = 0$ which implies that $\phi_{k,p} = \phi_{k,p,0}$. This in turn implies that $\xi_{p-1} = (\cos(\phi_{k,p-1} - \phi_{k,p-1,0}) - 1) \cos(\phi_{k,p}) \cos(\phi_{k,p,0}) = 0$ which can only be the case if $\phi_{k,p-1} = \phi_{k,p-1,0}$ since $\cos(\phi_{k,p}) \cos(\phi_{k,p,0}) \neq 0$ as $\phi_{k,p} = \phi_{k,p,0} \in (0, \pi/2)$. By similar arguments, we conclude that $(\phi_{k,k+1}, \ldots, \phi_{k,p}) = (\phi_{k,k+1,0}, \ldots, \phi_{k,p,0})$.

Lemma A.2. *With $l_t^\star(\theta)$ defined in (A.1), under Assumptions 3.1-3.2,3.4-3.9, it holds that*

$$E\left[\frac{\partial l_t^\star(\theta_0)}{\partial \theta} \middle| \mathcal{F}_{t-1} \right] = 0 \text{ almost surely,} \tag{A.17}$$

$$E\left[\left\| \frac{\partial l_t^\star(\theta_0)}{\partial \theta} \right\|^2 \right] < \infty, \quad \text{and} \tag{A.18}$$

$$\Sigma = E\left[\frac{\partial l_t^\star(\theta_0)}{\partial \theta} \frac{\partial l_t^\star(\theta_0)}{\partial \theta'} \right] \tag{A.19}$$

is non-negative definite.

Proof of Lemma A.2: With $Y_t(\theta) = V(\phi)' X_t$, for $i = 1, \ldots, d_\theta$, we have from Lemma A.3 that

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

with $\dot{\lambda}_{t,i}(\theta) := \partial \lambda_t^\star(\theta) / \partial \theta_i$ and $\dot{Y}_{t,i}(\theta) := \partial Y_t(\theta) / \partial \theta_i$. Evaluating at θ_0, we have

$$\begin{aligned} \frac{\partial l_t^\star(\theta_0)}{\partial \theta_i} &= \operatorname{tr}\left\{ \Lambda_t^{\star -1} \dot{\Lambda}_{t,i}^\star \left[I_p - \eta_t \eta_t' \right] \right\} + 2 \dot{Y}_{t,i}' \Lambda_t^{\star -1} Y_t \\ &=: M_{1,t,i} + M_{2,t,i}. \end{aligned} \tag{A.20}$$

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

$$E[M_{1,t,i}|\mathcal{F}_{t-1}] = 0 \text{ almost surely,} \tag{A.21}$$

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

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

which implies that

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

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

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

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

$$\text{tr}(S_i(\theta)) = 0. \tag{A.22}$$

For $\theta = \theta_0$ we then have

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

using $E[X_tX_t'|\mathcal{F}_{t-1}] = \Omega_t^{\star}$ and (A.22),

$$E[M_{t,2,i}|\mathcal{F}_{t-1}] = 2\text{tr}\{S_i\} = 0 \text{ almost surely.} \tag{A.23}$$

Combining (A.20), (A.21), and (A.23), we conclude that (A.17) holds. Turning to (A.18), 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.20) 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,

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

where we note that $E[\eta_{q,t}^4] < \infty$, $q = 1, \ldots, p$, by Assumption 3.8. 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.7, 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 $\tilde{S}_i = VS_i'V'$,

$$\begin{aligned}
M_{2,t,i}^2 &= 4X_t'S_iV\Lambda_t^{\star-1}V'X_tX_t'S_iV\Lambda_t^{\star-1}V'X_t \\
&= 4\text{tr}\left(S_i\Omega_t^{\star-1}X_tX_t'S_i\Omega_t^{\star-1}X_tX_t'\right) = 4\text{tr}(\tilde{S}_i'\Lambda_t^{\star-1}Y_tY_t'\tilde{S}_i'\Lambda_t^{\star-1}Y_tY_t') \\
&\leq K\|\Lambda_t^{\star-1}Y_tY_t'\|^2 = K\left(Y_t'Y_tY_t'\Lambda_t^{\star-2}Y_t\right) = K\left(\sum_{i=1}^p y_{it}^2\right)\left(\sum_{i=1}^p \frac{y_{it}^2}{\lambda_{it}^{\star 2}}\right). \qquad \text{(A.24)}
\end{aligned}$$

We note that (A.24) 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 3.8 and that for $\theta_0 \in \text{int}\Theta$,

$$\frac{\lambda_{k,t}^\star}{\lambda_{l,t}^\star} = \frac{\omega_{0,k} + \sum_{i=1}^p \alpha_{0,ki}y_{i,t-1}^2 + \sum_{i=1}^p \beta_{0,ki}\lambda_{i,t-1}^\star}{\omega_{0,l} + \sum_{i=1}^p \alpha_{0,li}y_{i,t-1}^2 + \sum_{i=1}^p \beta_{0,li}\lambda_{i,t-1}^\star} \leq \frac{\omega_{0,k}}{\omega_{0,l}} + \sum_{i=1}^p \frac{\alpha_{0,ki}}{\alpha_{0,li}} + \sum_{i=1}^p \frac{\beta_{0,ki}}{\beta_{0,li}} \leq K, \qquad \text{(A.25)}$$

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. The matrix Σ is non-negative definite by construction.

Lemma A.3. *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

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

Proof of Lemma A.3: 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,

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

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

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

Hence, the score with respect to θ_i is

$$\begin{aligned}\frac{\partial l_t^\star(\theta)}{\partial\theta_i} &= \mathrm{tr}\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\} - \mathrm{tr}\left\{Y_t(\theta)Y_t'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\right\} + 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)\\ &= \mathrm{tr}\left\{\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\left[I_p - \Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right]\right\} + 2\dot{Y}_{t,i}'(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta).\end{aligned}$$

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

$$\begin{aligned}\frac{\partial^2 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j} &= -\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}(\theta)\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\right)\\ &+ \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\ &- \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\ &- 2\mathrm{tr}\left(\tilde{S}_j'(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + 2\mathrm{tr}\left((\dot{S}_{i,j}(\theta) + S_i(\theta)S_j(\theta))\Omega_t^{\star-1}(\theta)X_tX_t'\right)\\ &+ 2\mathrm{tr}\left(V(\phi)'\left(\dot{S}_{i,j}(\theta) + S_i(\theta)S_j(\theta)\right)V(\phi)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\ &- 2\mathrm{tr}\left(\tilde{S}_i(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right) + 2\mathrm{tr}\left(\tilde{S}_i'(\theta)\Lambda_t^{\star-1}(\theta)\tilde{S}_j(\theta)Y_t(\theta)Y_t'(\theta)\right),\end{aligned} \tag{A.26}$$

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

$$S_i(\theta) = \frac{\partial V(\phi)}{\partial\theta_i}V(\phi)', \tag{A.27}$$

$$\tilde{S}_i'(\theta) = V(\phi)'S_i(\theta)V(\phi) = -V(\phi)'S_i'(\theta)V(\phi) = -\tilde{S}_i(\theta). \tag{A.28}$$

Proof of Lemma A.4: Throughout the proof, we suppress the dependence on θ. From the proof of Lemma A.3 we have that

$$\begin{aligned}\frac{\partial^2 l_t^\star(\theta)}{\partial\theta_i\partial\theta_j} &= \frac{\partial\mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star)}{\partial\theta_j} - \frac{\partial\mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\Lambda_t^{\star-1}Y_tY_t')}{\partial\theta_j} + 2\frac{\partial\dot{Y}_{t,i}'\Lambda_t^{\star-1}Y_t}{\partial\theta_j}\\ &= N_{1,t} - N_{2,t} + 2N_{3,t}.\end{aligned}$$

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

$$\frac{\partial\mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star)}{\partial\theta_j} = -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^\star\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^\star\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^\star\right). \tag{A.29}$$

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.27),

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

with $\tilde{S}_j = V'S_jV$ defined in (A.28). Hence, the second term of the Hessian, $N_{2,t}$, is

$$\frac{\partial \mathrm{tr}(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t')}{\partial\theta_j} = -\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}\ddot{\Lambda}_{t,i,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right)$$
$$- \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right) + 2\mathrm{tr}\left(\tilde{S}_j'\Lambda_t^{\star-1}\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}Y_tY_t'\right).$$

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,

$$\ddot{Y}_{t,i,j}'\Lambda_t^{\star-1}Y_t = X_t'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}V'X_t$$
$$= X_t'VV'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}V'X_t$$
$$= \mathrm{tr}\left(V'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{\star-1}Y_tY_t'\right)$$

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

$$\frac{\partial\dot{Y}'_{t,i}\Lambda_t^{\star-1}Y_t}{\partial\theta_j} = \mathrm{tr}\left(V'\left(\dot{S}_{i,j}+S_iS_j\right)V\Lambda_t^{\star-1}Y_tY'_t\right) - \mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\Lambda_t^{\star-1}Y_tY'_t\right) + \mathrm{tr}\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_jY_tY'_t\right) \tag{A.30}$$

Using (A.29)-(A.30), we have (A.26).

Lemma A.5. *With $l_t^\star(\theta)$ defined in (A.1), under Assumptions 3.1-3.2,3.4-3.9,, for $i,j = 1,\dots,d_\theta$,*

$$E\left[\left.\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta_i\partial\theta_j}\right|\mathcal{F}_{t-1}\right] = \mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\right) + 2\mathrm{tr}\left(S_iS_j\right) + 2\mathrm{tr}\left(\tilde{S}'_i\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^\star\right), \tag{A.31}$$

$$E\left[\left\|\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'}\right\|\right] < \infty, \tag{A.32}$$

and

$$J = E\left[\frac{\partial^2 l_t^\star(\theta_0)}{\partial\theta\partial\theta'}\right] \text{ is invertible.} \tag{A.33}$$

Proof of Lemma A.5: Using the expression for $\partial^2 l_t^\star(\theta)/\partial\theta_i\partial\theta_j$ from Lemma A.4, we immediately have that

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

This expression can be simplified further as both $\dot{S}_{i,j}$, $\mathrm{tr}\left(\tilde{S}_j\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\right)$ and $\mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\right)$ are skew-symmetric, and hence $\mathrm{tr}(\dot{S}_{i,j}) = \mathrm{tr}\left(\tilde{S}_j\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\right) = \mathrm{tr}\left(\tilde{S}_i\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\right) = 0$, and we obtain (A.31).

Turning to (A.32), we consider each term in (A.31). Notice that $E\left[|\mathrm{tr}\left(\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,i}\Lambda_t^{\star-1}\dot{\Lambda}^\star_{t,j}\right)|\right] < \infty$ by Lemma A.7. Trivially, $\mathrm{tr}(S_iS_j)$ is bounded, since Θ is compact and S_i is continuous in ϕ. Lastly, consider tr $\left(\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^{\star}\right)$,

$$\tilde{S}_i'\Lambda_t^{\star-1}\tilde{S}_j\Lambda_t^{\star} =$$
$$\begin{pmatrix} 0 & -\tilde{s}_{i,12} & \dots & -\tilde{s}_{i,1p} \\ \tilde{s}_{i,12} & 0 & \dots & -\tilde{s}_{i,2p} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{s}_{i,1p} & \tilde{s}_{i,2p} & \dots & 0 \end{pmatrix} \begin{pmatrix} \lambda_{1,t} & 0 & \dots & 0 \\ 0 & \lambda_{2,t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_{p,t} \end{pmatrix} \times$$
$$\begin{pmatrix} 0 & \tilde{s}_{j,12} & \dots & \tilde{s}_{j,1p} \\ -\tilde{s}_{j,12} & 0 & \dots & \tilde{s}_{j,2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{s}_{j,1p} & -\tilde{s}_{j,2p} & \dots & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_{1,t}} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_{2,t}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\lambda_{p,t}} \end{pmatrix},$$

which has the trace,

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

which is bounded in light of (A.25). We conclude that (A.32) 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

$$\sum_{i=1}^{d_\theta} \gamma_i \text{vec}\left(\frac{\partial \Omega_t^{\star}}{\partial \theta_i}\right) = 0_{p^2\times 1} \quad \text{a.s.}, \tag{A.34}$$

where we have suppressed the dependence on θ_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

$$\sum_{i=1}^{11} \gamma_i \text{vec}\left(\frac{\partial \Omega_t^{\star}}{\partial \theta_i}\right) = 0_4 \quad \text{a.s.} \tag{A.35}$$

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 $i = 1, \ldots, 10$,

$$\frac{\partial \Omega_t^\star}{\partial \theta_i} = V \frac{\partial \Lambda_t^\star}{\partial \theta_i} V' = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i} & 0 \\ 0 & \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i} \cos^2\phi + \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} \sin^2\phi & (\frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} - \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}) \cos\phi \sin\phi \\ (\frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} - \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i}) \cos\phi \sin\phi & \frac{\partial \lambda_{2,t}^\star}{\partial \theta_i} \cos^2\phi + \frac{\partial \lambda_{1,t}^\star}{\partial \theta_i} \sin^2\phi \end{pmatrix}, \tag{A.36}$$

and for $i = 11$

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

where

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

and

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

Hence, the first row of (A.35) 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 3.4 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

$$
\cos^2 \phi (\alpha_{12} - \alpha_{11}) + \sin^2 \phi (\alpha_{22} - \alpha_{21}) = 0. \tag{A.37}
$$

The same reasoning applied to the second and third rows of (A.35) yields that

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

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

$$
(\alpha_{22} - \alpha_{21}) - (\alpha_{12} - \alpha_{11}) = 0 \Leftrightarrow (\alpha_{22} - \alpha_{21}) = (\alpha_{12} - \alpha_{11}). \tag{A.38}
$$

Combining (A.37) and (A.38), we have that $\alpha_{12} = \alpha_{11}$ and $\alpha_{22} = \alpha_{21}$, which is ruled out by Assumption 3.9, and we conclude that (A.35) only holds whenever $\gamma_{11} = 0$. Hence

(A.35) 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 $\text{vec}(\partial \Lambda_t^\star / \partial \theta)_i$, $i = 1, \ldots, 10$, 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 γ such that (A.35) holds. We conclude that J is invertible.

Lemma A.6. *With $l_t^\star(\theta)$ defined in (A.1), suppose that Assumptions 3.1-3.2,3.4-3.9, hold. Then there exists a neighborhood around θ_0, $N(\theta_0) \subset \Theta$, such that*

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

Proof of Lemma A.6: 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 particular, with $[B^{j-1}]_s$ the sth row of $[B^{j-1}]$,

$$\underline{\omega} = \min_{r=1,\ldots,p} \inf_{\theta \in N(\theta_0)} \omega_r > 0, \tag{A.39}$$

$$\tilde{\omega} = \min_{s=1,\ldots,p} \inf_{\theta \in N(\theta_0)} \sum_{j=1}^{\infty} [B^{j-1}]_s W > 0. \tag{A.40}$$

$$\underline{\alpha} = \min_{r,s=1,\ldots,p} \inf_{\theta \in N(\theta_0)} A_{rs} > 0, \tag{A.41}$$

$$\overline{\alpha} = \max_{r,s=1,\ldots,p} \sup_{\theta \in N(\theta_0)} A_{rs} > 0. \tag{A.42}$$

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.26) with respect to some parameter θ_k:

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

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,

$$
\begin{aligned}
&\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}\right)\\
&= -2\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\right)\\
&+ \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^\star(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,k}^\star(\theta)\right) + \mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,j,k}^\star(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^\star(\theta)\right).
\end{aligned}
\tag{A.43}
$$

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.43) belongs to $\mathcal{L}_{N(\theta_0)}$. The same argument applies to the other terms in (A.43).

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 $N(\theta_0)$.

$$\begin{aligned}&\frac{\partial}{\partial\theta_k}\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)=\\&-3\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\&\quad+\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,j,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\&+\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\ddot{\Lambda}_{t,i,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)Y_t'(\theta)\right)\\&\quad+2\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)Y_t(\theta)\dot{Y}_{t,k}'(\theta)\right).\end{aligned}\tag{A.44}$$

Note that the first term in (A.44) we may use that $Y_t(\theta)=V(\phi)'X_t$, 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

$$\begin{aligned}&\mathrm{tr}\left(\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)V(\phi)'V\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}V'V(\phi)\right)=\\&\mathrm{vec}(V(\phi)'V)'(\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}\otimes\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,j}^{\star}(\theta)\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}(\theta)\Lambda_t^{\star-1}(\theta))\\&\times\mathrm{vec}(V'V(\phi)).\end{aligned}\tag{A.45}$$

Since $\mathrm{vec}(V(\phi)'V)$ consist of rotations based on trigonometric functions, it is bounded on $N(\theta_0)$. Next, note that the quantity $\Lambda_t^{\star 1/2}\eta_t\eta_t'\Lambda_t^{\star 1/2}\otimes\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,k}^{\star}\Lambda_t^{\star-1}(\theta)$ $\dot{\Lambda}_{t,j}^{\star}\Lambda_t^{\star-1}(\theta)\dot{\Lambda}_{t,i}^{\star}\Lambda_t^{\star-1}$ entering (A.45) is a symmetric $p^2\times p^2$ matrix, with $p\times p$ blocks, $Q_{g,h}$, $g,h=1,\dots,p$, each of which are diagonal,

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

for $s=1,\dots,p$, where $\dot{\lambda}_{s,t,i}^{\star}(\theta)\dot{\lambda}_{s,t,j}^{\star}(\theta)\dot{\lambda}_{s,t,k}^{\star}(\theta)/\lambda_{s,t}^{\star 3}(\theta)$ has finite rth moment for any $r>0$ by Lemma A.7. 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 θ_0 and θ 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 3.8, Lemma A.7, 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

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

which belongs to $\mathcal{L}_{N(\theta_0)}$, applying the same arguments as for (A.45).

Term #4 The derivative is,

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

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,

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

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 $\bar{S}_{i,j}(\theta) := V(\phi)'\left(\dot{S}_{i,j}(\theta)+S_i(\theta)S_j(\theta)\right)V(\phi)$

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

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

Term #9 Note that

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

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

Lemma A.7. *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 3.1-3.2,3.4-3.9,, 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 θ_0 such that

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

Proof of Lemma A.7: We start out by considering the first-order derivatives $\dot{\lambda}_{h,t,i}^\star(\theta)/\lambda_{h,t}^\star(\theta)$. With $Y_t = V(\phi)'X_t$, and suppressing the dependence on θ,

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

which has derivatives

$$\begin{aligned}
\frac{\partial \lambda_t^\star}{\partial \omega_i} &= \sum_{j=1}^{\infty}\dot{C}_{1,i}^{(j-1)}, & \dot{C}_{1,i}^{(j-1)} &= B^{j-1}\frac{\partial W}{\partial \omega_i},\\
\frac{\partial \lambda_t^\star}{\partial \alpha_i} &= \sum_{j=1}^{\infty}\dot{C}_{2,i}^{(j-1)}Y_{t-j}^{\odot 2}, & \dot{C}_{2,i}^{(j-1)} &= B^{j-1}\frac{\partial A}{\partial \alpha_i},\\
\frac{\partial \lambda_t^\star}{\partial \beta_i} &= \sum_{j=1}^{\infty}\dot{C}_{3,i}^{(j-1)}(W + AY_{t-j}^{\odot 2}), & \dot{C}_{3,i}^{(j-1)} &= \frac{\partial B^{j-1}}{\partial \beta_i} = \sum_{k=1}^{j-1}B^{k-1}\frac{\partial B}{\partial \beta_i}B^{j-1-k},\\
\frac{\partial \lambda_t^\star}{\partial \phi_i} &= 2\sum_{j=1}^{\infty}C_2^{(j-1)}(Y_{t-j}\odot \tilde{S}_iY_{t-j}), & & \qquad\qquad (\text{A.46})
\end{aligned}$$

where ω_i, α_i, β_i, ϕ_i denote arbitrary entries of, respectively, W, A, B, ϕ, and where $\tilde{S}_i$ is defined in (A.28).

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 ϕ_i in (i)–(iv) below.

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

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

where we have used that $\lambda^\star_{s,t} \geq \underline{\omega}$, with $\underline{\omega} > 0$ defined in (A.39), and $\sup_{\theta\in\Theta} \rho(B) < 1$.

(ii) Next, consider $\theta_i = \alpha_i$. Since $\partial \lambda^\star_t/\partial \alpha_i = \sum_{j=1}^\infty \dot{C}_{2,i}^{(j-1)} Y^{\odot 2}_{t-j}$, with $\dot{C}_{2,i}^{(j-1)} = B^{j-1}\partial A/\partial \alpha_i$. Here $\partial A/\partial \alpha_i$ is a matrix of zeros except for a 1 in the place of α_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 $s = 1, \ldots, p$,

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

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

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

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

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

which elementwise corresponds to,

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

Recall furthermore that,

$$\lambda^\star_{s,t} \geq \underline{\omega} + \sum_{h=1}^p [B^j]_{s,h} [\bar{C}_{t-j}]_h.$$

Lastly, we use the inequality $x/(1+x) \leq x^k$ for all $x \geq 0$ and $k \in (0,1)$, such that,

$$\beta_i \frac{\partial \lambda^\star_{s,t}/\partial \beta_i}{\lambda^\star_{s,t}} \leq \frac{\sum_{j=1}^{\infty} j \sum_{h=1}^{p} [B^j]_{s,h}[\bar{C}_{t-j}]_s}{\underline{\omega} + \sum_{h=1}^{p} [B^j]_{s,h}[\bar{C}_{t-j}]_s} \leq \sum_{j=1}^{\infty}\sum_{h=1}^{p} j \left(\frac{[B^j]_{sh}[\bar{C}_{t-j}]_h}{\underline{\omega}} \right)^k$$

$$= \sum_{j=1}^{\infty}\sum_{h=1}^{p} j [B^j]_{sh}^k \left(\frac{[\bar{C}_{t-j}]_h}{\underline{\omega}} \right)^k \leq K \sum_{j=1}^{\infty} j \varrho^j \sum_{h=1}^{p} \left(\frac{[\bar{C}_{t-j}]_h}{\underline{\omega}} \right)^k$$

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

(iv) Finally, consider $\theta_i = \phi_i$. The partial derivative $\partial \lambda^\star_t/\partial \phi_i$ in (A.46) contains the matrix product $\tilde{S}_i Y_{t-n}$, where the jth 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,

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

where ι_p is a p-dimensional column vector of ones, and $\overline{\alpha} > 0$ is defined in (A.42). Then, with $[B^{j-1}]_s$ the sth row of $[B^{j-1}]$, $\sum_{h=1}^{p} [C_2^{(j-1)}]_{s,h} \leq p\overline{\alpha}[B^{j-1}]_s \iota_p$, and we have that

$$\left| \frac{\partial \lambda^\star_{s,t}}{\partial \phi_i} \right| \leq K p^2 \overline{\alpha} \sum_{j=1}^{\infty} [B^{j-1}]_s \iota_p \|Y_{t-j}\|^2. \tag{A.47}$$

Moreover, with $\underline{\alpha} > 0$ defined in (A.41), we have that $[C_2^{(j-1)}]_{s,h} \geq \underline{\alpha}[B^{j-1}]_s\iota_p$ for $h, s = 1, \ldots, p$. Hence for any $j \geq 1$, and $s = 1, \ldots, p$,

$$\begin{aligned}\lambda^\star_{s,t} &= \sum_{j=1}^{\infty}[B^{j-1}]_s W + \sum_{j=1}^{\infty}\sum_{h=1}^{p}[C_2^{(j-1)}]_{s,h}y^2_{h,t-j} \geq \tilde{\omega} + \sum_{j=1}^{\infty}\sum_{h=1}^{p}\underline{\alpha}[B^{j-1}]_s\iota_p y^2_{h,t-j} \\ &= \tilde{\omega} + \underline{\alpha}\sum_{j=1}^{\infty}[B^{j-1}]_s\iota_p\|Y_{t-j}\|^2 \geq \tilde{\omega} + \underline{\alpha}[B^{j-1}]_s\iota_p\|Y_{t-j}\|^2, \end{aligned} \tag{A.48}$$

where $\tilde{\omega} > 0$ is defined in (A.40). Combining (A.47) and (A.48), we have that for $s = 1, \ldots, p$ and $k \in (0, 1)$

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

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 H_2^* in (4.2) 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 H_2^* in (4.2) that $\theta = (\tau, \delta) \in \Theta_\tau \times \Theta_\delta$, where $\delta = (\beta_{11}, \beta_{21}, \beta_{31})'$ denotes the unidentified parameters, while $\tau \in \Theta_\tau$ denotes the remaining $d_\tau = 21$ parameters. As in Appendix A.1, consider the stationary and ergodic version of the log-quasi-likelihood contributions given by,

$$
\begin{aligned}
l_t^{\star}(\tau,\delta) &= \log\det(\Omega_t^{\star}(\tau,\delta)) + X_t'\Omega_t^{\star -1}(\tau,\delta)X_t,\\
\Omega_t^{\star}(\tau,\delta) &= V(\phi)\Lambda_t^{\star}(\tau,\delta)V(\phi)', \quad \Lambda_t^{\star}(\tau,\delta) = \mathrm{diag}(\lambda_t^{\star}(\tau,\delta)),\\
\lambda_t^{\star}(\tau,\delta) &= W + A(V(\phi)'X_{t-1})^{\odot 2} + B\lambda_{t-1}^{\star}(\tau,\delta).
\end{aligned}
$$

The limiting distribution of the supLR statistic in (4.4) 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 $\tilde{\tau}_{T,\delta}$ and $\hat{\tau}_{T,\delta}$ defined in (4.3), assume that $\tilde{\tau}_{T,\delta}, \hat{\tau}_{T,\delta} \xrightarrow{p} \theta_0$.

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

$$
\Sigma_{\delta_1\delta_2} = E\left[\frac{\partial l_t^{\star}(\tau_0,\delta_1)}{\partial\tau}\frac{\partial l_t^{\star}(\tau_0,\delta_2)}{\partial\tau'}\right], \quad \text{for } \delta_1,\delta_2\in\Theta_\delta. \tag{B.1}
$$

(iii) For any $\delta\in\Theta_\delta$, $T^{-1}\partial^2 l_t^{\star}(\tau_0,\delta)/\partial\tau\partial\tau' \xrightarrow{p} J_\delta$, where

$$
J_\delta = E\left(\frac{\partial^2 l_t^{\star}(\tau_0,\delta)}{\partial\tau\partial\tau'}\right), \tag{B.2}
$$

with J_δ invertible uniformly on Θ_δ.

(iv) The sets $\Theta_\tau - \tau_0$ and $\Theta_\tau^* - \tau_0$ are locally equal to some convex cones C and C^*, respectively.[1]

(v) There exists a neighborhood $N(\tau_0)$ of τ_0 such that

$$
\sup_{\delta\in\Theta_\delta}\left\|T^{-1/2}\sum_{t=1}^{T}\left(\frac{\partial l_t(\tau_0,\delta)}{\partial\tau} - \frac{\partial l_t^{\star}(\tau_0,\delta)}{\partial\tau}\right)\right\| \xrightarrow{p} 0,
$$

and

$$
\sup_{\delta\in\Theta_\delta,\tau\in N(\tau_0)\cap\Theta_\tau}\left\|T^{-1}\sum_{t=1}^{T}\left(\frac{\partial^2 l_t(\tau,\delta)}{\partial\tau\partial\tau'} - \frac{\partial^2 l_t^{\star}(\tau,\delta)}{\partial\tau\partial\tau'}\right)\right\| \xrightarrow{p} 0.
$$

(vi) For any fixed $\delta\in\Theta_\delta$, and any deterministic scalar sequence $(\varepsilon_T : T = 1,2,...)$ with $\varepsilon_T \to 0$,

$$
\sup_{\tau\in\Theta_\tau:\|\tau-\tau_0\|\le\varepsilon_T}\left\|T^{-1}\sum_{t=1}^{T}\left(\frac{\partial^2 l_t^{\star}(\tau,\delta)}{\partial\tau\partial\tau'} - \frac{\partial^2 l_t^{\star}(\tau_0,\delta)}{\partial\tau\partial\tau'}\right)\right\| \xrightarrow{p} 0.
$$

[1] With $\Theta\subset\mathbb{R}^{\dim\theta}$ and $\theta_0\in\Theta$, 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 2ε.

By Andrews (2001, Theorem 4), under conditions (i)-(vi) and H_2^*,

$$\sup \mathrm{LR}_T(H_2^*) \xrightarrow{d} \sup_{\delta\in\Theta_\delta} \{\lambda_\delta' J_\delta \lambda_\delta\} - \sup_{\delta\in\Theta_\delta} \{\lambda_\delta^{*\prime} J_\delta \lambda_\delta^*\}, \tag{B.3}$$

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.3) depends on the cones C and C^*, and hence implicitly on the location of the nuisance parameters, see e.g., Cavaliere *et al.* (2020) 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 τ as

$$\tau = (\tau_1', \tau_2')',$$

with $\tau_1 = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \beta_{11}, \beta_{12})'$ of dimension $d_{\tau_1} = 5$, and with τ_2 containing the remaining $d_{\tau_2} = 16$ (nuisance) parameters in W, A and B.

(vii) Assume that $\tau_{2,0} \in \mathrm{int}\Theta_{\tau_2}$ and $\Theta_\tau = \Theta_{\tau_1} \times \Theta_{\tau_2}$, with $\tau_1 \in \Theta_{\tau_1}$ and $\tau_2 \in \Theta_{\tau_2}$.

Under the additional assumptions in (vii), $C = \mathbb{R}_+^{d_{\tau_1}} \times \mathbb{R}^{d_{\tau_2}}$ and $C^* = \{0_{d_{\tau_1}}\} \times \mathbb{R}^{d_{\tau_2}}$, which implies that

$$\sup \mathrm{LR}_T(H_2^*) \xrightarrow{d} \sup_{\delta\in\Theta_\delta} \left\{\lambda_\delta' \left(K J_\delta^{-1} K'\right)^{-1} \lambda_\delta\right\}, \tag{B.4}$$

where K is given by $K\tau = \tau_1$ and

$$\lambda_\delta = \arg \inf_{\eta\in\mathbb{R}_+^{d_{\tau_1}}} \left\{(\eta - Z_\delta)' \left(K J_\delta^{-1} K'\right)^{-1} (\eta - Z_\delta)\right\}. \tag{B.5}$$

and with $Z_\delta = K J_\delta^{-1} G_\delta$ such that Z_δ is a d_{τ_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, δ is $d_\delta = 3$ dimensional and we choose k different values for each entry of δ, such that we have a discrete grid Δ with $d_\Delta = k^{d_\delta}$ different values of δ.

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

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

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

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

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_{θ_1} dimensional λ_{δ_i} by solving the constrained minimization problem in (B.5), with Z_δ and J_δ replaced with Z_{δ_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.* (2020), we apply a restricted recursive bootstrap to obtain critical values for the likelihood ratio statistic, $LR_T(H_2)$, 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 H_2. That is, the likelihood function in (3.8) is maximized with $A = \gamma\alpha'$ and $B = \gamma\beta'$ where the (3×2) matrices γ, α and β have non-negative entries. With $\tilde{\theta}_T$ denoting the obtained restricted estimator, for $t = 1, ..., T$ compute the centered and standardized residuals,

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

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

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

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

$$X_t^* = V(\tilde{\phi}_T)\Lambda_t^{*1/2}(\tilde{\theta}_T)\eta_t^*, \qquad \Lambda_t^*(\tilde{\theta}_T) = \text{diag}(\lambda_t^*(\tilde{\theta}_T))$$
$$\lambda_t^*(\tilde{\theta}_T) = W(\tilde{\theta}_T) + A(\tilde{\theta}_T)(V(\tilde{\phi}_T)'X_{t-1}^*)^{\odot 2} + B(\tilde{\theta}_T)\lambda_{t-1}^*(\tilde{\theta}_T),$$

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

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

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

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

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,

$$LR_T^*(H_2) = 2(L_T^*(\hat{\theta}_T^*) - L_T^*(\tilde{\theta}_T^*)).$$

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 10. *Note that the bootstrap distribution approximates the $LR_T(H_2)$ statistic for the case where, under 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. *(2020).*

D Monte Carlo Simulations

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 ϕ being restricted to the interval $[\phi_L, \phi_U] = [0, \pi/2]$, which is sufficient for identification. The simulations indeed indicate that the conditions of finite second order moments and the restrictions on ϕ are not necessary.

D.1 Case (i): Sufficient Conditions for Asymptotic Normality Satisfied

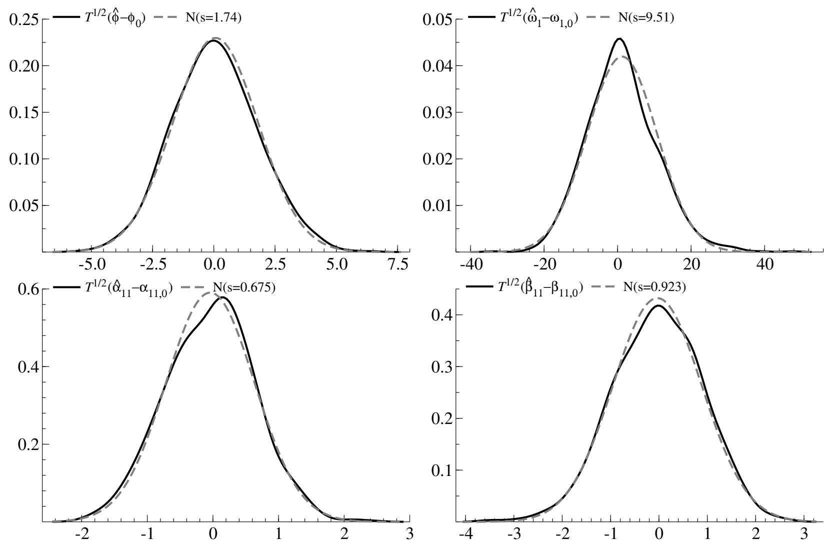


Figure D.1: Monte Carlo-based densities of estimated parameters $\hat{\theta}_T$ for the finite second-order moment case, $\rho\,(A + B) < 1$.

In Case (i), the bivariate λ-GARCH model is considered, where

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

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

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

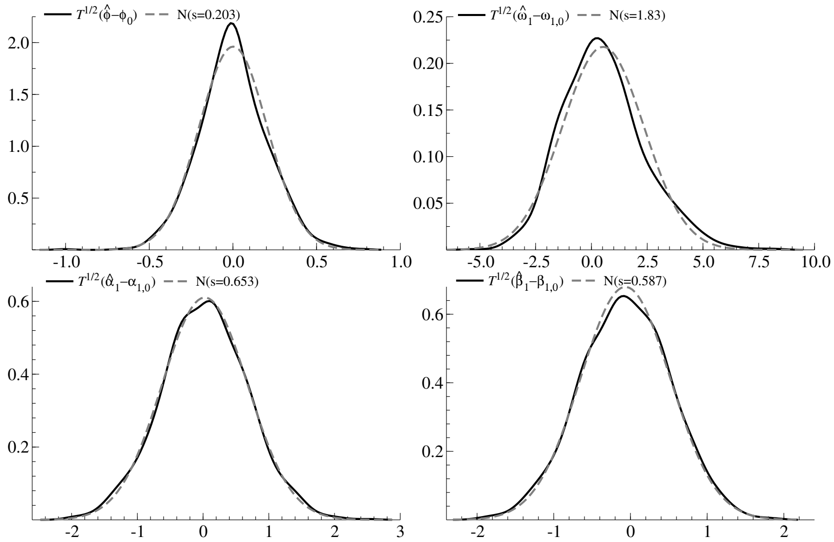


FIGURE D.2: Monte Carlo based densities of $\hat{\theta}_T$ for the "integrated" case of $\rho(A+B) = 1$.

satisfied.

We simulate $N = 1000$ realizations the process with $T = 10000$ observations, and estimate ϕ, W, A, B by QMLE. Figure D.1 contains kernel density estimates of the centered and scaled estimates of ϕ, ω_1, α_{11}, and β_{11}. The solid line is the estimated density, and the dashed line is the normal density. As expected Figure D.1 confirms asymptotic normality.

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

Consider again the model in (D.1) with A and B diagonal. For the dgp ϕ_0 is as before, $W_0 = (0.1, 0.1)'$

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

such that $\rho(A_0+B_0) = 1$. Hence, by definition, the stationary solution does not have finite second-order moments which violates the sufficient condition in Theorem 3.3. Figure D.2 contains kernel density estimates of the centered and scaled estimates of ϕ, ω_1, α_{11}, and β_{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 ϕ

Consider here the trivariate λ-GARCH,

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

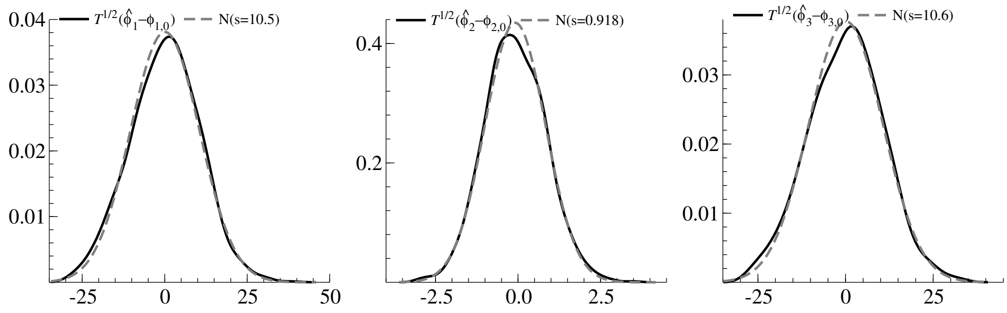


FIGURE D.3: Monte Carlo-based densities of estimated parameters $\hat{\theta}_T$ when $\phi_L = -\pi/2$ and $\phi_U = \pi/2$ in Θ_ϕ.

with $B = 0_{3\times3}$ 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

$$\phi_0 = \begin{pmatrix} 0.47 \\ 1.45 \\ -1.30 \end{pmatrix}, \quad W_0 = \begin{pmatrix} 0.45 \\ 1.50 \\ 0.95 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0.25 & 0.05 & 0.09 \\ 0.03 & 0.35 & 0.06 \\ 0.07 & 0.12 & 0.3 \end{pmatrix}, \quad B_0 = 0_{3\times3},$$

such that $\phi_{0,3} \notin [0, \pi/2]$. Figure D.3 contains standardized densities of $\hat{\phi}_1$, $\hat{\phi}_2$, and $\hat{\phi}_3$. Whereas, Lemma 3.1 restricts ϕ_i to be in the interval $[0, \pi/2]$ in order to ensure identification, we have that Figure D.3 indicates that the condition can be relaxed, as the densities seem to be centered around zero.

Chapter 2

Spectral Targeting Estimation of Dynamic Conditional Eigenvalue GARCH Models

Abstract

This paper investigates a two-step estimator of a class of orthogonal GARCH models, combining (eigenvalue and -vector) targeting estimation with stepwise (univariate) estimation. We denote this estimator the "spectral targeting estimator". This type of estimator has long been used in empirical modeling, and in this paper we present novel asymptotic theory. We find that the estimator is consistent under finite second order moments, while asymptotic normality holds under finite fourth order moments. The estimator is especially well suited for modeling larger portfolios: we compare the empirical performance of the spectral targeting estimator to that of the quasi-maximum likelihood estimator for five portfolios of 25 assets. The spectral targeting estimator dominates in terms of computational complexity, being up to 57 times faster in estimation, while both estimators produce similar out-of-sample forecasts, indicating that the spectral targeting estimator is well suited for high-dimensional empirical applications.

Keywords: Asymptotic theory, Multivariate GARCH, Variance targeting, Two-step estimation.
JEL: C32, C58.

1 Introduction

Multivariate conditionally heteroskedastic (MGARCH) models are a popular tool for risk management and dynamic portfolio allocation, where forecasts of conditional covariance matrices play an important role. As is well known in the literature, MGARCH models suffer from the "curse of dimensionality", making them difficult and time consuming to estimate for larger portfolios using quasi-maximum likelihood (QML) techniques. Many practitioners and academics alike have therefore preferred using alternative estimation methods: Two popular choices are the variance targeting (VT) estimator and the equation-by-equation (EbE) estimator, see e.g., Bauwens, Laurent, and Rombouts (2006).

In the context of orthogonal GARCH models, such as the Dynamic Conditional Eigenvalue GARCH (or λ-GARCH) model of Hetland, Pedersen, and Rahbek (2020), we can combine the idea behind the two methods in what we denote the spectral targeting estimator (STE): By estimating the unconditional eigenvalues and -vectors using a sample moment estimator, the remainder of the parameters of the GARCH model may be estimated univariately in a stepwise manner, in which we target the unconditional eigenvalues and -vectors. This estimation procedure dramatically reduces the computational complexity of the optimization problem and speeds up numerical estimation compared to the QML estimator.

In this paper, we derive large-sample properties (consistency and asymptotic normality) of this estimator under mild conditions. Our numerical illustrations show that the estimator is superior to the QML estimator in cross-sections larger than 10 financial assets, being up to 57 times faster in estimation, while the out-of-sample forecasts from the QML and ST estimator are similar in portfolios of 25 assets. Furthermore, because the second step of the estimator is based on univariate estimation, estimation of conditional covariance matrices of high-dimensional portfolios is feasible. In our numerical exercise, we estimate the model in dimensions up to $p = 500$ assets.

Two-step estimators such as the STE are well-known in the empirical multivariate GARCH literature, and the STE has been applied as early as Alexander and Chibumba (1997). Related multi-step estimators are discussed in e.g., Fan, Wang, and Yao (2008) and Boswijk and Weide (2011) (see also references therein), with the common element that a first step estimator utilizes (un-)conditional information from the matrix of second order moments using a loss-function, and a second step in which univariate estimation based on the Gaussian log-likelihood function is used. While Fan, Wang, and Yao (2008) and Boswijk and Weide (2011) show consistency of their respective first step estimators, both papers come short of showing the joint asymptotic behavior of the two-step estimator. In contrast to the aforementioned papers, we derive consistency and asymptotic normality of the joint estimator.

In general, asymptotic theory of QML estimation of MGARCH models is well-understood (large-sample theory for the λ-GARCH is covered in Hetland, Pedersen, and Rahbek (2020), while chapter 11 of Francq and Zakoïan (2019) contain a review of existing theory for other model specifications), whereas less attention has been paid to the theory of alternative estimation methods. Large-sample properties of the two-step VT estimator are considered in Pedersen and Rahbek (2014) and Francq, Horvath, and Zakoïan (2014) for the BEKK model (Engle and Kroner, 1995) and the (extended) CCC model (Bollerslev, 1990 and Jeantheau, 1998) respectively, while Francq and Zakoïan (2016) consider the two-step EbEE for various MGARCH specifications. Both the VT and EbE estimators are two-step estimators, which are quite common in econometrics, see e.g., Newey and McFadden (1994). The EbEE and VTE both aim at making high(er) dimensional estimation feasible, and do so in two distinct ways: The EbEE estimates univariate volatility models in a first step, and subsequently a (conditional) correlation dynamic in a second step, whereas the VTE estimates the unconditional covariance matrix using a moment estimator (or vector of unconditional variances in the case of the CCC model), followed by a joint (profiled) estimation of the volatility and covariance dynamics. The STE is related to both, as we recover sample eigenvalues and -vectors from the unconditional covariance matrix, and estimate univariate dynamics for "rotated" (orthogonalized) returns in a second step. The resulting estimator is well-behaved and easily implemented: Because the λ-GARCH model is specified using the spectral decomposition, the (profiled) log-likelihood, conditional on the initial estimator, can be rewritten as a sum of orthogonal univariate log-likelihood functions, making stepwise estimation feasible. Furthermore, by recovering the (constant conditional) eigenvectors we avoid having to parameterize the eigenvectors under the restriction of orthonormality.

The remainder of the paper proceeds as follows: Section 2 introduces the λ-GARCH model and spectral targeting. Section 3 presents the two-step estimator and Section 4 presents novel asymptotic results and discuss practical considerations for implementation. Section 5 investigates the empirical properties of the estimator compared to the QML estimator. Finally, Section 6 concludes. All proofs are relegated to the appendices.

1.1 Notation

Some notation used throughout the paper. $\mathbb{R}$ denotes the real numbers, $\mathbb{R}_+$ the positive real numbers, $\mathbb{R}_{++}$ the strictly positive real numbers. The absolute value of $a \in \mathbb{R}$ is denoted $|a|$. For $p, n \in \mathbb{N}$, I_p denotes the $(p \times p)$ identity matrix and $0_{n\times p}$ denotes a $n \times p$ matrix of zeros. The vector vec(A) stacks the columns of the matrix A. For a $p-$dimensional vector x, $\text{diag}(x) = \text{diag}((x_i)_{i=1}^p)$ is a diagonal matrix with x on the diagonal. The trace of a square matrix is denoted $\text{tr}(A)$, and the determinant $\det(A)$. Furthermore, denote by $\rho(A)$ the spectral radius of any square matrix A, i.e., $\rho(A) =$

$\max\{|\tilde{\lambda}_i| : \tilde{\lambda}_i$ is an eigenvalue of A$\}$. We use $||\cdot||$ as a matrix norm. Let $\odot$ denote the Hadamard product, with $A^{\odot 2} = A \odot A$, and $A \otimes B$ denotes the Kronecker product of A and B of suitable dimensions, and note that $A^{\otimes 2} = A \otimes A$. Elements of matrices or vectors are denoted by lower case letters, e.g., a_{ij} is the (i,j)'th element of the matrix A. We use three kinds of convergence of random variables, $\stackrel{a.s.}{\to}$ denotes almost sure convergence, $\stackrel{p}{\to}$ denotes convergence in probability and $\stackrel{D}{\to}$ denotes convergence in distribution.

2 The λ-GARCH Model

As in Hetland, Pedersen, and Rahbek (2020), we focus on the class of orthogonal GARCH (O-GARCH) models originally introduced by Alexander and Chibumba (1997). The presented model has more general dynamics than the O-GARCH, allowing for eigenvalue-spillovers, and we denote this version of the model λ-GARCH.

Let X_t be a $p \times 1$ vector of asset returns,

$$X_t = H_t^{1/2} Z_t, \tag{2.1}$$

where $t = 1, \ldots, T$ and Z_t is an $iid(0, I_p)$ sequence of random variables. $H_t^{1/2} = V\Lambda_t^{1/2}$ is the (asymmetric) matrix square root of the conditional covariance matrix, H_t (following the literature on MGARCH models, see e.g., Weide (2002) and Lanne and Saikkonen (2007)), which is decomposed using the spectral theorem,

$$H_t = V\Lambda_t V'. \tag{2.2}$$

Here, $V = \begin{pmatrix} V_1 & V_2 & \ldots & V_p \end{pmatrix}$ is an orthonormal matrix of eigenvectors, $VV' = I_p$, and Λ_t is a diagonal matrix with time-varying eigenvalues, λ_t, on the diagonal,

$$\Lambda_t = \text{diag}(\lambda_t). \tag{2.3}$$

The $p \times 1$ vector of dynamic eigenvalues are assumed to follow a GARCH dynamic,

$$\lambda_t = W + AY_{t-1}^{\odot 2} + B\lambda_{t-1}, \tag{2.4}$$

where $Y_t = V'X_t$ are "rotated" (or orthogonalized) returns: The orthonormal matrix V rotates the returns X_t to be orthogonal with conditional covariance Λ_t. To ensure that the covariance matrix is positive definite for all $t \in \mathbb{Z}$, we restrict $w_i > 0$, $a_{ij} \geq 0$, and $b_{ij} \geq 0$ for $i, j = 1, \ldots p$. Furthermore, to facilitate stepwise estimation, we restrict B to be a diagonal matrix, letting the i'th lagged eigenvalue enter equation i. The restriction on B is crucial for making the step-wise estimation feasible: In the context of a non-restricted B matrix, $\lambda_{i,t}$ is a function of $\lambda_{j,t}$ for $i, j = 1, \ldots, p$, and it is not possible to estimate the model sequentially as $\lambda_{i,t}$ depends on conditional eigenvalues, $\lambda_{j,t}$, $j \neq i$, that (potentially) have not been estimated yet. This restriction is common in equation-by-equation estimation of GARCH models, see e.g., Francq, Horvath, and Zakoïan (2014).[1]

In this restricted case, the λ-GARCH is equivalent to the model in Fan, Wang, and Yao (2008) and closely related to the model in Lanne and Saikkonen (2007) and Boswijk and Weide (2011), see Section 2.1 of Hetland, Pedersen, and Rahbek (2020) for a discussion and comparison.

By Lemma C.1 and C.2 in Appendix C, the stochastic process $\{X_t\}_{t\in\mathbb{Z}}$ can be initiated from the invariant distribution such that it is covariance stationary if and only if $\rho(A+B) < 1$. If this is the case, the unconditional covariance matrix, $H = V(X_t) = E[X_t X_t']$, exists almost surely and is given by,

$$H = V\mathrm{diag}(\lambda)V', \tag{2.5}$$

$$\lambda = (I_p - A - B)^{-1}W, \tag{2.6}$$

where $\lambda = E[\lambda_t]$ is the vector of unconditional eigenvalues.

To obtain the spectral targeting λ-GARCH, we re-parameterize the model by substituting (2.6) into (2.4),

$$\lambda_t = (I_p - A - B)\lambda + AY_{t-1}^{\odot 2} + B\lambda_{t-1}. \tag{2.7}$$

This implies that the i'th rotated return is driven by an augmented GARCH$(1,1)$ with spill-overs from the other squared rotated returns,

$$y_{i,t} = \lambda_{i,t}^{1/2} z_{i,t}, \tag{2.8}$$

$$\lambda_{i,t} = w_i + \sum_{j=1}^{p} a_{ij} y_{j,t-1}^2 + b_i \lambda_{i,t-1}, \tag{2.9}$$

with $y_{i,t} = V_i' X_t$ and $w_i = (1-b_i)\lambda_i - \sum_{j=1}^{p} a_{ij}\lambda_j$ for $i = 1, \ldots, p$. This specification is motivated by generality: it seems restrictive to assume that the conditional variance of a component is not influenced by the past of other components, and allowing for spill-overs between assets may improve the model fit and out-of-sample performance.

3 Spectral Targeting Estimation

While theory for classical joint QMLE of λ-GARCH type models have been considered in Hetland, Pedersen, and Rahbek (2020), we consider spectral targeting estimation. The

[1] In principle one could allow B to be non-restricted. The cost however, is that all equations have to be estimated jointly in a second step, similar to the VTE of ECCC-GARCH or BEKK-GARCH models, see Pedersen and Rahbek (2014) and Francq, Horvath, and Zakoïan (2014).

stepwise estimation procedure examined in this paper makes estimation and inference for the λ-GARCH feasible in (very) large systems, as long as the time series dimension dominates the cross-sectional dimension (see e.g., Ledoit and Wolf, 2004, 2012).

Define $\upsilon = \text{vec}(V)$, i.e., the vector of stacked eigenvectors, such that

$$\gamma = [\lambda', \upsilon']' \text{ and } \kappa^{(i)} = [a_{i1}, \ldots, a_{ip}, b_i]', \tag{3.1}$$

where γ contain the eigenvalues and -vectors of the unconditional covariance matrix, H. Hence, γ denote "static" and $\kappa^{(i)}$ the "dynamic" parameters of equation i, such that $\theta^{(i)} = [\gamma', \kappa^{(i)\prime}]'$ is the vector of parameters associated with the i'th rotated return, $i = 1, \ldots, p$, of size $e = p^2 + 2p + 1$. Likewise, define the parameter space $\Theta^{(i)} := \mathcal{L} \times \mathcal{V} \times \mathcal{K}^{(i)} \subset \mathbb{R}^p_{++} \times \mathbb{R}^{p^2} \times \mathbb{R}^{p+1}_{+}$ which is restricted such that $\rho(A + B) < 1$ and λ is element-wise strictly positive and eigenvectors are orthonormal, $VV' = I_p$, such that H is positive definite and symmetric. To ensure $W = (I_p - A - B)\lambda > 0$ it is sufficient to impose the restriction $\sum_{j=1}^{p} a_{ij} + b_i < 1$.

The vector of all the parameters in the model is

$$\theta = [\gamma', \kappa^{(1)\prime}, \ldots, \kappa^{(p)\prime}]',$$

which has $p(p+1)/2 + p^2 + p$ elements. To emphasize the dependence on the parameters in $\theta^{(i)}$, we restate the model for the i'th rotated return as,

$$\begin{aligned} y_{i,t}(\gamma) &= \lambda_{i,t}(\gamma, \kappa^{(i)}) z_{i,t} \\ \lambda_{i,t}(\gamma, \kappa^{(i)}) &= w_i + \sum_{j=1}^{p} a_{ij} y^2_{j,t-1}(\gamma) + b_i \lambda_{i,t-1}(\gamma, \kappa^{(i)}), \end{aligned}$$

which also explicitly states that the conditional eigenvalues are a non-linear function of the eigenvectors in γ, and are linear in the dynamic parameters in $\kappa^{(i)}$. Furthermore, $H_t(\theta) = V\Lambda_t(\theta)V'$, such that the (constant conditional) eigenvectors only depend on γ, whereas the diagonal matrix of conditional eigenvalues depend on the full vector of parameters, θ.

The STE consists of two steps: In the first step, we estimate γ using a sample estimator. In the second step, the dynamic parameters of the model are estimated by univariate QMLE for each equation in (2.8)-(2.9) for $i = 1, \ldots, p$. This procedure yields the STE for equation i, denoted $\theta^{(i)}$, and based on the joint vector of parameters, θ, the sequence of filtrated conditional covariance matrices, $H_t(\theta)$, can be recovered for $t = 1, \ldots, T$.

3.1 The moment estimator

The first step of the STE utilizes the (strong) law of large numbers for strictly stationary and ergodic processes, and we estimate H by the sample covariance matrix,

$$\mathrm{vec}(\hat{H}) = \mathrm{vec}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right). \tag{3.2}$$

If X_t is covariance stationary and ergodic, $\hat{H}$ is a strongly consistent estimator for H by the ergodic theorem. From $\hat{H}$ it is possible to recover the estimated eigenvalues, $\hat{\lambda}$, and estimated eigenvectors, $\hat{V}$, by solving the two equations,

$$|H - \lambda I_p| = 0, \tag{3.3}$$

$$HV_i = \lambda_i V_i, \qquad i = 1, \ldots, p, \tag{3.4}$$

and under Assumption 4.1 below, $\hat{\lambda}$ and $\hat{\upsilon}$ are strongly consistent estimators of λ and υ respectively by the continuous mapping theorem.

In applications, these two equations are solved using iterative procedures and for H symmetric and positive definite, all eigenvalues are almost surely strictly positive. Notice however, that the eigenvalue decomposition is not unique: the spectrum of H is unique only up to the ordering, and while the eigenspace of H is unique the eigenvectors are not. Furthermore, eigenvalues may not be unique. We discuss this further in the next subsection (see Assumption 4.2 and Remark 4.1).

Remark 3.1 (Alternative first step estimator)**.** *Instead of estimating the eigenvalues and -vectors implicitly using the moment estimator of H, we can estimate them directly using an approach similar to that proposed by Fan, Wang, and Yao (2008) (see also Boswijk and Weide (2011)), wherein V is specified using rotation matrices,*

$$V(\phi) = \prod_{1 \leq i < j \leq p} U_{ij}(\phi_{ij}),$$

with $U_{ij}(\phi_{ij})$ a p-dimensional identity matrix apart from four elements: (i,i) and (j,j) are $\cos(\phi_{ij})$, (i,j) *and* (j,i) *are* $\sin(\phi_{ij})$ *and* $-\sin(\phi_{ij})$ *respectively. ϕ is a $p(p-1)/2$ vector containing the rotation parameters, ϕ_{ij}. This parameterization ensures that $V(\phi)V'(\phi) = I_p$. The eigenvectors and eigenvalues can then be estimated by numerically solving the minimization problem,*

$$\underset{[\phi', \lambda']' \in \mathcal{C}}{\arg\min} \quad C_T(\phi, \lambda)$$

where $\mathcal{C}$ is an appropriate parameter space and $C_T(\phi, \lambda)$ is a cost function, e.g., the

Gaussian log-likelihood,

$$C_T(\phi,\lambda) = \frac{1}{T}\sum_{t=1}^{T}\left(\log\det(\Lambda) + X_t'V(\phi)\Lambda^{-1}V'(\phi)X_t\right).$$

The asymptotic theory for this estimator can be derived with relative ease, see e.g., Hetland, Pedersen, and Rahbek (2020) who parameterize the joint QMLE of the λ*-GARCH in a similar fashion. One should however, keep in mind that the rotation parameters in* ϕ *are not uniquely identified unless we impose restrictions on the parameter space. A sufficient condition is* $\phi_{ij} \in (0, \pi/2)$*, see Lemma 1 of Hetland, Pedersen, and Rahbek (2020).*

The alternative first step estimator outlined in Remark 3.1 requires numerical optimization of a cost function, and may therefore run into numerical problems as p increase, such as failure of a Newton-type optimization procedure to converge, or the possibility of ending up in a local maximum – problems similar to those of the joint QML estimator. We therefore choose to work with the sample moment estimator as it has a closed form solution and is the preferred first step estimator in the variance targeting literature.

3.2 The profiled maximum likelihood estimator

In the second step of the STE, we consider the profiled quasi log-likelihood function based on the multivariate Gaussian distribution. Conditional on a fixed X_0 and H_0, the joint Gaussian log-likelihood of the model is, up to a constant,

$$\begin{aligned} L_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\log\det(H_t(\theta)) + X_t'H_t^{-1}(\theta)X_t \\ &= \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p}\left(\log(\lambda_{i,t}(\gamma,\kappa^{(i)})) + \frac{y_{i,t}^2(\gamma)}{\lambda_{i,t}(\gamma,\kappa^{(i)})}\right) \\ &= \sum_{i=1}^{p} L_T^{(i)}(\gamma,\kappa^{(i)}), \end{aligned} \tag{3.5}$$

using $H_t^{-1}(\theta) = V\Lambda_t^{-1}(\theta)V'$, $\log\det(H_t(\theta)) = \sum_{i=1}^{p}\log(\lambda_{i,t}(\gamma,\kappa^{(i)}))$, and $Y_t(\gamma) = V'X_t$. That is, because the rotated returns are orthogonal, the log-likelihood function can be decomposed as the sum of p univariate log-likelihood functions, each of which depend on $\theta^{(i)} = [\gamma', \kappa^{(i)\prime}]'$,

$$L_T^{(i)}(\gamma,\kappa^{(i)}) = \frac{1}{T}\sum_{t=1}^{T} l_t^{(i)}(\gamma,\kappa^{(i)}), \tag{3.6}$$

$$l_t^{(i)}(\gamma,\kappa^{(i)}) = \log(\lambda_{i,t}(\gamma,\kappa^{(i)})) + \frac{y_{i,t}^2(\gamma)}{\lambda_{i,t}(\gamma,\kappa^{(i)})}, \tag{3.7}$$

where $y_{i,t}(\gamma) = V_i' X_t$ and $\lambda_{i,t}(\gamma, \kappa^{(i)})$ is given in (2.9). Conditional on γ, each of the i univariate log-likelihood functions are orthogonal and do not depend on $\kappa^{(j)}$ for $j \neq i$. The parameters of the model can therefore be estimated sequentially, and we define the STE of $\kappa^{(i)}$ as,

$$\hat{\kappa}^{(i)} = \underset{\kappa^{(i)} \in \mathcal{K}^{(i)}}{\arg\min}\ L_T^{(i)}(\hat{\gamma}, \kappa^{(i)}), \tag{3.8}$$

and the two-step procedure yields the STE of θ,

$$\hat{\theta} = [\hat{\gamma}', \hat{\kappa}^{(1)\prime}, \ldots, \hat{\kappa}^{(p)\prime}]'.$$

Similar to quasi-maximum likelihood estimation of multivariate GARCH models, we use the Gaussian log-likelihood function, but we do not assume that the vector of innovations Z_t are Gaussian, only that they are centered with unit variance: Even if the innovations are drawn from a different distribution, the results in Theorem 4.1 and 4.2 below still hold, as long as the assumptions are satisfied.

Compared to joint QMLE, which estimates all $\frac{3}{2}(p^2 + p)$ parameters jointly, the STE procedure vastly reduces the number of parameters estimated in each step: In the first step $p(p+1)/2$ parameters are estimated by method of moments and in the second step $p+1$ parameters are estimated for each rotated return, making the estimation procedure suitable in high-dimensional systems and less vulnerable to numerical problems.

4 Large-Sample Properties of Spectral Targeting Estimation

In this section we establish consistency and asymptotic normality of the STE and discuss practical considerations for implementation. A novelty of the asymptotic theory presented here is that we parameterize the moment estimator in terms of the unconditional eigenvalues and vectors, rather than the vectorized covariance matrix. In doing so, we apply the mean-value theorem on the eigenvectors, which otherwise do not have a closed form solution as a function of the unconditional covariance matrix. This, in conjunction with the continuous mapping theorem, allows us to study the asymptotic behavior of both the first step estimator, γ, and the joint parameter vector of the i'th rotated return, $\theta^{(i)}$.

The two-step estimator is consistent under finite second order moments, and it has a limiting Gaussian distribution under the assumption of finite fourth order moments. Both of these moment conditions stem from the first step moment estimator, and are more strict that the moment conditions for the joint QML estimator (for which we need $E||X_t||^{2+\delta} < \infty$, $\delta > 0$, see Theorem 3.3 in Hetland, Pedersen, and Rahbek (2020)). These results are novel and extend the existing literature on targeting and stepwise estimation, see e.g., Francq, Horvath, and Zakoïan (2014), Pedersen and Rahbek (2014) and Francq and Zakoïan (2016). All proofs are relegated to Appendix A.

Before discussing the asymptotic properties in detail, we make the following assumptions. First, we assume that the process is covariance stationary and ergodic.

Assumption 4.1. *The process $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary, ergodic and has finite second order moments.*

Assumption 4.1 is in line with the literature for variance-targeting estimation in the univariate and multivariate case, see e.g., Pedersen and Rahbek (2014) or Francq, Horvath, and Zakoïan (2011, 2014), and is needed to ensure that the moment estimator converge to a well-defined unconditional covariance matrix for $T \to \infty$.

Furthermore, we need the following assumption on identification of the first step estimator.

Assumption 4.2. *The characteristic polynomial of the unconditional covariance matrix, H, has an algebraic multiplicity of 1. Furthermore, the eigenvalues are sorted from smallest to largest, $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$ and each eigenvector is normalized such that the first non-zero element of V_i is positive for $i = 1, \ldots, p$.*

Assumption 4.2 is novel in the (variance) targeting literature and is needed to ensure identification of the first step estimator: We assume that the unconditional eigenvalues are simple, i.e., that the characteristic polynomial of the unconditional covariance matrix has an algebraic multiplicity of one. This is needed for two reasons: First, in the case of repeated eigenvalues, the associated eigenvectors are not uniquely determined and the parameters in the first step estimator are not uniquely identified, and hence the first step estimator is not consistent. Second, it is a requirement for λ and υ to be continuously differentiable (see Theorem 1 of Magnus, 1985), which is needed to apply the mean-value theorem when considering the asymptotic distribution of the estimator. The second part of Assumption 4.2 imposes an identifying normalization on the first step estimator and ensures a unique identification of the unconditional eigenvalues and -vectors. This is needed since the ordering of the eigenvalues is not fixed and the sign of the eigenvectors is unidentified without imposing a normalization.

Remark 4.1 (Consequences of failure of Assumption 4.2)**.** *When the first part of Assumption 4.2 is violated, that is, the case of repeated eigenvalues, the eigenvectors can only be identified up to an orthogonal transformation. Furthermore, Ledoit and Wolf (2004) present simulation evidence that the estimates of λ_i tend to be under-/overstated, when eigenvalues are repeated. It should be noted, and as emphasized by Jolliffe (2002) (Section 2.4), that repeated eigenvalues are a problem that (paraphrasing slightly)* "arise in theory, but are relatively uncommon in practice"*, and we note that assumptions similar to our Assumption 4.2 are commonly made in both the statistics and econometrics* *literature, see e.g., Aït-Sahalia and Xiu (2019) and references therein. Anderson (1963), Eaton and Tyler (1991), and Jolliffe (2002) (Section 3.7) discuss methods for testing of repeated eigenvalues.*

We also assume that the dynamic parameters of the model are identified and that the true parameter vector is a subset of the parameter space.

Assumption 4.3. *The true parameter vector* $\theta_0^{(i)} \in \Theta^{(i)}$*, with* $\Theta^{(i)}$ *compact.*

Assumption 4.4. *For* $\kappa^{(i)} \in \mathcal{K}^{(i)}$*, if* $\kappa^{(i)} \neq \kappa_0^{(i)}$*, then* $\lambda_{i,t}(\gamma_0, \kappa^{(i)}) \neq \lambda_{i,t}(\gamma_0, \kappa_0^{(i)})$ *(almost surely).*

Assumptions 4.3-4.4 are standard for (multivariate) GARCH models, see e.g., Comte and Lieberman (2003) or Hafner and Preminger (2009a). Moreover, the normalization imposed on the first step estimator ensures that the eigenvalues and -vectors of the first step estimator are uniquely identified. Assumption 4.4 is high-level, and Assumption 3.4 and 3.5 of Hetland, Pedersen, and Rahbek (2020) provide sufficient and primitive conditions under which it holds. Most importantly, the matrix $2p \times p$ $[A_0, B_0]$ must have full rank p.

These assumptions lead us to the following theorem from Hetland, Pedersen, and Rahbek (2020) (Theorem 3.2) on strong consistency of the STE.

Theorem 4.1. *Under Assumptions 4.1-4.4, as* $T \to \infty$*, the ST estimator is consistent,*

$$\hat{\theta}^{(i)} \overset{a.s}{\to} \theta_0^{(i)}.$$

Next, we show that the estimator is asymptotically normal. To do so, we need two additional assumptions on existence of moments and the true parameter vector.

Assumption 4.5. *The process* $\{X_t\}_{t\in\mathbb{Z}}$ *has finite fourth order moments,* $E||X_t||^4 < \infty$*.*

Assumption 4.5 is required to ensure that the first step estimator, $\sqrt{T}(\hat{\gamma} - \gamma_0)$, converges to a Gaussian distribution with a finite variance. This assumption is common in the variance targeting literature and is also needed when reparameterizing the moment estimator in terms of the spectral decomposition. In fact, the moment requirement is not needed in the probability analysis of the profiled log-likelihood function, but it is needed in the sense that the second step estimator depends on convergence of $\hat{\gamma}$.. Lemma C.2 in Appendix C can be used to check the moment condition in Assumption 4.5. Based on the simulations included in Appendix D, the moment conditions for consistency and asymptotic normality are sufficient and necessary.

Assumption 4.6. $\theta_0^{(i)}$ *is in the interior of* $\Theta^{(i)}$*.*

Assumption 4.6 is standard in the literature, and is a technical requirement to ensure that the mean-value theorem can be applied on the optimality condition for the profiled log-likelihood function.

This leads us to the next theorem on asymptotic normality of the estimator for the i'th rotated return,

Theorem 4.2. *Under Assumptions 4.1-4.6, for $T \to \infty$*

$$\sqrt{T}\left(\hat{\theta}^{(i)} - \theta_0^{(i)}\right) \xrightarrow{D} N\left(0, \Sigma_0^{(i)}\right),$$

where $\Sigma_0^{(i)}$ is the asymptotic covariance matrix, given by,

$$\underset{e\times e}{\Sigma_0^{(i)}} = \begin{pmatrix} I_{p(p+1)} & 0_{p(p+1)\times p+1} \\ -(J_0^{(i)})^{-1}K_0^{(i)} & -(J_0^{(i)})^{-1} \end{pmatrix} \Omega_0^{(i)} \begin{pmatrix} I_{p(p+1)} & -(J_0^{(i)})^{-1}(K_0^{(i)})' \\ 0_{p+1\times p(p+1)} & -(J_0^{(i)})^{-1} \end{pmatrix}. \tag{4.1}$$

where $J_0^{(i)}$ and $K_0^{(i)}$ are defined in (A.12) *and $\Omega_0^{(i)}$ is given in* (B.22)*.*

In the derivation of the asymptotic distribution of the first step (and consequently the second step) estimator, we restate the moment estimator as the average of the conditional eigenvalues. However, as $\text{vec}(\hat{H}) = V^{\otimes 2}\text{vec}(\frac{1}{T}\sum_{t=1}^{T}\Lambda_{0,t}^{1/2}Z_tZ_t'\Lambda_{0,t}^{1/2})$ is in terms of Λ_t, and not the vectorized eigenvalues, λ_t, we restate the dynamics of the conditional eigenvalues in (2.4) as a (restricted) BEKK$(p^2, 1, 1, 1)$ model for Λ_t. This parametrization is present in $\Omega_0^{(i)}$ in (B.22). In doing so, $\hat{\gamma} - \gamma_0$ is shown to (asymptotically) be a martingale difference, allowing us to use a central limit theorem for martingale differences on $\sqrt{T}\left(\hat{\gamma} - \gamma_0 \quad \frac{\partial L_T^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}}\right)'$ jointly to show normality and find the expression for $\Omega_0^{(i)}$. The proof of joint normality of $\sqrt{T}(\hat{\theta}^{(i)} - \theta_0^{(i)})$ applies the mean-value theorem on the optimality condition of the second step estimator, stacked with the moment estimator from step one, and Lemmata B.6-B.10 in Appendix B verify that the mean-value theorem can be applied.

Remark 4.2 (Fixed initial values)**.** *Assumption 4.1 assumes that the process is strictly stationary, implying that the process $\{X_t\}_{t\in\mathbb{Z}}$ is initiated in the invariant distribution or in the infinite past. In practice the observed process (in the likelihood function) is initiated in some fixed values, X_0 and H_0, which by definition makes the process non-stationary. However, Lemmata B.4 and B.10 in Appendix B verifies that the choice of initial values are asymptotically irrelevant for both consistency and asymptotic normality of the estimator.*

In (very) large portfolios, practitioners may prefer a "diagonal" specification of the model, in which both A and B are restricted to be diagonal, to reduce dimensionality of the model. In this case the parameter vector is $\kappa^{(i)} = [a_i, b_i]'$, and the lemmata in the appendix can easily be modified such that the estimator is still consistent and asymptotically normal by Theorems 4.1-4.2.

In applications, the asymptotic variance matrix for the stepwise estimator may be approximated using plug-in sample estimators. That is,

$$\underset{p+1\times p+1}{\hat{J}_T^{(i)}} = \frac{1}{T}\sum_{t=1}^{T} \left.\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial\kappa^{(i)}\partial\kappa^{(i)\prime}}\right|_{\theta^{(i)}=\hat{\theta}^{(i)}}, \qquad \underset{p+1\times p^2+p}{\hat{K}_T^{(i)}} = \frac{1}{T}\sum_{t=1}^{T} \left.\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial\kappa^{(i)}\partial\gamma'}\right|_{\theta^{(i)}=\hat{\theta}^{(i)}}, \tag{4.2}$$

$$\underset{e\times e}{\hat{\Omega}_T^{(i)}} = \frac{1}{T}\sum_{t=1}^{T}\hat{\omega}_t\hat{\omega}_t' \quad \text{where} \quad \underset{e\times 1}{\hat{\omega}_t} := \begin{pmatrix} D\hat{V}^{\otimes 2}\left(\text{vec}(X_tX_t) - \text{vec}(\hat{H})\right) \\ \hat{V}_1' \otimes (\hat{\lambda}_1 I_p - \hat{H})^+\left(\text{vec}(X_tX_t) - \text{vec}(\hat{H})\right) \\ \vdots \\ \hat{V}_p' \otimes (\hat{\lambda}_p I_p - \hat{H})^+\left(\text{vec}(X_tX_t) - \text{vec}(\hat{H})\right) \\ \left.\frac{\partial l_t^{(i)}(\theta^{(i)})}{\partial\kappa^{(i)}}\right|_{\theta^{(i)}=\hat{\theta}^{(i)}} \end{pmatrix} \tag{4.3}$$

where $(\hat{\lambda}_i I_p - \hat{H})^+$ denotes the Penrose-Moore pseudo-inverse of $\hat{\lambda}_i I_p - \hat{H}$ for $i = 1, \ldots, p$ and D is defined in Lemma B.5. The expressions in (4.2)-(4.3) converge almost surely to their population counterparts due to the ergodic theorem. Note that we may substitute the estimators by their true value, as we have established strong consistency of $\theta^{(i)}$. This makes estimation of the asymptotic covariance matrix only slightly more cumbersome than that of the well-known "sandwich" covariance matrix estimator from joint QMLE.

Once we know the asymptotic distribution of $\hat{\theta}^{(i)}$, it is possible to derive the asymptotic distribution of the original parameters of the model in (2.9), $\phi^{(i)} = [w_i, a_{i1}, \ldots, a_{ip}, b_i]'$ using the delta method.

Corollary 1 (Limiting distribution of $\phi^{(i)} = [w_i, a_{i1}, \ldots, a_{ip}, b_i]'$). *Under Assumption 4.1-4.6, for $T \to \infty$,*

$$\sqrt{T}(\hat{\phi}^{(i)} - \phi_0^{(i)}) \xrightarrow{D} N(0, \varphi_0^{(i)}\Sigma_0^{(i)}\varphi_0^{(i)\prime})$$

for $i = 1, \ldots, p$ with $\Sigma_0^{(i)}$ given in (4.1) *and,*

$$\underset{p+2\times e}{\varphi_0^{(i)}} = \left.\frac{\partial\phi^{(i)}}{\partial\theta^{(i)\prime}}\right|_{\theta^{(i)}=\theta_0^{(i)}} = \begin{pmatrix} (1-b_{0,i})\mathbb{I}\{i\} - A_{0,i}' & 0_{1\times p^2} & -\lambda_0' & -\lambda_{0,i} \\ 0_{p\times p} & 0_{p\times p^2} & I_p & 0_{p\times 1} \\ 0_{1\times p} & 0_{1\times p^2} & 0_{1\times p} & 1 \end{pmatrix} \tag{4.4}$$

where $\mathbb{I}\{i\}$ is a $1 \times p$ vector of zeros, apart from a "1" in the i'th column and A_i is the i'th row of A.

Note that we present the asymptotic theory in terms of $\theta^{(i)}$ rather than θ, following Francq and Zakoïan (2016) and their notation for an equation-by-equation estimator of various MGARCH models. The theorems listed above could easily be restated in terms of θ, as the asymptotic results hold simultaneously due to the orthogonality of the conditional univariate log-likelihood functions, but we refrain from doing so for two reasons: First, the present formulation is coherent with the step-wise approach of the estimator. Second, this presentation makes it straightforward to parallelize estimation, computing the asymptotic variance matrix in each iteration, which speeds up the estimation procedure.

As already emphasized, the STE reduces the risk of numerical issues in estimation compared to the QML estimator, and in the context of the λ-GARCH model, the ST estimator is closely related to the variance targeting estimator: Because the profiled log-likelihood consists of p orthogonal terms, the STE and VTE of the λ-GARCH are theoretically equivalent, and in practice is expected to produce similar estimates and standard errors. Note however, that we expect the STE to have a smaller computational burden, as it minimizes the log-likelihood function over a smaller parameter space.

Remark 4.3 (Variance targeting estimator of the λ-GARCH)**.** *An alternative to the stepwise estimation of the λ-GARCH is the variance targeting (VT) estimator, in which all $\kappa = [\kappa^{(1)\prime}, \ldots, \kappa^{(p)\prime}]'$ are estimated jointly. This estimator still relies on the first step estimator of $\hat{\gamma}$ in* (3.2)*, and we denote the full VT estimator $\tilde{\theta} = [\hat{\gamma}', \tilde{\kappa}']'$. Because of the orthogonal structure of the log-likelihood function, consistency and asymptotic normality of the VT estimator of the λ-GARCH can be derived using similar techniques as in Appendix A and B.*

5 Empirical Illustrations

In the following we compare the empirical performance of the ST estimator to that of the joint QML estimator. First, we consider the relative efficiency of the two estimators in a simulation setting for different portfolio sizes. This exercise lets us compare the (empirical) efficiency of the STE against the QMLE. Second, we consider the out-of-sample performance of the the two estimation methods in a recursive value-at-risk application for portfolios of $p = 25$ assets. The empirical fit is assessed using the likelihood ratio tests of Christoffersen (1998). In both these exercises, we also consider the computational complexity (i.e., time spent on estimating the model) of the two methods. Finally, we briefly summarize the results.

5.1 Relative efficiency: STE vs. QMLE

We now compare the relative efficiency and the time complexity of the STE against the joint QMLE. This is done for the diagonal model of dimension p, where we simulate a data-generating process $N = 399$ times with $A_0 = 0.05I_p$, $B_0 = 0.85I_p$, such that the process has finite fourth order moments. The unconditional eigenvalues are specified as $\lambda_{0,i} = (p + 1 - i)/10$ for $i = 1, \ldots, p$ and the eigenvectors are constructed using rotation matrices with all rotation parameters $\phi_{0,i} = 0.5$ for $i = 1, \ldots, p(p-1)/2$ (see Remark

3.1). The innovations, Z_t, are drawn *iid* from a standard normal distribution, and each path of the simulated process has $T = 2000$ observations. The model has $p(p-1)/2 + 3p$ parameters, and the STE procedure estimates $p(p+1)/2$ parameters in the first step, and the remaining $2p$ parameters sequentially for each rotated return. The QMLE on the other hand estimates all $p(p-1)/2 + 3p$ parameters simultaneously.

In comparing the two estimators, we employ the same methodology as Francq and Zakoïan (2016) who use the quadratic form $T(\hat{\vartheta} - \vartheta_0)'\mathcal{I}(\hat{\vartheta} - \vartheta_0)$ as a measure of accuracy of an estimator $\hat{\vartheta}$, where $\mathcal{I}$ is the (numerically) approximated information matrix and the parameter vector is constructed identically for both estimators with $\vartheta = [\text{vec}(H)', \text{diag}(A)', \text{diag}(B)']'$. Because $\mathcal{I}$ is computationally demanding to compute in higher dimensions, we instead use the simulated information matrix, which is obtained as $\mathcal{I} = var(\hat{\vartheta}^n_{QMLE} - \vartheta_0) \approx \frac{1}{N}\sum_{n=1}^{N}(\hat{\vartheta}^n_{QMLE} - \vartheta_0)(\hat{\vartheta}^n_{QMLE} - \vartheta_0)'$ for $n = 1, \ldots, N$, where $\hat{\vartheta}^n_{QMLE}$ is the QMLE parameter vector for the n'th simulated path. The relative efficiency is then computed as,

$$RE = \frac{(\bar{\vartheta}_{STE} - \vartheta_0)'\mathcal{I}(\bar{\vartheta}_{STE} - \vartheta_0)}{(\bar{\vartheta}_{QMLE} - \vartheta_0)'\mathcal{I}(\bar{\vartheta}_{QMLE} - \vartheta_0)},$$

where $\bar{\vartheta}_{STE} = \frac{1}{N}\sum_{n=1}^{N}\hat{\vartheta}^n_{STE}$ with $\bar{\vartheta}_{QMLE}$ defined analogously. By this definition, if $RE < 1$, the ST estimator is relatively more efficient than the QML estimator.

TABLE 5.1: Relative efficiency: QMLE vs. STE.

Dimension, p	# parameters	Time (s), QMLE	Time (s), STE	RE
2	7	12.82	0.69	2.59
4	18	52.93	1.36	1.42
6	33	161.70	1.92	2.05
8	52	358.16	2.74	2.06
10	75	505.75	3.76	1.13
12	102	546.40	4.86	0.11
15	150	617.31	6.41	0.01
20	250	765.87	10.03	0.00
50	1,375	2,987.88	33.00	0.03
100	5,250	15,775.08	141.27	0.04
200	20,500	N/A	618.29	N/A
500	126,250	N/A	4,084.79	N/A

For the case $p = 200$ and $p = 500$ the QMLE failed to converge.
The time complexity and RE is the average over $N = 399$ simulations.
All simulations/estimations are done using a single core.

The (average) computation times and the relative efficiency for the two estimators are contained in Table 5.1. For the larger systems, $p > 20$, the computation time for QMLE is very big, on average 50 minutes for $p = 50$ and 260 minutes for $p = 100$, whereas STE remains feasible in all but the $p = 500$ case, in which the computation time is roughly one hour. Considering the relative efficiency of the two estimators, the QMLE performs favorably for $p \leq 10$, after which its performance deteriorate drastically compared to the STE. For portfolios larger than 10 assets, the STE is preferred.

Here, estimations are initiated in $\vartheta_{init} = \vartheta_0 - 0.025$. However, one could argue that initiating both estimators in $\vartheta_{init} \neq \vartheta_0$ gives the joint QMLE a disadvantage, as it performs numerical optimization over a much larger parameter space. We therefore repeat the exercise, initiating in $\vartheta_{init} = \vartheta_0$. This yields almost identical results (available upon request) and leads to the same conclusion, namely that the STE is relatively more efficient than joint QMLE for systems larger than $p > 10$ assets, and that it always has a lower computational complexity than joint QMLE.

5.2 An application in risk management

We now turn our attention to the empirical performance of the STE of the λ-GARCH, and compare it to the joint QML estimator. The out-of-sample performance is assessed by considering the conditional 5% value-at-risk (VaR) for five different medium-sized portfolios consisting of $p = 25$ assets from the SP100 index.

Methodology and data

We consider the out-of-sample performance by considering the conditional 5% value-at-risk at 1 and 5-day horizons for five different portfolios. The first of the five portfolios is equally weighted while the weights of the remaining portfolios are drawn randomly such that the second and third portfolios are long-only, with the third portfolio 50% geared. The fourth and fifth portfolios are long-short portfolios. The constituents of the portfolios are drawn randomly from the SP100 index and can, along with their weighting, be found in Appendix E.

The ST and the QML estimators are fitted on a (rolling window) sample of $T = 1200$ daily observations, with the initial sample starting on December 28th 2010, ending on December 29th 2015. The out-of-sample consists of 3 years of data from December 30th 2015 to December 31st 2018, leading to $\tau = 756$ out-of-sample observations for the 1-day forecast and $\tau = 189$ observations for the 5-day (non-overlapping) forecasts. The out-of-sample forecasts are computed using a filtered historical simulation in which we draw innovations *iid* with replacement from the standardized residuals, $\hat{Z}_t$, see e.g., Christoffersen (2009).

Recall that the conditional VaR at risk level α for the h-period return of portfolio i, denoted $\text{VaR}^i_{t,h}(\alpha)$ is defined as,

$$P_t(R^i_{t+h|t} < -\text{VaR}^i_{t,h}(\alpha)) = \alpha, \tag{5.1}$$

where P_t is the conditional distribution of the ex ante h-period return of portfolio i, $R^i_{t+h|h}$. We define the (unconditional) "hit" variable for portfolio i as,

$$I^i_t = \mathbf{1}\{R^i_{t+h|t} > -\text{VaR}^i_{t,h}(\alpha)\}, \tag{5.2}$$

such that the unconditional coverage for portfolio i is $\pi_i = \frac{1}{\tau}\sum_{n=1}^{\tau} I^i_n$. Similarly, we define the conditional hit variable as $I^i_{t|t-1} = \mathbf{1}\{I^i_t = 1|I^i_{t-1} = 1\}$, denoting two hits in a row.

When assessing the adequacy of the VaR forecasts we consider the three likelihood ratio (LR) tests proposed by Christoffersen (1998). The first LR test examines the hypothesis that the unconditional coverage is correct, $E[I^i_t] = \alpha$, but fails to account for potential clustering in the VaR hits. This is rectified by the second test, in which $I^i_{t|t-1}$ follows a two-state Markov chain, and we test the hypothesis of independence between hits. However, this test does not test for correct coverage, and as a consequence, we also consider the third test of correct conditional coverage, which lets $I^i_{t|t-1}$ follow the two-state Markov chain, and tests it against the null of independence between hits and correct coverage. The tests are denoted LR_{uc}, LR_{ind} and LR_{cc} respectively.

Out-of-sample results

The results of the out-of-sample exercise is given in Table 5.2. Importantly, the STE procedure is roughly 57 times faster than the QMLE. We note that the estimated λ-GARCH (on average) has finite second order moments but not fourth order moments. Intuitively, this means that both estimators are consistent, but only the QML estimator has a limiting Gaussian distribution.

The two estimation methods have a similar performance based on unconditional coverage and the LR-tests: In general, the unconditional coverage is slightly different from the hypothesized 5% and most of the LR-tests do not reject. Similar results are found for the 1 and 5-day 1% VaR (not reported here). The rejected LR-tests relate to the equally weighted portfolio P_1.

In general, the VaR estimates produced by the two estimation methods are similar, but not identical: Consider Figure 5.1 which plots the estimated VaR for the two estimation methods along with the realized return of portfolio P_4. As shown, the VaR estimates are, for the majority of the sample, very similar, but the QML estimator sometimes produce more extreme VaR estimates than the STE. We note, however, that while the two VaR estimates at times differ, the unconditional and conditional hit sequences are almost identical, and based on the LR-test in Table 5.2, none of the estimation methods seem to dominate the other empirically. We therefore conclude that the estimation procedures seem to yield similar results, with the STE having the clear advantage that it is much faster in practice.[2]

[2] As an alternative empirical exercise, we also consider the one-step ahead minimum-variance portfolio

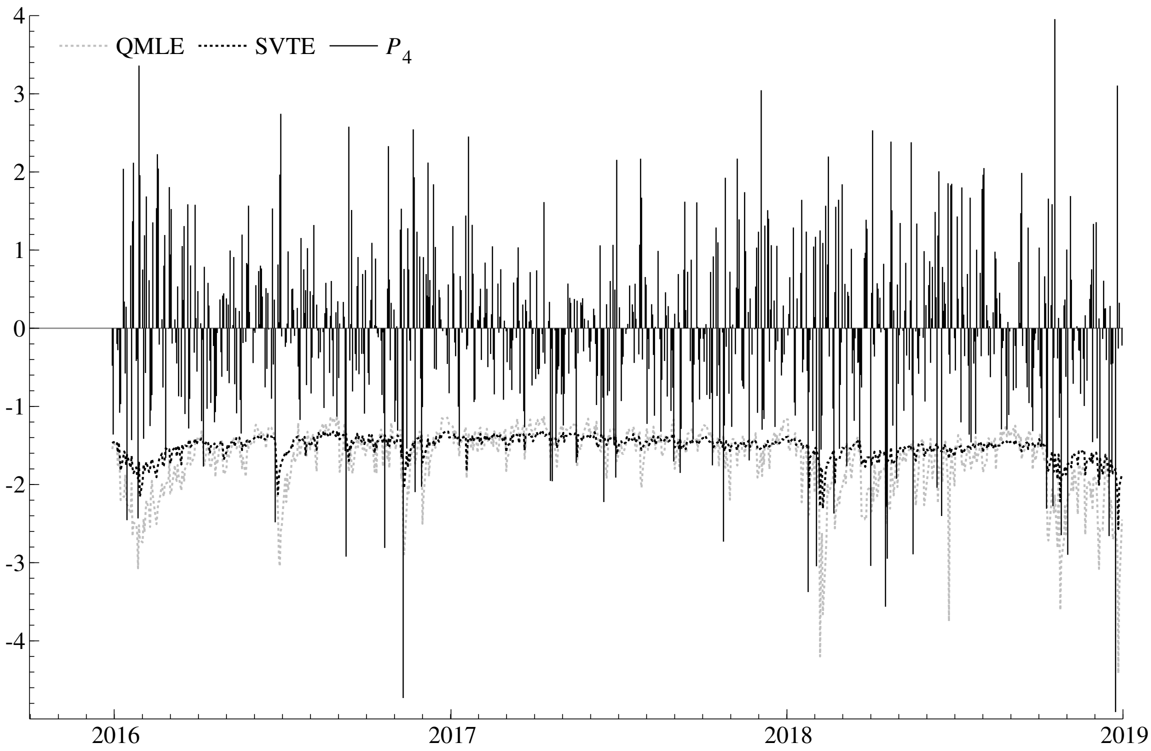
Table 5.2: Empirical exercise: 95% VaR at 1 and 5-day forecast horizons.

		1-day		5-day	
		STE	QMLE	STE	QMLE
P_1	$\hat{\pi}_1$	0.044	0.038	0.048	0.048
	LR_{uc}	0.413	0.126	0.880	0.880
	LR_{ind}	0.013	0.910	0.427	0.056
	LR_{cc}	0.033	0.309	0.721	0.160
P_2	$\hat{\pi}_2$	0.046	0.040	0.053	0.058
	LR_{uc}	0.636	0.178	0.856	0.614
	LR_{ind}	0.093	0.478	0.090	0.136
	LR_{cc}	0.218	0.313	0.234	0.290
P_3	$\hat{\pi}_3$	0.042	0.038	0.053	0.058
	LR_{uc}	0.321	0.126	0.856	0.614
	LR_{ind}	0.050	0.910	0.537	0.136
	LR_{cc}	0.089	0.309	0.813	0.290
P_4	$\hat{\pi}_4$	0.065	0.052	0.074	0.069
	LR_{uc}	0.073	0.842	0.155	0.261
	LR_{ind}	0.307	0.409	0.078	0.269
	LR_{cc}	0.119	0.697	0.077	0.288
P_5	$\hat{\pi}_5$	0.065	0.050	0.058	0.053
	LR_{uc}	0.073	0.973	0.614	0.856
	LR_{ind}	0.453	0.449	0.243	0.290
	LR_{cc}	0.152	0.751	0.446	0.562

	STE	QMLE
Time complexity	9.1	526.5

P_i refers to the i'th portfolio, with $\hat{\pi}_i$ being the unconditional hit ratio $i = 1, ..., 5$. LR_{uc}, LR_{ind} and LR_{cc} are the asymptotic p-values for the LR test for unconditional coverage, independence and conditional coverage respectively. The time complexity is given in seconds and is computed using a single core for one out-of-sample iteration.

FIGURE 5.1: 1-day 95% VaR for portfolio P_4



Note: STE and QMLE denote the estimated 1-day 5% VaR, P_4 denotes the realized return.

5.3 Brief summary of numerical exercises

The simulation evidence in Section 5.1 indicates that not only is the STE relatively more efficient than QMLE in cross-sections of more than $p > 10$ assets, it is also much more time efficient. This is verified in by the empirical study in Section 5.2. One potential explanation is that the λ-GARCH is a non-linear function of the parameters in γ through Y_{t-1}. By using a stepwise estimator, in which γ is estimated using a closed form estimator, we mitigate the potential issues due to non-linearity, which seem to cause issues for large p in the QML estimator.

In regards to the asymptotic results in Theorems 4.1-4.2, the simulation study in Appendix D suggests that the estimator is consistent in the case of finite second order moments of X_t. Furthermore, the simulations indicate that the asymptotic normality of the STE holds when X_t has finite fourth moments. Hence, the moment requirements in Assumption 4.1 and 4.5 appear to be sufficient and necessary for consistency and asymptotic normality of the STE.

for $p \in (5, 10, 25, 50)$, where we find that the STE yields portfolios with a lower ex post variance compared to the QML estimator as p increases, such that the STE outperforms QMLE for portfolios of $p \geq 25$. These results are not reported for brevity, but are available upon request.

6 Extensions and Concluding Remarks

We have derived asymptotic properties of the spectral targeting estimator (STE) for the λ-GARCH, an extended version of the multivariate orthogonal GARCH (O-GARCH). This two-step estimator is consistent under finite second order moments, while it has a limiting Gaussian distribution when fourth order moments are finite. Simulations indicate that these moment conditions are sufficient and necessary. Moreover, we compare the empirical performance of the STE to that of the quasi-maximum likelihood estimator (QMLE) for five portfolios of 25 assets. The STE dominates QMLE in terms of computational complexity, being up to 57 times faster in estimation, while both estimators produce similar out-of-sample forecasts. Finally, simulations indicate that in portfolios of more than 10 assets, the stepwise estimator is relatively more efficient than QMLE. The STE is therefore well suited for practitioners as it alleviates numerical problems and speeds up numerical optimization, while being easy to implement.

We note that while the STE delivered promising results in this exposition, the first step (sample) estimator may not be well-behaved when the ratio p/T approaches one. This is discussed in e.g., Ledoit and Wolf (2004, 2012), who derive shrinkage estimators for the sample covariance matrix, minimizing the estimation error. An extension could therefore consider the asymptotic analysis of a spectral targeting estimator where the first step estimator is based on shrinkage. Another extension would be to consider spectral targeting estimation with infinite fourth order moments, in a similar fashion to the exposition in Pedersen (2016) who consider the variance targeting estimator of the extended constant conditional correlation GARCH model.

APPENDIX

A Proofs

Recall the log-likelihood function for the i'th equation,

$$L_T^{(i)}(\gamma, \kappa^{(i)}) = \frac{1}{T}\sum_{t=1}^{T} l_t^{(i)}(\gamma, \kappa^{(i)}), \tag{A.1}$$

$$l_t^{(i)}(\gamma, \kappa^{(i)}) = \log(\lambda_{i,t}(\gamma, \kappa^{(i)})) + \frac{y_{i,t}^2(\gamma)}{\lambda_{i,t}(\gamma, \kappa^{(i)})}, \tag{A.2}$$

which has first and second order derivatives,

$$\frac{\partial l_t^{(i)}(\theta^{(i)})}{\partial \theta_n^{(i)}} = \left(1 - \frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})}\right) \frac{1}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}} + 2\frac{y_{i,t}(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial y_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}}, \tag{A.3}$$

$$\begin{aligned}
\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial \theta_n^{(i)} \partial \theta_m^{(i)}} &= \left(1 - \frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})}\right) \frac{1}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial^2 \lambda_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)} \partial \theta_m^{(i)}} \\
&+ \left(2\frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})} - 1\right) \frac{1}{\lambda_{i,t}^2(\theta^{(i)})} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_m^{(i)}} \\
&+ 2\left(\frac{\partial^2 y_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)} \partial \theta_m^{(i)}} y_{i,t}(\theta^{(i)}) + \frac{\partial y_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}} \frac{\partial y_{i,t}(\theta^{(i)})}{\partial \theta_m^{(i)}}\right) \frac{1}{\lambda_{i,t}(\theta^{(i)})} \\
&- 2\left(\frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}} \frac{\partial y_{i,t}(\theta^{(i)})}{\partial \theta_m^{(i)}} + \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_m^{(i)}} \frac{\partial y_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}}\right) \frac{y_{i,t}(\theta^{(i)})}{\lambda_{i,t}^2(\theta^{(i)})},
\end{aligned} \tag{A.4}$$

for $n, m = 1, \ldots, p(p+1)+1$.

Throughout the proofs, we let $L_t^{(i)}(\theta^{(i)})$ ($l_t^{(i)}(\theta^{(i)})$) denote the log-likelihood function (-contribution) initiated in the infinite past, and we let $L_{t,h}^{(i)}(\theta^{(i)})$ ($l_{t,h}^{(i)}(\theta^{(i)})$) denote the log-likelihood function (-contribution) initiated in a fixed X_0 and H_0 (with H_0 positive definite),

$$L_{T,h}^{(i)}(\gamma, \kappa^{(i)}) = \frac{1}{T} \sum_{t=1}^{T} l_{t,h}^{(i)}(\gamma, \kappa^{(i)}), \tag{A.5}$$

$$l_{t,h}^{(i)}(\gamma, \kappa^{(i)}) = \log(\lambda_{i,t,h}(\gamma, \kappa^{(i)})) + \frac{y_{i,t}^2(\gamma)}{\lambda_{i,t,h}(\gamma, \kappa^{(i)})}, \tag{A.6}$$

where $\lambda_{i,t}(\theta^{(i)})$ and $\lambda_{i,t,h}(\theta^{(i)})$ is defined analogously. Because $\lambda_{i,t}(\theta^{(i)})$ and $\lambda_{i,t,h}(\theta^{(i)})$ are defined for the same strictly stationary and ergodic sequence, $\{X_t\}_{t \in \mathbb{Z}}$, we may write,

$$\lambda_{i,t}(\theta^{(i)}) - \lambda_{i,t,h}(\theta^{(i)}) = b_i^t(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)})), \tag{A.7}$$

for $t \geq 1$.

The structure of the main proofs and the accompanying lemmata follow that of Pedersen and Rahbek (2014) (proof of Theorems 4.1-4.2 and Lemmata B.1-B.11). In order to make the proofs readable, most steps rely on lemmata stated and proved in Appendix B. In the following, we let the letters K and ϕ denote generic constants, whose value can vary along the text, but always satisfy $K > 0$ and $0 < \phi < 1$. Furthermore, let $H_{0,t} := H_t(\theta_0^{(i)})$, $V_0 := V(\theta_0^{(i)})$ and $\Lambda_{0,t} := \Lambda_t(\theta_0^{(i)})$.

A.1 Proof of consistency

Initially, observe that by the ergodic theorem (Theorem 20.3 of Jacod and Protter (2012)), along with Assumption 4.1 the sample estimator is strongly consistent, $\hat{H} \overset{a.s.}{\to} H_0$, for $T \to \infty$. Since λ_0 is assumed to be simple, we may use the continuous mapping theorem (Theorem 17.5 of Jacod and Protter (2012)) to establish strong consistency of the first stage estimation,

$$\hat{\lambda} \to \lambda_0 \qquad \text{a.s.} \tag{A.8}$$

$$\hat{\upsilon} \to \upsilon_0 \qquad \text{a.s.} \tag{A.9}$$

We now show that $\hat{\kappa}^{(i)}$ is consistent. The proof follows that of Theorem 4.1 in Pedersen and Rahbek (2014).

For any $\varepsilon > 0$, it holds almost surely for large T,

$$\begin{aligned}
E[l_t^{(i)}(\gamma_0, \hat{\kappa}^{(i)})] &< L_T^{(i)}(\gamma_0, \hat{\kappa}^{(i)}) + \frac{\varepsilon}{5} && \text{By Lemma B.2,} \\
L_T^{(i)}(\gamma_0, \hat{\kappa}^{(i)}) &< L_{T,h}^{(i)}(\hat{\gamma}, \hat{\kappa}^{(i)}) + \frac{\varepsilon}{5} && \text{By Lemma B.4,} \\
L_{T,h}^{(i)}(\hat{\gamma}, \hat{\kappa}^{(i)}) &< L_{T,h}^{(i)}(\hat{\gamma}, \kappa_0^{(i)}) + \frac{\varepsilon}{5} && \text{By (3.8),} \\
L_{T,h}^{(i)}(\hat{\gamma}, \hat{\kappa}^{(i)}) &< L_T^{(i)}(\gamma_0, \kappa_0^{(i)}) + \frac{\varepsilon}{5} && \text{By Lemma B.4,} \\
L_T^{(i)}(\gamma_0, \kappa_0^{(i)}) &< E[l_t^{(i)}(\gamma_0, \kappa_0^{(i)})] + \frac{\varepsilon}{5} && \text{By Lemma B.2 .}
\end{aligned}$$

That is, for any $\varepsilon > 0$,

$$E[l_t^{(i)}(\gamma_0, \hat{\kappa}^{(i)})] < E[l_t^{(i)}(\gamma_0, \kappa_0^{(i)})] + \varepsilon,$$

and by Lemma B.3 along with standard arguments for two-step estimators (Newey and McFadden, 1994), it follows that for $T \to \infty$, $\hat{\kappa}^{(i)} \overset{a.s.}{\to} \kappa_0^{(i)}$. Hence, the two-step estimator is strongly consistent, $\hat{\theta}^{(i)} \overset{a.s.}{\to} \theta^{(i)}$. □

A.2 Proof of asymptotic normality

Compared to asymptotic theory for the joint QMLE of multivariate GARCH models additional difficulties arise from the fact that STE is a multi-step estimator. Conversely, the proof is simplified by the additional assumption of $E||X_t||^4 < \infty$ and the fact that we treat individual $\{y_{i,t}\}_{t \in \mathbb{Z}}$ separately.

The proof of asymptotic normality is based on an application of the mean-value theorem on the optimality condition of the score vector, $\theta^{(i)} = \theta_0^{(i)}$ along with Assumption

4.6 and (3.8),

$$
\begin{aligned}
0_{p+1\times 1} &= \frac{\partial L_{T,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} + \left(\frac{\partial^2 L_{T,h}^{(i)}(\tilde{\theta}^{(i)})}{\partial \kappa^{(i)} \partial \theta^{(i)\prime}}\right)(\hat{\theta}^{(i)} - \theta_0^{(i)}) \\
&= \frac{\partial L_{T,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} + J_{T,h}^{(i)}(\tilde{\theta}^{(i)})(\hat{\kappa}^{(i)} - \kappa_0^{(i)}) + K_{T,h}^{(i)}(\tilde{\theta}^{(i)})(\hat{\gamma} - \gamma_0),
\end{aligned}
\tag{A.10}
$$

where

$$
\frac{\partial L_{t,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} = \left.\frac{\partial L_{t,h}^{(i)}(\theta^{(i)})}{\partial \kappa^{(i)}}\right|_{\theta^{(i)}=\theta_0^{(i)}}, \quad J_{T,h}^{(i)}(\tilde{\theta}^{(i)}) = \left.\frac{\partial^2 L_{T,h}^{(i)}(\theta^{(i)})}{\partial \kappa^{(i)} \partial \kappa^{(i)\prime}}\right|_{\theta^{(i)}=\tilde{\theta}^{(i)}}, \quad K_{T,h}^{(i)}(\tilde{\theta}^{(i)}) = \left.\frac{\partial^2 L_{T,h}^{(i)}(\theta^{(i)})}{\partial \kappa^{(i)} \partial \gamma'}\right|_{\theta^{(i)}=\tilde{\theta}^{(i)}}.
$$

Here $\kappa^{(i)} = [a_{i1}, \ldots, a_{ip}, b_i]'$, $\gamma = [\lambda', \upsilon']'$, and $\tilde{\theta}^{(i)}$ is on the line between $\theta_0^{(i)}$ and $\hat{\theta}^{(i)}$.

$J_T^{(i)}$ is finite and invertible with probability approaching one (Lemma B.7 and B.9) and $\hat{\theta}^{(i)} \overset{a.s.}{\to} \theta^{(i)}$ (Theorem 4.1). Hence, by Lemma B.10, (A.10) can be rewritten as,

$$
\sqrt{T}\left(\hat{\kappa}^{(i)} - \kappa_0^{(i)}\right) = -\left(J_T^{(i)}(\tilde{\theta}^{(i)})\right)^{-1}\sqrt{T}\frac{\partial L_t^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} - \left(J_T^{(i)}(\tilde{\theta}^{(i)})\right)^{-1} K_T^{(i)}(\tilde{\theta}^{(i)})\sqrt{T}\left(\hat{\gamma} - \gamma_0\right) + o_p(1),
$$

which we rewrite for the (joint) parameter vector of equation i,

$$
\sqrt{T}\begin{pmatrix} \hat{\gamma} - \gamma_0 \\ \hat{\kappa}^{(i)} - \kappa_0^{(i)} \end{pmatrix} = \begin{pmatrix} I_{p(p+1)} & 0_{p(p+1)\times p+1} \\ -(J_T^{(i)}(\tilde{\theta}^{(i)}))^{-1} K_T^{(i)}(\tilde{\theta}^{(i)}) & -(J_T^{(i)}(\tilde{\theta}^{(i)}))^{-1} \end{pmatrix} \begin{pmatrix} \sqrt{T}(\hat{\gamma} - \gamma_0) \\ \sqrt{T}\frac{\partial L_t^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} \end{pmatrix} + o_p(1).
$$

The asymptotic normality then follows from Lemma B.6 together with Slutsky's theorem,

$$
\sqrt{T}\left(\hat{\theta}^{(i)} - \theta_0^{(i)}\right) \overset{D}{\to} N(0, \Sigma_0^{(i)}),
$$

with

$$
\underset{e\times e}{\Sigma_0^{(i)}} = \begin{pmatrix} I_{p(p+1)} & 0_{p(p+1)\times p+1} \\ -(J_0^{(i)})^{-1} K_0^{(i)} & -(J_0^{(i)})^{-1} \end{pmatrix} \Omega_0^{(i)} \begin{pmatrix} I_{p(p+1)} & -(J_0^{(i)})^{-1}(K_0^{(i)})' \\ 0_{p+1\times p(p+1)} & -(J_0^{(i)})^{-1} \end{pmatrix}, \tag{A.11}
$$

where $\Omega_0^{(i)}$ is defined in Lemma B.6, and $J_T^{(i)}(\tilde{\theta}^{(i)}) \overset{a.s.}{\to} J_0^{(i)}$, $K_T^{(i)}(\tilde{\theta}^{(i)}) \overset{a.s.}{\to} K_0^{(i)}$ by Lemma B.8, with

$$
\underset{p+1\times p+1}{J_0^{(i)}} = E\left[\left.\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial \kappa^{(i)} \partial \kappa^{(i)\prime}}\right|_{\theta^{(i)}=\theta_0^{(i)}}\right], \qquad \underset{p+1\times p^2+p}{K_0^{(i)}} = E\left[\left.\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial \kappa^{(i)} \partial \gamma'}\right|_{\theta^{(i)}=\theta_0^{(i)}}\right]. \tag{A.12}
$$

□

A.3 Proof of Corollary 1

The asymptotic distribution of the vector $\phi^{(i)} = [w_i, a_{i1}, \dots, a_{ip}, b_i]'$ can be found using the delta method, for which we need the partial derivative of $\phi^{(i)}$ with respect to the parameter vector $\theta^{(i)}$,

$$\underset{p+2\times e}{\varphi_0^{(i)}} = \left.\frac{\partial \phi^{(i)}}{\partial \theta^{(i)\prime}}\right|_{\theta^{(i)}=\theta_0^{(i)}} = \begin{pmatrix} (1-b_{0,i})\mathbb{I}\{i\} - A'_{0,i} & 0_{1\times p^2} & -\lambda'_0 & -\lambda_{0,i} \\ 0_{p\times p} & 0_{p\times p^2} & I_p & 0_{p\times 1} \\ 0_{1\times p} & 0_{1\times p^2} & 0_{1\times p} & 1 \end{pmatrix} \tag{A.13}$$

where $\mathbb{I}\{i\}$ is a $1 \times p$ vector of zeros, apart from a 1 in the i'th column and A_i is the i' row of A. Hence, the asymptotic distribution of $\phi^{(i)}$ is,

$$\sqrt{T}(\hat{\phi}^{(i)} - \phi_0^{(i)}) \xrightarrow{D} N(0, \varphi_0^{(i)} \Sigma_0^{(i)} \varphi_0^{(i)\prime}). \tag{A.14}$$

□

B Lemmata

B.1 Lemmata for the proof of consistency

Lemma B.1 (Finite expectation of likelihood contributions)**.** *Under Assumptions 4.1-4.4,*

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}} \left| l_t^{(i)}(\gamma, \kappa^{(i)})\right|\right] \leq K,$$

where $l_t^{(i)}(\gamma, \kappa^{(i)})$ *is defined in* (A.2).

Proof. Notice that the i'th conditional eigenvalue may be rewritten as an ARCH(∞) process,

$$\lambda_{i,t}(\theta^{(i)}) = w_i + \sum_{j=1}^{p} a_{ij} y_{j,t-1}^2(\theta^{(i)}) + b_i \lambda_{i,t-1}(\theta^{(i)}) = \sum_{l=0}^{\infty} b_i^l \left(w_i + \sum_{j=1}^{p} a_{ij} y_{j,t-l-1}^2(\theta^{(i)}) \right).$$

Using $\rho(B) < 1$ (By Assumption 4.1 and Lemma 4.1 of Ling and McAleer (2003)), along with Theorem 9.2 of Jacod and Protter (2012),

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}} \left|\log(\lambda_{i,t}(\theta^{(i)}))\right|\right] \leq E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}} \left|\lambda_{i,t}(\theta^{(i)})\right|\right] \leq K \sum_{t=1}^{\infty} \phi^t (1 + E||X_t||^2) \leq K.$$

Furthermore,

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{\lambda_{i,t}(\theta^{(i)})}\right|\right] \leq \sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{w_i}\right| \leq K, \tag{B.1}$$

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|y_{i,t}^2(\theta^{(i)})\right|\right] \leq E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|(V_i'(\theta^{(i)})X_t)^2\right|\right] \leq KE||X_t||^2 \leq K. \tag{B.2}$$

This, along with the triangle inequality, means that the log-likelihood contribution for the i'th rotated return is then bounded by a constant by Assumption 4.1,

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|l_t^{(i)}(\gamma,\kappa^{(i)})\right|\right] \leq E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\lambda_{i,t}(\theta^{(i)})\right|\right] + E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})}\right|\right] \leq K.$$

□

Lemma B.2 (Uniform convergence of likelihood function)**.** *Under Assumptions 4.1-4.4,*

$$\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|L_T^{(i)}(\gamma,\kappa^{(i)}) - E[l_t^{(i)}(\gamma,\kappa^{(i)})]\right| \overset{a.s.}{\to} 0,$$

where $L_T^{(i)}(\theta^{(i)})$ *is the log-likelihood function for the i'th rotated return and* $l_t^{(i)}(\theta^{(i)})$ *is the log-likelihood contribution for the i'th rotated return at time t, stated in* (A.1) *and* (A.2)*.*

Proof. Follows from Lemma B.1 and the uniform law of large numbers (Theorem A.2.2. of White (1994)) □

Lemma B.3 (Likelihood uniquely minimized)**.** *Under Assumptions 4.1-4.4,*

$$E|l_t^{(i)}(\gamma_0,\kappa_0^{(i)})| < \infty,$$

and if $\kappa^{(i)} \neq \kappa_0^{(i)}$*,*

$$E[l_t^{(i)}(\gamma_0,\kappa^{(i)})] > E[l_t^{(i)}(\gamma_0,\kappa_0^{(i)})],$$

where $l_t^{(i)}(\gamma,\kappa^{(i)})$ *is defined in* (A.2)*.*

Proof. The first statement follows directly from Lemma B.1. For the second statement, consider

$$\begin{aligned}
E[l_t^{(i)}(\gamma_0,\kappa^{(i)})]-&E[l_t^{(i)}(\gamma_0,\kappa_0^{(i)})]\\
&=E\left[\log\left(\frac{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}\right)+y_{i,t}^2(\gamma_0)\left(\frac{1}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}-\frac{1}{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}\right)\right]\\
&=E\left[\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})\left(\frac{1}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}-\frac{1}{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}\right)-\log\left(\frac{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}\right)\right]\\
&=E\left[\frac{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}-1-\log\left(\frac{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}\right)\right]\geq 0,
\end{aligned}$$

which uses $y_{i,t}(\gamma_0)=\lambda_{i,t}^{1/2}(\gamma_0,\kappa_0^{(i)})z_{i,t}$, where $z_{i,t}$ is *iid* with unit variance. Notice that $\log x\leq x-1$ for $x>0$, and that $\log x=x-1$ only if $x=1$. This inequality is strict unless $\frac{\lambda_{i,t}(\gamma_0,\kappa_0^{(i)})}{\lambda_{i,t}(\gamma_0,\kappa^{(i)})}=1$, which is ruled out, as it violates Assumption 4.3 on identification. □

Lemma B.4 (Asymptotic irrelevance of initial values)**.** *Under Assumptions 4.1-4.4,*

$$\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\gamma_0,\kappa^{(i)})-L_{T,h}^{(i)}(\hat\gamma,\kappa^{(i)})\right|\overset{a.s.}{\to}0,$$

where $L_t^{(i)}(\gamma,\kappa^{(i)})$ *is defined in* (A.1) *and* $L_{t,h}^{(i)}(\gamma,\kappa^{(i)})$ *is defined in* (A.5)*.*

Proof. We want to show that the initial values are asymptotically irrelevant. As in the proof of Theorem 4.1 in Pedersen (2016), we use the triangle inequality as follows,

$$\begin{aligned}
&\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\gamma_0,\kappa^{(i)})-L_{T,h}^{(i)}(\hat\gamma,\kappa^{(i)})\right|\leq\\
&\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\gamma_0,\kappa^{(i)})-L_T^{(i)}(\hat\gamma,\kappa^{(i)})\right|+\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\hat\gamma,\kappa^{(i)})-L_{T,h}^{(i)}(\hat\gamma,\kappa^{(i)})\right|.
\end{aligned}\tag{B.3}$$

In line with the aforementioned proof, we apply the mean-value theorem to the first term of (B.3),

$$\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}|L_T^{(i)}(\gamma_0,\kappa^{(i)})-L_T^{(i)}(\hat\gamma,\kappa^{(i)})|\leq\sum_{j=1}^{p(p+1)}(\hat\gamma_j-\gamma_{0,j})\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta^{(i)}\in\tilde{\mathcal{L}}\times\mathcal{V}\times\mathcal{K}^{(i)}}\left|\frac{\partial l_t(\gamma,\kappa^{(i)})}{\partial\gamma_j}\right|,\tag{B.4}$$

where $\tilde{\mathcal{L}}$ is chosen to be a compact subset of $(0,\infty)^p$ such that $(I_p-A-B)\lambda\in(0,\infty)^p$ and bounded away from zero on $\tilde{\mathcal{L}}\times\mathcal{V}$, with λ_0 in the interior of $\tilde{\mathcal{L}}$. An expression for $\partial l_t(\gamma,\kappa^{(i)})/\partial\theta^{(i)}$ can be found in (A.3) along with derivatives of $\lambda_{i,t}(\theta^{(i)})$ in (B.23)-(B.25) in Lemma B.7. Notice also that $E[\sup_{\gamma\in\mathcal{L}\times\mathcal{V}}||\partial y_{i,t}^2(\gamma)/\partial\gamma||]\leq KE||X_t||^2$. By Assumption 4.1 and $\rho(B)<1$ (follows Assumption 4.1, see Lemma 4.1 in Ling and McAleer (2003)) $\sup_{\theta^{(i)}\in\tilde{\mathcal{L}}\times\mathcal{V}\times\mathcal{K}^{(i)}}|\partial\lambda_{i,t}(\theta^{(i)})/\partial\gamma_j|<\infty$. This, along with $\sup_{\theta^{(i)}\in\tilde{\mathcal{L}}\times\mathcal{V}\times\mathcal{K}^{(i)}}|1/\lambda_{i,t}(\kappa^{(i)})|\leq K$, ensure that $E\left[\sup_{\theta^{(i)}\in\tilde{\mathcal{L}}\times\mathcal{V}\times\mathcal{K}^{(i)}} \partial l_t^{(i)}(\gamma,\kappa^{(i)})/\partial\gamma_j|\right] < \infty$. By the ergodic theorem and (A.8)-(A.9) we find that, $\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\gamma_0,\kappa^{(i)}) - L_T^{(i)}(\hat{\gamma},\kappa^{(i)})\right| \overset{a.s.}{\to} 0$.

Next, consider the second term of (B.3),

$$\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\hat{\gamma},\kappa^{(i)}) - L_{T,h}^{(i)}(\hat{\gamma},\kappa^{(i)})\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|b_i^t(\lambda_{i,0}(\hat{\gamma},\kappa^{(i)}) - \lambda_{i,0,h}(\hat{\gamma},\kappa^{(i)}))\right| +$$
$$\frac{1}{T}\sum_{t=1}^{T}K\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|||X_t||^2\left[\frac{1}{\lambda_{i,t,h}(\hat{\gamma},\kappa^{(i)})}\left(\lambda_{i,t,h}(\hat{\gamma},\kappa^{(i)}) - \lambda_{i,t}(\hat{\gamma},\kappa^{(i)})\right)\frac{1}{\lambda_{i,t}(\hat{\gamma},\kappa^{(i)})}\right]\right| \leq$$
$$K\frac{1}{T}\sum_{t=1}^{T}\phi^t(1+||X_t||^2) \tag{B.5}$$

by $\log x \leq x-1$ for $x \geq 1$ along with (A.7), (B.1) and $\sup_{\gamma\in\mathcal{L}\times\mathcal{V}}|V_i'X_t| \leq K||X_t||$. Additionally, we use that for any $j \geq 0$, $\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}|b_i^j| \leq K\phi^j$ (see e.g., p.616 of Francq and Zakoïan (2004) or p.611 of Francq, Horvath, and Zakoïan (2011)), along with A.9 and the compactness of $\mathcal{K}^{(i)}$ for $T\to\infty$,

$$K\frac{1}{T}\sum_{t=1}^{T}\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|b_i^t(\lambda_{i,0}(\hat{\gamma},\kappa^{(i)}) - \lambda_{i,0,h}(\hat{\gamma},\kappa^{(i)}))\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t \qquad \text{a.s.} \tag{B.6}$$

Hence,

$$\sup_{\kappa^{(i)}\in\mathcal{K}^{(i)}}\left|L_T^{(i)}(\hat{\gamma},\kappa^{(i)}) - L_{T,h}^{(i)}(\hat{\gamma},\kappa^{(i)})\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t(1+||X_t||^2) \qquad \text{a.s.}$$

For any $\varepsilon > 0$, we use Markov's inequality and Assumption 4.1,

$$\sum_{t=1}^{\infty}P(\phi^t||X_t||^2 > \varepsilon) \leq \sum_{t=1}^{\infty}\frac{\phi^t E||X_t||^2}{\varepsilon} < \infty,$$

Next, by the Borel-Cantelli lemma, $\phi^t||X_t||^2 \overset{a.s.}{\to} 0$, and finally, by Cesaro's mean theorem

$$\frac{1}{T}\sum_{t=1}^{T}\phi^t||X_t||^2 \overset{a.s.}{\to} 0,$$

we conclude that the initial values are asymptotically irrelevant for consistency of the estimator. □

B.2 Lemmata for the proof of asymptotic normality

Lemma B.5 (Rewriting the two-step estimator in vector form)**.** *Under Assumptions 4.1-4.6, for $T \to \infty$, the two-step estimator can be written jointly as,*

$$\sqrt{T}\begin{pmatrix} \hat{\gamma} - \gamma_0 \\ \frac{\partial L_T^{(i)}(\theta_0)}{\partial \kappa^{(i)}} \end{pmatrix} =$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix} D(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)(\Lambda_{0,t}^{1/2})^{\otimes 2} vec(Z_t Z_t' - I_p) \\ V_{0,1}' \otimes (\lambda_{0,1} I_p - H_0)^{+} V_0^{\otimes 2}(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)(\Lambda_{0,t}^{1/2})^{\otimes 2} vec(Z_t Z_t' - I_p) \\ \vdots \\ V_{0,p}' \otimes (\lambda_{0,p} I_p - H_0)^{+} V_0^{\otimes 2}(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)(\Lambda_{0,t}^{1/2})^{\otimes 2} vec(Z_t Z_t' - I_p) \\ -\frac{1}{\lambda_{i,t}(\theta_0^{(i)})}\frac{\partial \lambda_{i,t}(\theta_0^{(i)})}{\partial \kappa^{(i)}}(z_{i,t}^2 - 1) \end{pmatrix} + o_p(1). \tag{B.7}$$

where $(\cdot)^{+}$ is the Moore-Penrose inverse, D is $p \times p^2$ with all elements zero, apart from one element on each row, which is $d_{j,j+(j-1)p} = 1$ for $j = 1, \ldots, p$ and $\tilde{A}$ and $\tilde{B}$ are given in (B.13).

Proof. In rewriting the estimator in vector form, we partly follow Pedersen and Rahbek (2014) (proof of Lemma B.8) and rewrite $\text{vec}(\hat{H}) - \text{vec}(H_0)$ in terms of the GARCH parameters. The remainder of the proof is distinctly different, as we have to recast the vector of dynamic eigenvalues, λ_t, in a BEKK$(p^2, 1, 1, 1)$ parameterization, and state the first step estimator in terms of $\hat{\gamma} - \gamma_0$ rather than $\text{vec}(\hat{H}) - \text{vec}(H_0)$.

First, consider the moment estimator of the unconditional covariance matrix,

$$\hat{H} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t' = \frac{1}{T}\sum_{t=1}^{T} H_{0,t}^{1/2} Z_t Z_t' (H_{0,t}^{1/2})' = \frac{1}{T}\sum_{t=1}^{T} V_0 \Lambda_{0,t}^{1/2} Z_t Z_t' \Lambda_{0,t}^{1/2} V_0'.$$

Recall that $Y_{0,t} = V_0' X_t$ and define $\hat{\Lambda} = \frac{1}{T}\sum_{t=1}^{T} Y_{0,t} Y_{0,t}' = \frac{1}{T}\sum_{t=1}^{T} \Lambda_{0,t}^{1/2} Z_t Z_t' \Lambda_{0,t}^{1/2}$, such that,

$$\begin{aligned} \text{vec}(\hat{\Lambda}) = (V_0')^{\otimes 2}\text{vec}\left(\hat{H}\right) &= \text{vec}\left(\frac{1}{T}\sum_{t=1}^{T} \Lambda_{0,t}^{1/2} Z_t Z_t' \Lambda_{0,t}^{1/2}\right) \\ &= \left(\frac{1}{T}\sum_{t=1}^{T}(\Lambda_{0,t}^{1/2})^{\otimes 2}\text{vec}(Z_t Z_t' - I_p) + \text{vec}\left(\frac{1}{T}\sum_{t=1}^{T}\Lambda_{0,t}\right)\right). \end{aligned} \tag{B.8}$$

Next, we need to rewrite the conditional eigenvalues in a "vec"-reparameterization. That is, we first write the dynamics of $\Lambda_{0,t}$ to be nested in the BEKK-GARCH, and then we apply the vec-operator to obtain the vec-parameterization of the conditional eigenvalues.

Hence,

$$\Lambda_{0,t} = C + \sum_{i=1}^{p^2} A_{i,0} Y_{0,t-1} Y_{0,t-1}' A_{i,0}' + B_{1,0} \Lambda_{0,t-1} B_{1,0}', \tag{B.9}$$

with $C = (\Lambda_0 - \sum_{i=1}^{p^2} A_{i,0}\Lambda_0 A_{i,0}' + B_{1,0}\Lambda_0 B_{1,0}')$, and A_i are restricted parameter matrices, e.g., for the bivariate case,

$$A_1 = \begin{pmatrix} \sqrt{a_{11}} & 0 \\ 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & \sqrt{a_{12}} \\ 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 \\ \sqrt{a_{21}} & 0 \end{pmatrix}, \quad A_4 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{a_{22}} \end{pmatrix}, \tag{B.10}$$

and B_1 is

$$B_1 = B^{1/2}. \tag{B.11}$$

The vec-reparameterization is therefore,

$$\text{vec}(\Lambda_{0,t}) = (I_{p^2} - \tilde{A}_0 - \tilde{B}_0)\text{vec}(\Lambda_0) + \tilde{A}_0 \text{vec}(Y_{0,t-1}Y_{0,t-1}') + \tilde{B}_0 \text{vec}(\Lambda_{0,t-1}), \tag{B.12}$$

where

$$\tilde{A} = \sum_{i=1}^{p^2} A_i^{\otimes 2}, \qquad \tilde{B} = B_1^{\otimes 2}. \tag{B.13}$$

We now use this reparameterzation of the model to rewrite $\text{vec}\left(\frac{1}{T}\sum_{t=1}^T \Lambda_{0,t}\right)$ as follows,

$$\begin{aligned}
\text{vec}\left(\frac{1}{T}\sum_{t=1}^T \Lambda_{0,t}\right) &= \left(I_{p^2} - \tilde{A}_0 - \tilde{B}_0\right)\Lambda_0 + \tilde{A}_0 \text{vec}\left(\frac{1}{T}\sum_{t=1}^T Y_{0,t-1}Y_{0,t-1}'\right) + \tilde{B}_0 \text{vec}\left(\frac{1}{T}\sum_{t=1}^T \Lambda_{0,t-1}\right) \\
&= \left(I_{p^2} - \tilde{A}_0 - \tilde{B}_0\right)\Lambda_0 + \tilde{A}_0 \text{vec}\left(\frac{1}{T}\sum_{t=1}^T Y_{0,t}Y_{0,t}'\right) + \tilde{B}_0 \text{vec}\left(\frac{1}{T}\sum_{t=1}^T \Lambda_{0,t}\right) \\
&\quad + \frac{1}{T}\left(\tilde{A}_0(\text{vec}(Y_{0,0}Y_{0,0}') - \text{vec}(Y_{0,T}Y_{0,T}')) + \tilde{B}_0(\text{vec}(\Lambda_{0,0}) - \text{vec}(\Lambda_{0,T}))\right).
\end{aligned}$$

Collecting terms, and noting that $\hat{\Lambda} = \frac{1}{T}\sum_{t=1}^T Y_{0,t}Y_{0,t}'$, yields,

$$\begin{aligned}
\text{vec}\left(\frac{1}{T}\sum_{t=1}^T \Lambda_{0,t}\right) &= (I_{p^2} - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)\text{vec}(\Lambda_0) + (I_{p^2} - \tilde{B}_0)^{-1}\tilde{A}_0 \text{vec}(\hat{\Lambda}) \\
&\quad + \frac{1}{T}(I_{p^2} - \tilde{B}_0)^{-1}\left(\tilde{A}_0(\text{vec}(Y_{0,0}Y_{0,0}') - \text{vec}(Y_{0,T}Y_{0,T}') + \tilde{B}_0(\text{vec}(\Lambda_{0,0}) - \text{vec}(\Lambda_{0,T}))\right),
\end{aligned}$$

where $(I_{p^2} - \tilde{B}_0)$ is invertible since B is diagonal with $\rho(B_0) = \rho(\tilde{B}_0) < 1$. Insert this into

(B.8) and rearrange,

$$\text{vec}(\hat{\Lambda}) - \text{vec}(\Lambda_0) = \left(I_{p^2} - \tilde{A}_0 - \tilde{B}_0\right)^{-1} \left(I_{p^2} - \tilde{B}_0\right) \frac{1}{T} \sum_{t=1}^{T} (\Lambda_{0,t}^{1/2})^{\otimes 2} \text{vec}(Z_t Z_t' - I_p)$$
$$+ \frac{1}{T} \left(I_{p^2} - \tilde{A}_0 - \tilde{B}_0\right)^{-1} \left(\tilde{A}_0(\text{vec}(Y_{0,0}Y_{0,0}') - \text{vec}(Y_{0,T}Y_{0,T}')) + \tilde{B}_0(\text{vec}(\Lambda_{0,0}) - \text{vec}(\Lambda_{0,T}))\right)$$

By Markov's inequality it holds that for $\varepsilon > 0$,

$$P\left(\left|\left|\frac{1}{T}(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1} \left(\tilde{A}_0(\text{vec}(Y_{0,0}Y_{0,0}') - \text{vec}(Y_{0,T}Y_{0,T}')) + \tilde{B}_0(\text{vec}(\Lambda_{0,0}) - \text{vec}(\Lambda_{0,T}))\right)\right|\right| > \varepsilon\right)$$
$$\leq \frac{KE||X_t||^2}{T\varepsilon} \to 0,$$

as $T \to \infty$. This yields,

$$\text{vec}(\hat{\Lambda}) - \text{vec}(\Lambda_0) = (I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)\frac{1}{T}\sum_{t=1}^{T}(\Lambda_{0,t}^{1/2})^{\otimes 2}\text{vec}(Z_t Z_t' - I_p) + o_p(1).$$

Recall that $\text{vec}(\hat{H}) - \text{vec}(H_0) = V_0^{\otimes 2}\left(\text{vec}(\hat{\Lambda}) - \text{vec}(\Lambda_0)\right)$, and we find that,

$$\sqrt{T}(\text{vec}(\hat{H}) - \text{vec}(H_0)) =$$
$$V_0^{\otimes 2}(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\Lambda_{0,t}^{1/2})^{\otimes 2}\text{vec}(Z_t Z_t' - I_p) + o_p(1). \tag{B.14}$$

As the model is parameterized in terms of the eigenvalues and -vectors, we now restate (B.14) in terms of λ and υ. Notice that $\hat{\lambda} - \lambda_0 = D(V_0')^{\otimes 2}\left(\text{vec}(\hat{H}) - \text{vec}(H_0)\right)$, where D is a $p \times p^2$ matrix of zeros, apart from p elements, $d_{i,i+(i-1)p} = 1$ for $i = 1, \ldots, p$, such that $D\text{vec}(\Lambda) = \lambda$, and we find that,

$$\sqrt{T}(\hat{\lambda} - \lambda_0) = D(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1}(I_{p^2} - \tilde{B}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\Lambda_{0,t}^{1/2})^{\otimes 2}\text{vec}(Z_t Z_t' - I_p) + o_p(1). \tag{B.15}$$

Next, since υ does not have a closed form solution as a function of H, we apply the mean-value theorem, and use the following result from Magnus (1985) (Theorem 1),

$$\frac{\partial V_j}{\partial \text{vec}(H)} = V_j' \otimes (\lambda_j I_p - H)^+, \tag{B.16}$$

where $(\lambda_j I_p - H)^+$ is the Moore-Penrose (pseudo-) inverse of $(\lambda_j I_p - H)$. From this, we can apply the mean-value theorem to the j'th eigenvector,

$$\begin{aligned}\sqrt{T}(\hat{V}_j - V_{0,j}) &= V_{0,j}' \otimes (\lambda_{0,j} I_p - H_0)^{+} \sqrt{T}(\mathrm{vec}(\hat{H}) - \mathrm{vec}(H_0)) \\ &\quad + \underbrace{(V_{0,j}' \otimes (\lambda_{0,j} I_p - H_0)^{+} - \tilde{V}_j' \otimes (\tilde{\lambda}_j I_p - \tilde{H}_0)^{+})\sqrt{T}(\mathrm{vec}(\hat{H}) - \mathrm{vec}(H_0))}_{=o_p(1)} \\ &= V_{0,j}' \otimes (\lambda_{0,j} I_p - H_0)^{+} \sqrt{T}(\mathrm{vec}(\hat{H}) - \mathrm{vec}(H_0)) + o_p(1), \qquad j = 1, \dots, p,\end{aligned} \tag{B.17}$$

where $\tilde{H} = \tilde{V}\tilde{\Lambda}\tilde{V}'$ is on the line between H_0 and $\hat{H}$. Hence, by (B.14) and (B.17),

$$\sqrt{T}(\hat{\upsilon} - \upsilon_0) =$$
$$\begin{pmatrix} V_{0,1}' \otimes (\lambda_{0,1} I_p - H_0)^{+} V_0^{\otimes 2} (I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1} (I_{p^2} - \tilde{B}_0) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Lambda_{0,t}^{1/2})^{\otimes 2} \mathrm{vec}(Z_t Z_t' - I_p) \\ \vdots \\ V_{0,p}' \otimes (\lambda_{0,p} I_p - H_0)^{+} V_0^{\otimes 2} (I_{p^2} - \tilde{A}_0 - \tilde{B}_0)^{-1} (I_{p^2} - \tilde{B}_0) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\Lambda_{0,t}^{1/2})^{\otimes 2} \mathrm{vec}(Z_t Z_t' - I_p) \end{pmatrix} + o_p(1). \tag{B.18}$$

Finally, note that

$$\frac{\partial l_t^{(i)}(\theta_0^{(i)})}{\partial \kappa^{(i)}} = -\frac{1}{\lambda_{i,t}(\theta_0^{(i)})} \frac{\partial \lambda_{i,t}(\theta_0^{(i)})}{\partial \kappa^{(i)}} (z_{i,t}^2 - 1). \tag{B.19}$$

Hence, by (B.15), (B.18) and (B.19), we conclude that (B.7) holds. □

Lemma B.6 (Joint normality of parameter vector)**.** *Under Assumptions 4.1-4.6, for $T \to \infty$,*

$$\sqrt{T} \begin{pmatrix} \hat{\gamma} - \gamma_0 \\ \frac{\partial L_T^{(i)}(\theta_0)}{\partial \kappa^{(i)}} \end{pmatrix} \xrightarrow{D} N(0, \Omega_0^{(i)}), \tag{B.20}$$

with $\Omega_0^{(i)}$ defined in (B.22).

Proof. Similar to the variance targeting literature (e.g., Pedersen and Rahbek (2014) proof of Lemma B.8 and B.9), we use that (B.7) is a martingale difference sequence (asymptotically) to show convergence in distribution. From (B.7), define

$$\underset{e \times e}{\mathcal{Y}_t}(\theta_0^{(i)}) := \begin{pmatrix} \mathcal{Y}_{1,t}(\theta_0^{(i)}) \\ \mathcal{Y}_{2,t}(\theta_0^{(i)}) \\ \mathcal{Y}_{3,t}(\theta_0^{(i)}) \end{pmatrix},$$

which has elements,

$$\underset{p\times 1}{\mathcal{Y}_{1,t}}(\theta_0^{(i)}) = D(V_0')^{\otimes 2}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t,$$

$$\underset{p^2\times 1}{\mathcal{Y}_{2,t}}(\theta_0^{(i)}) = \begin{pmatrix} \chi_{0,1}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t \\ \vdots \\ \chi_{0,p}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t \end{pmatrix},$$

$$\underset{(p+1)\times 1}{\mathcal{Y}_{3,t}}(\theta_0^{(i)}) = -\frac{1}{\lambda_{i,t}(\theta_0^{(i)})}\frac{\partial\lambda_{i,t}(\theta_0^{(i)})}{\partial\kappa^{(i)}}(z_{i,t}^2 - 1),$$

where we use the following definitions,

$$\varepsilon_t := \mathrm{vec}(Z_t Z_t' - I_p),$$

$$\Gamma_0 := (V_0)^{\otimes 2}\left[(I_{p^2} - \tilde{A}_0 - \tilde{B}_0)\right]^{-1}(I_{p^2} - \tilde{B}_0),$$

$$\chi_{0,i} := V_{0,i}' \otimes (\lambda_{0,i} I_p - H_0)^+,$$

Notice that

$$E\left[\mathcal{Y}(\theta_0^{(i)})\right] = 0_{p(p+2)+1\times 1}, \tag{B.21}$$

since Z_t is *iid* with $E[Z_t Z_t'] = I_p$.

Next, consider the covariance matrix,

$$\underset{e\times e}{\Omega_0}^{(i)} := E\left[\mathcal{Y}_t(\theta_0^{(i)})\mathcal{Y}_t(\theta_0^{(i)})'\right] = E\left[\begin{pmatrix} \Omega_{0,11}^{(i)} & \Omega_{0,12}^{(i)} & \Omega_{0,13}^{(i)} \\ (\Omega_{0,12}^{(i)})' & \Omega_{0,22}^{(i)} & \Omega_{0,23}^{(i)} \\ (\Omega_{0,13}^{(i)})' & (\Omega_{0,23}^{(i)})' & \Omega_{0,33}^{(i)} \end{pmatrix}\right], \tag{B.22}$$

with

$$\Omega_{0,11}^{(i)} = D(V_0')^{\otimes 2}\Gamma\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\varepsilon_t\varepsilon_t'\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\Gamma'(V_0)^{\otimes 2}D',$$

$$\Omega_{0,22}^{(i)} = \begin{pmatrix} \chi_{0,1}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,1}' & \dots & \chi_{0,1}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,p}' \\ \vdots & \ddots & \vdots \\ \chi_{0,p}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,1}' & \dots & \chi_{0,p}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t\varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,p}' \end{pmatrix},$$

$$\Omega_{0,33}^{(i)} = \frac{1}{\lambda_{i,t}^2(\theta_0^{(i)})}\frac{\partial\lambda_{i,t}(\theta_0^{(i)})}{\partial\kappa^{(i)}}\frac{\partial\lambda_{i,t}(\theta_0^{(i)})}{\partial\kappa^{(i)\prime}}(z_{i,t}^2-1)^2,$$

$$\Omega_{0,12}^{(i)} = D(V_0')^{\otimes 2}\Gamma\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\varepsilon_t\left(\varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,1}' \quad \dots \quad \varepsilon_t'(\Lambda_{0,t}^{1/2})^{\otimes 2}\Gamma_0'\chi_{0,p}'\right),$$

$$\Omega_{0,13}^{(i)} = D(V_0')^{\otimes 2}\Gamma\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\varepsilon_t\frac{1}{\lambda_{i,t}(\theta_0^{(i)})}\frac{\partial\lambda_{i,t}(\theta_0^{(i)})}{\partial\kappa^{(i)\prime}}(1-z_{i,t}^2),$$

$$\Omega_{0,23}^{(i)} = \begin{pmatrix} \chi_{0,1}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t \\ \dots \\ \chi_{0,p}\Gamma_0(\Lambda_{0,t}^{1/2})^{\otimes 2}\varepsilon_t \end{pmatrix}\frac{1}{\lambda_{i,t}(\theta_0^{(i)})}\frac{\partial\lambda_{i,t}(\theta_0^{(i)})}{\partial\kappa^{(i)\prime}}(1-z_{i,t}^2),$$

which are $p\times p$, $p^2\times p^2$, $p+1\times p+1$, $p\times p^2$, $p\times p+1$ and $p^2\times p+1$ respectively.

To show that $\mathcal{Y}_t(\theta_0^{(i)})$ is square integrable, we verify that all elements of $\Omega_0^{(i)}$ are finite. By independence of Z_t and $\Lambda_{0,t}$,

$$E\left[||\Omega_{0,11}||\right] \le KE\left[\left|\left|\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\right|\right|^2\right]E\left[||\varepsilon_t||^2\right],$$

$$E\left[||\Omega_{0,22}||\right] \le KE\left[\left|\left|\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\right|\right|^2\right]E\left[||\varepsilon_t||^2\right],$$

and using the euclidean matrix norm, $||A|| = \sqrt{\text{tr}(A'A)}$ along with $\text{tr}(A\otimes B) = \text{tr}(A)\text{tr}(B)$ (for A and B square),

$$E\left[\left|\left|\left(\Lambda_{0,t}^{1/2}\right)^{\otimes 2}\right|\right|^2\right] = E\left[\text{tr}\left(\Lambda_{0,t}\right)^2\right] = E\left[\left(\sum_{i=1}^{p}\lambda_{i,t}(\theta_0^{(i)})\right)^2\right] \le K < \infty,$$

by Assumption 4.5. Moreover,

$$E\left[||\varepsilon_t||^2\right] \le E\left[||Z_t||^4\right] + K < \infty,$$

as $E\left[||Z_t||^4\right] \le KE\left[||X_t||^4\right]$. Hence $E\left[||\Omega_{0,11}^{(i)}||\right] < \infty$ and $E\left[||\Omega_{0,22}^{(i)}||\right] < \infty$. Next,

$$E\left[||\Omega_{0,33}^{(i)}||\right] \le K < \infty,$$

by Assumption 4.5. Finally, $E\left[||\Omega_{0,12}^{(i)}||\right]$, $E\left[||\Omega_{0,13}^{(i)}||\right]$ and $E\left[||\Omega_{0,23}^{(i)}||\right]$ are finite by the

Cauchy-Schwarz inequality.

Because $\mathcal{Y}_t(\theta_0^{(i)})$ is a square integrable ergodic martingale difference sequence, we can invoke the central limit theorem for martingale differences, see e.g., Brown (1971), implying that (B.20) holds. □

Lemma B.7 (Finite expectation of second order derivative)**.** *Under Assumptions 4.1-4.6*

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)}\partial\theta^{(i)\prime}}\right|\right] < \infty.$$

Proof. The second order derivative of $l_{i,t}(\theta^{(i)})$ is given in (A.4), and we note that in order to show that $E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial\theta^{(i)}\partial\theta^{(i)\prime}}\right|\right] < \infty$, if suffices to verify that,

1. $E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})}\frac{1}{\lambda_{i,t}(\theta^{(i)})}\frac{\partial^2\lambda_{i,t}(\theta^{(i)})}{\partial\theta^{(i)}\partial\theta^{(i)\prime}}\right|\right] < \infty$,
2. $E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\left(\frac{\partial^2 y_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}}y_{i,t}(\theta^{(i)}) + \frac{\partial y_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\frac{\partial y_{i,t}(\theta^{(i)})}{\partial\theta_m^{(i)}}\right)\frac{1}{\lambda_{i,t}(\theta^{(i)})}\right|\right] < \infty$,
3. $E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\left(\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\frac{\partial y_{i,t}(\theta^{(i)})}{\partial\theta_m^{(i)}} + \frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_m^{(i)}}\frac{\partial y_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\right)\frac{y_{i,t}(\theta^{(i)})}{\lambda_{i,t}^2(\theta^{(i)})}\right|\right] < \infty$,
4. $E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{y_{i,t}^2(\theta^{(i)})}{\lambda_{i,t}(\theta^{(i)})}\frac{1}{\lambda_{i,t}^2(\theta^{(i)})}\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_m^{(i)}}\right|\right] < \infty$,

for $n, m = 1, \ldots, p(p+1)+p+1$. Inequalities 1.-3. are finite by Assumption 4.5. The last inequality requires finite $2+s$ moments for $s > 0$, as,

$$E\left[\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\frac{1}{\lambda_{i,t}(\theta^{(i)})}\right|\right] \leq K.$$

To see this, recall that the process for the eigenvalues can be written as,

$$\lambda_{i,t} = \sum_{l=0}^{\infty} b_i^l\big(w_i + \sum_{j=1}^{p} a_{ij}y_{j,t-l-1}^2(\theta^{(i)})\big) = \frac{w_i}{1-b_i} + \sum_{l=0}^{\infty}\sum_{j=1}^{p} b_i^l a_{ij}y_{j,t-l-1}^2(\theta^{(i)})$$

with $w_i = (1-b_i)\lambda_i - \sum_{j=1}^{p} a_{ij}\lambda_j$ and $\theta^{(i)} = [\gamma', a_{i1}, \ldots, a_{ip}, b_i]'$. The derivatives of the eigenvalues are,

$$\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\gamma_k} = \frac{\partial w_i(1-b_i)^{-1}}{\partial\gamma_k} + 2\sum_{l=0}^{\infty}\sum_{j=1}^{p} b_i^l a_{ij}y_{j,t-l-1}(\theta^{(i)})\frac{\partial V_i{}'}{\partial\gamma_k}X_t, \tag{B.23}$$

$$\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial a_{ij}} = \frac{\partial w_i(1-b_i)^{-1}}{\partial a_{ij}} + \sum_{l=0}^{\infty} b_i^l y_{j,t-l-1}^2(\theta^{(i)}), \tag{B.24}$$

$$\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial b_i} = \frac{\partial w_i(1-b_i)^{-1}}{\partial b_i} + \sum_{l=1}^{\infty}\sum_{j=1}^{p} l b_i^{l-1} a_{ij}y_{j,t-l-1}^2(\theta^{(i)}), \tag{B.25}$$

for $k = 1, \dots, p(p+1)$ and $i, j = 1, \dots, p+1$. By Assumption 4.6 along with the inequality $z/(1+z) \leq z^s$ for $s \in (0,1)$ for all $z \geq 0$, for all interior $\theta^{(i)} \in \Theta^{(i)}$,

$$\begin{aligned}
\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \gamma_j} \right| &\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \sum_{l=0}^{\infty} \frac{\sum_{j=1}^{p} b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})}{K + \sum_{j=1}^{p} b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})} \right| \\
&\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \sum_{l=0}^{\infty} \sum_{j=1}^{p} b_i^{ls} a_{ij}^s y_{j,t-l-1}^{2s}(\theta^{(i)}) \right| \leq K, \qquad \text{(B.26)}
\end{aligned}$$

using $\sup_{\theta^{(i)} \in \Theta^{(i)}} |\frac{\partial y_{i,t-1}^2(\theta^{(i)})}{\partial \gamma_j}| = \sup_{\theta^{(i)} \in \Theta^{(i)}} |2V_i(\theta^{(i)})' X_{t-1} \frac{\partial V_i'(\theta^{(i)})}{\partial \gamma_j} X_{t-1}| \leq \sup_{\theta^{(i)} \in \Theta^{(i)}} |K y_{i,t-1}^2|$, due to the orthonormality of V_i. Similarly,

$$\begin{aligned}
\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial a_{ij}} \right| &\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \frac{1}{a_{ij}} \sum_{l=0}^{\infty} \frac{b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})}{K + b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})} \right| \\
&\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \sum_{l=0}^{\infty} b_i^{ls} a_{ij}^s y_{j,t-l-1}^{2s}(\theta^{(i)}) \right| \leq K, \qquad \text{(B.27)} \\
\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{\lambda_{i,t}(\theta^{(i)})} \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial b_i} \right| &\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \frac{1}{b_i} \sum_{l=1}^{\infty} l \frac{\sum_{j=1}^{p} b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})}{K + \sum_{j=1}^{p} b_i^l a_{ij} y_{j,t-l-1}^2(\theta^{(i)})} \right| \\
&\leq \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| K + K \sum_{l=1}^{\infty} \sum_{j=1}^{p} l b_i^{sl} a_{ij}^s y_{j,t-l-1}^{2s}(\theta^{(i)}) \right| \leq K, \qquad \text{(B.28)}
\end{aligned}$$

such that (B.26)-(B.28) are finite when $\{X_t\}_{t \in \mathbb{Z}}$ is stationary, ergodic and has finite fractional moments. Hence the last inequality holds under Assumptions 4.1-4.6. □

Lemma B.8 (Uniform convergence of second order derivative)**.** *Under Assumptions 4.1-4.6, for $T \to \infty$,*

$$\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{\partial^2 L_t^{(i)}(\theta^{(i)})}{\partial \theta^{(i)} \partial \theta^{(i)\prime}} - E\left[\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial \theta^{(i)} \partial \theta^{(i)\prime}} \right] \right| \overset{a.s.}{\to} 0.$$

Proof. Since $\frac{\partial^2 l_{i,t}(\theta^{(i)})}{\partial \theta^{(i)} \partial \theta^{(i)\prime}}$ is a function of $(X_t, X_{t-1}, \dots)$ and $\theta^{(i)}$, it is strictly stationary and ergodic. The result then follows by Lemma B.7 and the uniform law of large numbers for stationary and ergodic processes, see Theorem A.2.2 of White (1994). □

Lemma B.9 (Non-singular $J_0^{(i)}$)**.** *Under Assumption 4.1-4.6, $J_0^{(i)}$, given in* (A.12)*, is non-singular.*

Proof. $J_0^{(i)}$ is identical to the Hessian of a univariate (extended) GARCH model (for the rotated returns). The non-singularity of $J_0^{(i)}$ therefore follows from Berkes, Horváth, and Kokoszka (2003) Lemma 5.7.

□

Lemma B.10 (Asymptotic irrelevance of (fixed) initial values). *Under Assumptions 4.1-4.6,*

$$\left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\partial l_t^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} - \frac{\partial l_{t,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} \right) \right| \xrightarrow{p} 0 \tag{B.29}$$

for $n = 1, \dots, p+1$, and

$$\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial \theta_n^{(i)} \partial \theta_m^{(i)}} - \frac{\partial^2 l_{t,h}^{(i)}(\theta^{(i)})}{\partial \theta_n^{(i)} \partial \theta_m^{(i)}} \right) \right| \xrightarrow{p} 0, \tag{B.30}$$

for $n, m = 1, \dots, p(p+1) + p + 1$.

Proof. Consider first (B.29) concerning the elements of the score vector. By (A.3), the triangle inequality and $\sup_{\theta^{(i)} \in \Theta^{(i)}} |y_{i,t}| = \sup_{\gamma \in \mathcal{H}} |V_i'(\gamma) X_t| \le K ||X_t||$,

$$\begin{aligned}
&\left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\partial l_t^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} - \frac{\partial l_{t,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} \right) \right| \le \\
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left| \frac{1}{\lambda_{i,t}(\theta_0^{(i)})} \frac{\partial \lambda_{i,t}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} - \frac{1}{\lambda_{i,t,h}(\theta_0^{(i)})} \frac{\partial \lambda_{i,t,h}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} \right| + \\
&K \frac{1}{\sqrt{T}} \sum_{t=1}^{T} ||X_t||^2 \left| \frac{1}{\lambda_{i,t}(\theta_0^{(i)})} \frac{\partial \lambda_{i,t}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} - \frac{1}{\lambda_{i,t,h}(\theta_0^{(i)})} \frac{\partial \lambda_{i,t,h}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} \right|,
\end{aligned}$$

for $n = 1, \dots, p+1$. Following Francq, Horvath, and Zakoïan (2011) (p. 649),

$$\begin{aligned}
&\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{\partial \lambda_{i,t}(\theta^{(i)})}{\partial \theta_n^{(i)}} - \frac{\partial \lambda_{i,t,h}(\theta^{(i)})}{\partial \theta_n^{(i)}} \right| = \\
&\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{\partial b_i^t}{\partial \theta_n^{(i)}} \left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)}) \right) + b_i^t \frac{\partial \left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)}) \right)}{\partial \theta_n^{(i)}} \right| \le K \phi^t,
\end{aligned} \tag{B.31}$$

as both $b_i^t \to 0$ and $\partial b_i^t / \partial b_i \to 0$ for $t \to \infty$. Furthermore, notice that

$$\sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{\lambda_{i,t}(\theta^{(i)})} - \frac{1}{\lambda_{i,t,h}(\theta^{(i)})} \right| \le \sup_{\theta^{(i)} \in \Theta^{(i)}} \left| \frac{1}{w_i} - \frac{1}{w_i} \right| \le K. \tag{B.32}$$

Using (B.31)-(B.32), we find that B.29 can be bounded as,

$$\left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\partial l_t^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} - \frac{\partial l_{t,h}^{(i)}(\theta_0^{(i)})}{\partial \kappa_n^{(i)}} \right) \right| \le K \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi^t \left(1 + ||X_t||^2 \right).$$

As $\sum_{t=1}^{T} \phi^t \to (1-\phi)^{-1}$ it holds that $KT^{-1/2} \sum_{t=1}^{T} \phi_t \to 0$ for $T \to \infty$, and by the Markov inequality for $\varepsilon > 0$,

$$P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi^t(1+||X_t||^2) > \varepsilon\right) \leq \varepsilon^{-1}(1+E\,||X_t||^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi^t \xrightarrow{p} 0,$$

as $E||X_t||^2 < \infty$ by Assumption 4.1. This proves the first statement of the lemma.

Next, we consider the requirement in (B.30). The second order derivative of the log-likelihood function is given in (A.4), and to show that the expression in (B.30) converges to zero in probability we need three additional results. First,

$$\begin{aligned}
&\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{\partial^2\lambda_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}} - \frac{\partial^2\lambda_{i,t,h}(\theta^{(i)})}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}}\right| = \\
&\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{\partial^2 b_i^t}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}}\left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)})\right) + b_i^t\frac{\partial^2\left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)})\right)}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}}\right. \\
&\qquad\left. + \frac{\partial b_i^t}{\partial\theta_n^{(i)}}\frac{\partial\left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)})\right)}{\partial\theta_m^{(i)}} + \frac{\partial b_i^t}{\partial\theta_m^{(i)}}\frac{\partial\left(\lambda_{i,0}(\theta^{(i)}) - \lambda_{i,0,h}(\theta^{(i)})\right)}{\partial\theta_n^{(i)}}\right| \leq K\phi^t,
\end{aligned} \tag{B.33}$$

for $n, m = 1, \ldots, p(p+1)+p+1$. Second, by (B.26)-(B.28),

$$\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{\lambda_{i,t}(\theta^{(i)})}\frac{\partial\lambda_{i,t}(\theta^{(i)})}{\partial\theta_n^{(i)}}\right| \leq K, \qquad \sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{\lambda_{i,t,h}(\theta^{(i)})}\frac{\partial\lambda_{i,t,h}(\theta^{(i)})}{\partial\theta_n^{(i)}}\right| \leq K. \tag{B.34}$$

Third,

$$\begin{aligned}
&\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{\lambda_{i,t}(\theta^{(i)})} - \frac{1}{\lambda_{i,t,h}(\theta^{(i)})}\right| = \\
&\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{\lambda_{i,t}(\theta^{(i)})}(\lambda_{i,t}(\theta^{(i)}) - \lambda_{i,t,h}(\theta^{(i)}))\frac{1}{\lambda_{i,t,h}(\theta^{(i)})}\right| \leq K\phi^t.
\end{aligned} \tag{B.35}$$

Hence, by (B.31)-(B.35) along with the triangle inequality, (B.30) can be bounded as,

$$\sup_{\theta^{(i)}\in\Theta^{(i)}}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 l_t^{(i)}(\theta^{(i)})}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}} - \frac{\partial^2 l_{t,h}^{(i)}(\theta^{(i)})}{\partial\theta_n^{(i)}\partial\theta_m^{(i)}}\right)\right| \leq K\frac{1}{T}\sum_{t=1}^{T}\phi^t\left(1+||X_t||^2+||X_t||^4\right),$$

which by Markov's inequality converges to zero in probability,

$$P\left(K\frac{1}{T}\sum_{t=1}^{T}\phi^t\left(1+||X_t||^2+||X_t||^4\right) > \varepsilon\right) \leq \varepsilon^{-1}\left(K(1+E[||X_t||^2]+E[||X_t||^4])\frac{1}{T}\sum_{t=1}^{T}\phi^t\right) \xrightarrow{p} 0,$$

which concludes the proof, and we conclude that the (fixed) initial values used in the estimation do not matter for $T \to \infty$.

□

C Stationarity, ergodicity and existence of moments

The following two lemmata provide sufficient and necessary conditions for strict stationarity, ergodicity, and finite moments for the multivariate GARCH model and are both stated without proofs. They were originally stated in the context of the extended CCC model but are also applicable in the present model, as the λ-GARCH, conditional on the eigenvectors, is an extended CCC model for the rotated returns.

Rewrite the process of the eigenvalues as a stochastic recurrence equation,

$$\lambda_t = W + \mathcal{A}_{t-1}\lambda_{t-1},$$

where $\mathcal{A}_{t-1} = A \text{ diag}\left(Z_{t-1}^{\odot 2}\right) + B$ is an *iid* $p \times p$ sequence for $t \in \mathbb{Z}$.

Lemma C.1 (Francq and Zakoïan (2019) Theorem 10.6)**.** *A necessary and sufficient condition for the existence of a unique, non-anticipative, strictly stationary and ergodic solution to the process* $(X_t : t \in \mathbb{Z})$ *is* $\gamma < 0$*, with* γ *defined as the top Lyapunov coefficient,* $\gamma = \lim_{t\to\infty} \frac{1}{t} E\left[\log || \prod_{i=1}^{t} \mathcal{A}_i||\right]$

Notice that Lemma C.1 only ensures the existence of fractional moments, $E||X_t||^s < \infty$, $0 < s < 1$. We next restate a result from Pedersen (2017) (Proposition 2.1), which contain necessary and sufficient conditions for finite (non-fractional) moments.

Lemma C.2 (Pedersen (2017) Proposition 2.1)**.** *Let* $(X_t : t \in \mathbb{Z})$ *denote a strictly stationary and ergodic process. Then* $E\left[||X_t^{\odot 2}||^k\right] < \infty$*,* $k \in \mathbb{N}$ *if and only if* $\rho\left(E\left[\mathcal{A}_t^{\otimes k}\right]\right) < 1$*.*

D Simulation study

This appendix illustrates the theoretical results through simulations: we simulate the large-sample distribution of the STE in three cases: In the first, we illustrate the sufficiency of finite fourth order moments, and show that both steps of the STE are consistent and asymptotically normal when $E||X_t||^4 < \infty$. The second case considers the distribution of the STE when the data-generating process (DGP) does not admit finite fourth order moments, but rather has finite second order moments, $E||X_t||^2 < \infty$, $E||X_t||^4 = \infty$, indicating that the STE should be consistent, but have a non-normal limiting distribution. Finally, the third simulation considers the STE when the DGP only admits a finite mean, $E||X_t|| < \infty$, $E||X_t||^k = \infty$ for $k = 2, 4$.

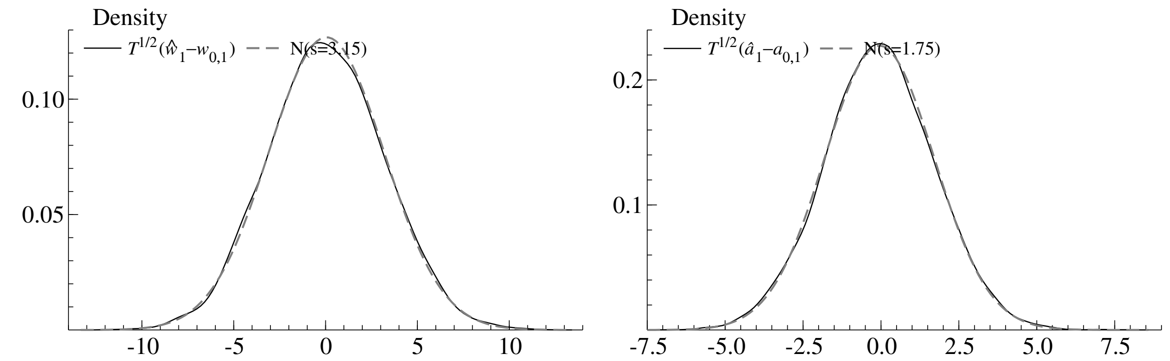


FIGURE D.1: Densities of estimated parameters when $E||X_t||^4 < \infty$. The solid line is the estimated density, and the grey dashed line is the normal distribution.

D.1 Case 1: The DGP satisfies the sufficient condition for asymptotic normality

Consider the bivariate λ-GARCH with Gaussian innovations,

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

with parameters

$$V_0 = \begin{pmatrix} 0.89 & 0.45 \\ -0.45 & 0.89 \end{pmatrix}, \quad W_0 = \begin{pmatrix} 1.5 \\ 0.46 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0.33 & 0 \\ 0 & 0.25 \end{pmatrix}, \quad B_0 = 0_{2\times 2}, \qquad \text{(D.2)}$$

such that $\rho(E[A \text{ diag}(Z_t^2) + B]^{\otimes 2}) = \max\left(a_{ii}^2 E[z_{i,t}^4]\right) < 1$ for $i = 1, 2$. For Z_t iid $N(0, I_2)$ this corresponds to $\max(a_i) < 1/\sqrt{3}$, and by Lemma C.2 (with $k = 2$), the stationary solution of the process has finite fourth order moments, and the moment restrictions of Theorem 4.2 are satisfied.

We simulate $N = 10.000$ realizations of (D.1)-(D.2) with $T = 10.000$ observations, and estimate W and A using STE. Figure D.1 contains standardized densities of w_1 and a_{11}. The figure suggests that the STE is indeed consistent and asymptotically normal, in line with the findings in Theorem 4.2.

D.2 Case 2: The DGP satisfies the sufficient condition for consistency

Next, we consider the case where the DGP has finite second order moments, $E||X_t||^2 < \infty$ but does not admit finite fourth order moments, $E||X_t||^4 = \infty$. We consider (D.1), with parameters

$$V_0 = \begin{pmatrix} 0.89 & 0.45 \\ -0.45 & 0.89 \end{pmatrix}, \quad W_0 = \begin{pmatrix} 1.5 \\ 0.46 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 0.60 & 0 \\ 0 & 0.55 \end{pmatrix}, \quad B_0 = 0_{2\times 2}, \qquad \text{(D.3)}$$

such that $\rho(E[A_0 \operatorname{diag}(Z_t^2) + B_0]) = \max\left(a_{0,ii}^2\right) < 1$ for $i = 1, 2$. By Lemma C.2 (with $k = 1$), the stationary solution of the process admits finite second order moments, and the moment restrictions for asymptotic normality (Theorem 4.2) are not satisfied. However, by Theorem 4.1, the estimator should be consistent.

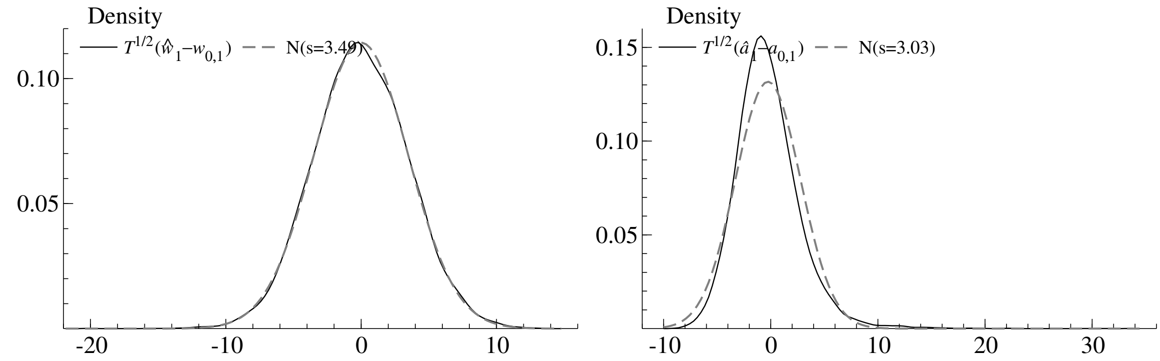


FIGURE D.2: Densities of estimated parameters when $E||X_t||^4 < \infty$. The solid line is the estimated density, and the grey dashed line is the normal distribution.

We simulate $N = 10.000$ realizations of (D.1) and (D.3) with $T = 10.000$ observations, and estimate W and A using STE. Figure D.2 contains standardized densities of w_1 and a_{11}. The figure suggests that in this case, the estimator is indeed consistent, but it is not asymptotically normal. Surprisingly, the density of w_1 seem to behave almost like a normal distribution, albeit with a heavy left tail, whereas that of a_{11} is clearly non-normal. This is similar to the findings of Pedersen and Rahbek (2014), who consider variance-targeting in the BEKK-model, and we conclude that $E||X_t||^4 < \infty$ is a necessary condition for the joint normality of the ST estimator.

D.3 Case 3: The DGP does not satisfy the sufficient condition for consistency

Finally, we consider the case where the DGP only admits a finite mean, $E||X_t|| < \infty$. Here we set parameter matrices to

$$V_0 = \begin{pmatrix} 0.89 & 0.45 \\ -0.45 & 0.89 \end{pmatrix}, \quad W_0 = \begin{pmatrix} 1.5 \\ 0.46 \end{pmatrix}, \quad A_0 = \begin{pmatrix} 1.01 & 0 \\ 0 & 0.90 \end{pmatrix}, \quad B_0 = 0_{2\times 2}. \qquad \text{(D.4)}$$

As $\max(a_{0,ii}) < \pi/2$ for $i = 1, 2$ the DGP is strictly stationary, ergodic and has a finite mean, but does not admit any higher order moments (by Lemma C.2). As before, we simulate $N = 10.000$ realizations of (D.1) and (D.4) with $T = 10.000$ observations, and estimate W and A using STE. Figure D.3 contains standardized densities of w_1 and a_{11}. Clearly, the estimator is neither consistent nor asymptotically normal when $E||X_t||^2 = \infty$, and we conclude that the moment condition in Theorem 4.1 is indeed necessary for consistency of the estimator.

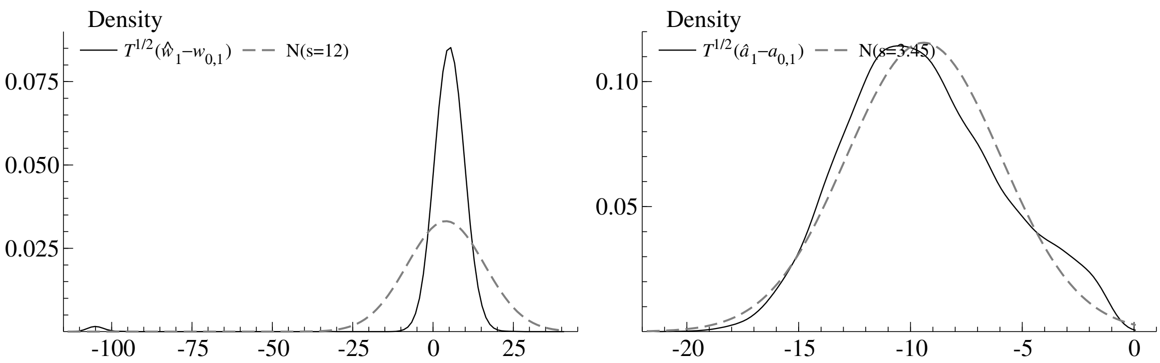


FIGURE D.3: Densities of estimated parameters when $E||X_t||^4 < \infty$. The solid line is the estimated density, and the grey dashed line is the normal distribution.

E Empirical exercise: Portfolio constituents and weights

Table E.1: Portfolio constituents and weights.

Bloomberg ticker	Company name	Portfolio weights				
		P_1	P_2	P_3	P_4	P_5
DIS US Equity	Walt Disney Co.	0.040	0.016	0.010	0.068	- 0.347
HD US Equity	Home Depot	0.040	0.002	0.085	- 0.082	- 0.099
ABT US Equity	Abbott	0.040	0.042	0.071	- 0.195	- 1.099
CVX US Equity	CV Sciences	0.040	0.066	0.007	- 0.038	- 1.243
EXC US Equity	Exelon	0.040	0.038	0.057	0.137	0.671
MCD US Equity	McDonalds	0.040	0.044	0.048	- 0.257	- 0.122
MMM US Equity	3M	0.040	0.001	0.101	- 0.282	0.318
AAPL US Equity	Apple	0.040	0.045	0.086	0.159	0.621
UNH US Equity	United Health Group	0.040	0.086	0.072	- 0.118	- 0.210
TXN US Equity	Texas Instruments	0.040	0.078	0.066	0.118	- 0.342
JPM US Equity	JPMorgan Chase	0.040	0.067	0.113	0.138	- 0.856
IBM US Equity	IBM	0.040	0.010	0.118	0.255	0.760
DVN US Equity	Devon Energy	0.040	0.082	0.040	- 0.059	0.861
GD US Equity	General Dynamics	0.040	0.078	0.029	- 0.083	0.606
CPB US Equity	Campbell Soup Company	0.040	0.015	0.095	0.180	0.970
PEP US Equity	PepsiCo	0.040	0.058	0.008	0.156	0.383
MRK US Equity	Merck & Co.	0.040	0.002	0.043	0.090	0.507
NKE US Equity	NantKwest	0.040	0.011	0.008	0.183	- 0.794
COST US Equity	Costco	0.040	0.029	0.038	0.084	0.692
T US Equity	AT&T	0.040	0.014	0.008	0.269	- 0.429
RF US Equity	Regions Financial Corporation	0.040	0.069	0.035	- 0.098	- 0.160
SLB US Equity	Schlumberger	0.040	0.075	0.094	0.140	- 0.468
PG US Equity	Procter & Gamble	0.040	0.032	0.108	0.265	0.177
HON US Equity	Honeywell	0.040	0.029	0.094	- 0.097	0.410
HPQ US Equity	Hewlett-Packard	0.040	0.010	0.067	0.066	0.195

Chapter 3

Bootstrap-Based Inference and Testing in Multivariate Dynamic Conditional Eigenvalue GARCH Models

ABSTRACT

We study fixed-design bootstrap for quasi-maximum likelihood estimation of multivariate GARCH processes. Specifically, we extend the univariate bootstrap of Cavaliere, Pedersen, and Rahbek (2018) to the Dynamic Conditional Eigenvalue GARCH model of Hetland, Pedersen, and Rahbek (2020). We show, under fairly mild conditions, that the bootstrap Wald test statistic is consistent, conditional on the original data. The theoretically investigated fixed-design bootstrap is contrasted to a recursive bootstrap, and the asymptotic test statistic. Through Monte Carlo simulations, we find evidence that the fixed-design bootstrap is superior to the recursive bootstrap and the asymptotic test in small samples. In larger samples, both bootstrap designs and the asymptotic test share properties, as expected from the asymptotic theory. An empirical application illustrates the empirical merits of the bootstrap in multivariate GARCH models. The appealing theoretical properties, along with the excellent finite sample properties, suggest that the fixed-design bootstrap is an important tool for small sample inference in multivariate GARCH models.

KEYWORDS: Multivariate GARCH, fixed-design bootstrap, fixed regressor bootstrap, hypothesis testing.
JEL: C32, C58.

1 Introduction

Multivariate GARCH models are a popular choice for modeling the conditional covariance matrix of financial returns. They are typically applied in the context of Markowitz type optimization procedures (see e.g., Engle, Ledoit, and Wolf, 2019) and for the estimation of market risk models, such as value-at-risk and expected shortfall (see e.g., Francq and Zakoïan, 2020). In both cases, inference is an integral part of empirical applications of multivariate GARCH models.

In small samples, inference based on asymptotic distributions may lead to an incorrect nominal size. As well known, see e.g., Cavaliere, Nielsen, and Rahbek (2017), Cavaliere, Pedersen, and Rahbek (2018), and Cavaliere et al. (2020), such small sample issues can be corrected by the bootstrap. The bootstrap is a simulation-based approach in which we generate new samples of the data and the test statistic of interest, and use these to compute finite-sample critical values.

This paper extends the univariate fixed-design (or fixed volatility) bootstrap of Cavaliere, Pedersen, and Rahbek (2018) of ARCH(q)-type models to the multivariate Dynamic Conditional Eigenvalue GARCH (λ-GARCH) model. The λ-GARCH was introduced in Hetland, Pedersen, and Rahbek (2020) as a generalized autoregressive score model (Creal, Koopman, and Lucas, 2011, 2013) for the conditional eigenvalues, and the authors establish full asymptotic theory of the quasi-maximum likelihood estimator (QMLE). The λ-GARCH is closely related to the GO-GARCH of Weide (2002) (see also Lanne and Saikkonen, 2007, Fan, Wang, and Yao, 2008 and Boswijk and Weide, 2011), as discussed in detail in Hetland, Pedersen, and Rahbek (2020). A general introduction to multivariate GARCH models can be found in e.g., Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009). For a treatment of stochastic properties along with estimation of multivariate GARCH models, see Francq and Zakoïan (2019) (chapter 10).

To our knowledge, we are the first to consider fixed-design bootstrap for multivariate GARCH processes in the context of full QMLE, and we explore the theoretical and numerical properties of this bootstrap. Within the realm of GARCH models, the fixed-design bootstrap is one of two popular residual bootstrap methods, the other being the recursive bootstrap: the recursive bootstrap generates new data recursively using the estimated dynamics, whereas the fixed-design bootstrap keeps lagged variables fixed at the (observed) sample values. That is, the recursive bootstrap generates new paths for the conditional eigenvalues in each simulation, whereas the fixed-design bootstrap keeps these fixed across all simulations, such that the only source of variation is the re-sampling of standardized residuals.

The fixed-design bootstrap has previously been studied in a univariate (G)ARCH context: Shimizu (2010) considers a one-step Newton-Raphson estimator, while Cavaliere, Pedersen, and Rahbek (2018) consider likelihood based tests for a class of ARCH(q)

models. Beutner, Heinemann, and Smeekes (2020) study a fixed-design bootstrap in the context of two-step estimation of value-at-risk, based on univariate GARCH models, while Cavaliere et al. (2020) consider a shrinkage-based fixed-design bootstrap for testing hypotheses on the boundary of the parameter space (see also Cavaliere, Nielsen, and Rahbek (2017)). Recursive bootstrap methods, in the context of univariate GARCH models, have been studied in e.g., Hall and Yao (2003), Pascual, Romo, and Ruiz (2006), Hidalgo and Zaffaroni (2007) and Jeong (2017). Considering multivariate GARCH models, Francq, Horvath, and Zakoïan (2014) apply the one-step Newton-Raphson fixed-design methodology from Shimizu (2010) in variance targeting estimation of the CCC-GARCH, while Hetland, Pedersen, and Rahbek (2020) use a recursive bootstrap to test parameter restrictions in the λ-GARCH. In contrast, our exposition is a complete asymptotic analysis based on the QMLE.

We provide a full asymptotic analysis of the fixed-design bootstrap for the λ-GARCH model, and show that the bootstrap Wald test statistic is consistent. That is, we establish validity of the bootstrap Wald test statistic and show that the bootstrap mimics the correct asymptotic distribution under both the null and alternative hypotheses. While the theoretical exposition assumes stationarity, ergodicity and finite fourth order moments of the return vector, our simulations imply that only a finite mean of the return process (in addition to finite fourth order moments for the residuals) are necessary for bootstrap consistency. Furthermore, our simulations indicate that the fixed-design bootstrap has excellent nominal coverage compared with both the recursive bootstrap and the asymptotic test statistic in small samples. As expected from the asymptotic theory, all three statistics perform comparably for larger sample sizes.

The remainder of the paper is as follows. Section 2 introduces the λ-GARCH, stochastic properties and estimation by QMLE. Section 3 present the fixed-design bootstrap, implementation and theoretical results, and Section 4 contain Monte Carlo simulations, while Section 5 considers a small empirical application. Finally, Section 6 concludes. All proofs can be found in Appendix A.

1.1 Notation

We denote by $\mathbb{R}, \mathbb{R}_+, \mathbb{R}_{++}$ the real, positive real and strictly positive real numbers respectively. The absolute value of $a \in \mathbb{R}$ is denoted $|a|$. For $p, n \in \mathbb{N}$, I_p denotes the $(p \times p)$ identity matrix and $0_{n \times p}$ denotes a $n \times p$ matrix of zeros. The trace of a square matrix is denoted $\text{tr}(A)$, and the determinant $\det(A)$. For a $p-$dimensional vector x, $\text{diag}(x) = \text{diag}((x_i)_{i=1}^p)$ is a diagonal matrix with x on the diagonal. Denote by $\rho(A)$ the spectral radius of any square matrix A, i.e., $\rho(A) = \max\{|\tilde{\lambda}_i| : \tilde{\lambda}_i \text{ is an eigenvalue of A}\}$. We use $|| \cdot ||$ as a matrix norm. Elements of matrices or vectors are denoted by lower case letters, e.g., a_{ij} is the $(i,j)'$th element of the matrix A. In the context of the Hadamard product, we let $X^{\odot 2} = X \odot X$ for the vector X. Let P^* and E^* denote probability and expectation conditional on the original sample. Furthermore, $\xrightarrow{w^*}_p$ denotes weak convergence in probability, i.e., $X_T^* \xrightarrow{w^*}_p X$ means that, as the sample size T diverges, the cumulative distribution function G_T^* of X_T^* conditional on the original data, converges in probability to the cumulative distribution function G of X at all continuity points of G. Moreover, for some sequence X_T^* computed from the bootstrap data, $X_T^* \xrightarrow{p^*}_p X$ or $X_T^* - X = o_p^*(1)$, mean that for any $\varepsilon > 0$, $P^*(||X_T^* - X|| > \varepsilon) \xrightarrow{p} 0$ for $T \to \infty$. Similarly, $X_T^* = O_p^*(1)$ in probability, means that, for any $\varepsilon > 0$, there exists some constant $M > 0$ such that, for large T, $P(P^*(||X_T^*|| > M) < \varepsilon)$ is arbitrarily close to one.

2 The λ-GARCH Model | Properties and Estimation

We now present the λ-GARCH model. Let X_t be a $p \times 1$ vector of asset returns,

$$X_t = V\Lambda_t^{1/2} Z_t, \tag{2.1}$$

$$\Lambda_t = \text{diag}((\lambda_{i,t})_{i=1}^p), \tag{2.2}$$

$$\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{p,t})' = W + AY_{t-1}^{\odot 2} + B\lambda_{t-1}, \tag{2.3}$$

for $t = 1, \ldots, T$. The innovations, Z_t, are $i.i.d.(0, I_p)$ and θ is a vector of parameters (to be specified). Here, V is an orthonormal matrix, $VV' = V'V = I_p$, and we let $Y_t = V'X_t$ denote the orthogonalized returns which have conditional covariance Λ_t. Let W be a $p \times 1$ vector of strictly positive elements, $w_i > 0$ for $i = 1, \ldots, p$, and let A and B be $p \times p$ with non-negative entries, $\alpha_{ij}, \beta_{ij} \geq 0$ for $i, j = 1, \ldots, p$. Furthermore, the (constant conditional) eigenvectors V are parametrized by the $p(p-1)/2$ dimensional vector $\phi = [\phi_{12}, \ldots, \phi_{(p-1)p}]'$. Specifically, we parameterize V as a series of Givens rotation matrices (see also Pinheiro and Bates (1996) and Hetland, Pedersen, and Rahbek, 2020)

$$V = V(\phi) = \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} R(i,j;\phi), \tag{2.4}$$

where $R(i,j;\phi)$ is a $(p \times p)$ matrix with elements

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

$$[R(i,j;\phi)]_{ii} = [R(i,j;\phi)]_{jj} = \cos(\phi_{ij}), \qquad [R(i,j;\phi)]_{ij} = -[R(i,j;\phi)]_{ji} = \sin(\phi_{ij}).$$

Before we discuss estimation of the λ-GARCH, we briefly state a condition for strict stationarity and ergodicity of the process in (2.1)-(2.3).

Remark 2.1 (Stationarity and ergodicity)**.** *As detailed in Theorem 3.1 of Hetland, Pedersen, and Rahbek (2020), the* λ*-GARCH is strictly stationary and ergodic if and only if the top Lyapunov exponent,* γ*, defined as*

$$\gamma = \lim_{T\to\infty} T^{-1} E\left(\log || \prod_{t=1}^{T} A \ \mathrm{diag}((z_{i,t}^2)_{i=1}^p) + B||\right), \tag{2.5}$$

is strictly negative, $\gamma < 0$*.*

In the following, we consider QMLE based on the Gaussian probability density function, which has log-likelihood, up to a constant,

$$L_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta), \tag{2.6}$$

$$l_t(\theta) = \log(\det(\Lambda_t(\theta))) + Y_t'(\theta)\Lambda_t^{-1}(\theta)Y_t(\theta). \tag{2.7}$$

The parameters of the model are given by $\theta = [W', \mathrm{vec}(A)', \mathrm{vec}(B)', \phi']'$ which has dimension $d_\theta = p + 2p^2 + p(p-1)/2$. Note that Λ_0 and $Y_0(\theta) = V'(\theta)X_0$ are fixed in the statistical analysis.

The QMLE, $\hat{\theta}_T$, is defined as,

$$\hat{\theta}_T = \arg\min_{\theta\in\Theta} L_T(\theta), \tag{2.8}$$

where Θ is the parameter space,

$$\Theta = \Theta_W \times \Theta_A \times \Theta_B \times \Theta_\phi, \tag{2.9}$$

with $\Theta_W \subset \mathbb{R}^p_{++}, \Theta_A \subset \mathbb{R}^{p^2}_{+}, \Theta_B \subset \mathbb{R}^{p^2}_{+}$ and $\Theta_\phi \subset \mathbb{R}^{p(p-1)/2}$.

To investigate the stochastic properties of the QMLE we make the following assumption on the parameter space Θ in (2.9).

Assumption 2.1. *The true value of the parameter vector* $\theta_0 = [W_0, vec(A_0)', vec(B_0)', \phi_0']'$ *belongs to* Θ*. Moreover, assume that* $\Theta_W = [w_L, w_U]^p$ *for some* $0 < w_L < w_U < \infty$*,* $\Theta_A = [0, a_U]^{p^2}$ *for some* $a_U < \infty$ *and* $\Theta_B \in \mathbb{R}^{p^2}_{+}$ *such that* $\sup_{vec(B)\in\Theta_B} \rho(B) < 1$*, and* $\Theta_\phi = [0, \phi/2]^{p(p-1)/2}$*.*

In particular, Assumption 2.1 implies that Θ is compact, and ensures that the rotation parameters in ϕ are identified. We also make the following assumption about the data-generating process, $\{X_t\}_{t\in\mathbb{Z}}$.

Assumption 2.2. *The process defined in* (2.1)-(2.3) *is strictly stationary and ergodic with* $E||X_t||^s < \infty$ *for some* $s > 0$*.*

Next, we state an assumption ensuring that the log-likelihood function is well-defined for $\theta \in \Theta$, along with a low-level identifying assumption.

Assumption 2.3. *The i.i.d. innovations, Z_t, has finite second order moments, $E||Z_t||^2 < \infty$.*

Assumption 2.4. *The $(p \times p^2)$ parameter matrix $[A_0, B_0]$ has full rank p.*

We now have the following result on consistency for the QMLE from Hetland, Pedersen, and Rahbek (2020) (Theorem 3.2).

Theorem 2.1 (Consistency of QMLE)**.** *Suppose that Assumptions 2.1-2.3 hold, then for $T \to \infty$, the QMLE in* (2.8) *is strongly consistent,*

$$\hat{\theta}_T \overset{a.s.}{\to} \theta_0.$$

In order to show that the estimator is asymptotically normal, we make the following additional assumptions.

Assumption 2.5. *θ_0 is an interior point of Θ, and the matrix A_0 has a row with a unique entry.*

Assumption 2.6. *The data-generating process satisfies $E||Z_t||^4 < \infty$ and $E||X_t||^{2+s} < \infty$ for $s > 0$.*

The moment requirements in Assumption 2.6 are sufficient conditions to ensure that the derivatives of the log-likelihood function in (2.6)-(2.7) are well-behaved in the limit.

We are now ready to state the following theorem from Hetland, Pedersen, and Rahbek (2020) (Theorem 3.3) which establishes asymptotic normality of the QMLE.

Theorem 2.2 (Asymptotic normality of QMLE)**.** *Suppose that Assumptions 2.1-2.6 hold. Then, for $T \to \infty$, the QMLE in* (2.8) *is asymptotically normal,*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{D}{\to} N(0, \Sigma),$$

with $\Sigma = J^{-1} V J^{-1}$, where $J = E\left[\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}\right]$ and $V = E\left[\frac{\partial l_t(\theta)}{\partial\theta}\frac{\partial l_t(\theta)}{\partial\theta'}\Big|_{\theta=\theta_0}\right]$.

Proofs of Theorem 2.1 and Theorem 2.2 are given in Hetland, Pedersen, and Rahbek (2020). Some additional comments on the assumptions. Assumptions 2.1-2.4 ensure that the λ-GARCH are uniquely identified, and that the QML estimator is well-defined and attains a unique minimum. The restrictions on Θ_ϕ, namely that all rotation parameters are restricted to the interval $[0, \pi/2]$ ensure that the eigenvalues cannot permute and has a fixed order. Assumption 2.2 on strict stationarity and ergodicity is needed to invoke a law of large numbers for the estimator. Finally, Assumptions 2.5-2.6 are needed to ensure that $\hat{J}_T$ is invertible and $\hat{V}_T$ is finite for $T \to \infty$.

In practice, we use the sample counterparts of J and V,

$$\hat{J}_T = T^{-1} \sum_{t=1}^{T} \left. \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} \right|_{\theta=\hat{\theta}_T} \qquad \hat{V}_T = T^{-1} \sum_{t=1}^{T} \left. \frac{\partial l_t(\theta)}{\partial\theta} \frac{\partial l_t(\theta)}{\partial\theta'} \right|_{\theta=\hat{\theta}_T} \tag{2.10}$$

and note that $\hat{\Sigma}_T := \hat{J}_T^{-1} \hat{V}_T \hat{J}_T^{-1} \overset{a.s.}{\to} \Sigma$.

Remark 2.2 (Moment requirement for QML estimation). *A simulation study in Hetland, Pedersen, and Rahbek (2020) indicates that the moment condition for X_t, outlined in Assumption 2.6 above, may not be necessary, and that stationarity and ergodicity of the process in* (2.1)-(2.3)*, in addition to $E||Z_t||^4 < \infty$, is sufficient and necessary for asymptotic normality of the QML estimator.*

In the following, we are interested in testing a hypothesis of the form,

$$H_0 : R\theta = r \tag{2.11}$$

where R is $k \times d_\theta$ of full rank, r is $k \times 1$, and $d_\theta \geq k \geq 1$, and k denotes the number of restrictions. Both R and r are known.

The hypothesis given in (2.11) can be tested using a standard Wald test, based on the statistic,

$$W_T = T(R\hat{\theta}_T - r)'(R\hat{\Sigma}_T R')^{-1}(R\hat{\theta}_T - r), \tag{2.12}$$

where $\hat{\Sigma}_T$ are defined in Theorem 2.2. Under H_0, Theorem 2.2 implies that,

$$W_T \overset{D}{\to} \chi^2(k) \tag{2.13}$$

where $\chi^2(k)$ denotes a chi-squared distribution with k degrees of freedom.

Alternatives to the Wald test statistic include the likelihood ratio (LR) and the Lagrange multiplier (LM) test statistics both of which are also based on the log-likelihood function. However, additional considerations are needed for the probability analysis of the LR and LM test statistics. In particular, both require the analysis of the estimator under the null hypothesis, complicating the theoretical exposition. In contrast, the Wald statistic only requires the analysis of the alternative (unrestricted) estimator. We conjecture, however, that the theory presented in this paper can, with relative ease, be extended to also cover the LR and LM statistics, by employing e.g., the exposition in Cavaliere, Nielsen, and Rahbek (2017) or Cavaliere et al. (2020). Recall that the Wald, LR and LM statistics are asymptotically equivalent, and the choice of test statistic therefore does not matter for $T \to \infty$.

Notice, that the theory presented here is not valid for testing hypotheses on the boundary of the parameter space, examples of which include testing for no-ARCH ($A_0 = B_0 = 0$) or no eigenvalue spillovers ($A_0 = \text{diag}(a_{0,i}), B_0 = \text{diag}(b_{0,i})$) for $i = 1, \ldots, p$). Bootstrap-based testing for no-ARCH in a univariate setup is covered in Cavaliere, Nielsen, and Rahbek (2017) and Cavaliere et al. (2020), while Pedersen (2017) considers (non-bootstrap) testing for no volatility spill-overs in the extended CCC-GARCH. Extending our theory to allow for parameters on the boundary considerably complicates the exposition, and is left for future work. We can, however, test if the eigenvalues are integrated ($\rho(A_0 + B_0) = 1$), which is highly useful when forecasting the conditional covariance matrix.

3 The Fixed-Design Bootstrap

We now introduce the fixed-design bootstrap algorithm for the multivariate λ-GARCH, motivated by Cavaliere, Pedersen, and Rahbek (2018). The key feature of the fixed-design bootstrap is that the conditional eigenvalues, λ_t, are kept fixed, and are not random, conditional on the original data. Therefore, the re-sampled residuals serve as the only source of variation in the fixed-design bootstrap. This is in contrast to the recursive bootstrap, in which both the eigenvalues and innovations are generated recursively by drawing from the standardized estimated residuals.

The estimated residuals are given by,

$$\hat{Z}_t = \Lambda_t^{-1/2}(\hat{\theta}_T)V'(\hat{\theta}_T)X_t, \tag{3.1}$$

and can be obtained after computing the QMLE in (2.8). Similar to Francq and Zakoïan (2016), and analogously to the literature on univariate bootstrap inference in GARCH models (see e.g., Cavaliere, Nielsen, and Rahbek, 2017, Cavaliere, Pedersen, and Rahbek, 2018, or Cavaliere et al., 2020), we standardize the residuals as,

$$\hat{Z}_t^s = \hat{\Omega}_Z^{-1/2}(\hat{Z}_t - \bar{Z}), \tag{3.2}$$

where $\bar{Z} = \frac{1}{T}\sum_{t=1}^{T}\hat{Z}_t$ and $\hat{\Omega}_Z = \frac{1}{T}\sum_{t=1}^{T}(\hat{Z}_t - \bar{Z})(\hat{Z}_t - \bar{Z})'$.

Algorithm 1 (Fixed-design bootstrap). *The fixed-design bootstrap can be outlined as follows:*

1. *Generate the bootstrap series* $\{X_t^*\}_{t=1}^T$ *as* $X_t^* = \hat{V}\hat{\Lambda}_t^{1/2}Z_t^*$, *where* Z_t^* *are drawn independently with replacement from the standardized residuals,* $\{\hat{Z}_t^s\}_{t=1}^T$ *in* (3.2).

2. *Compute the bootstrap QMLE,*

$$\hat{\theta}_T^* = \arg\min_{\theta\in\Theta} L_T^*(\theta), \tag{3.3}$$

with

$$L_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log(\det(\Lambda_t(\theta))) + X_t^* V(\theta)\Lambda_t^{-1}(\theta)V(\theta)' X_t^*. \tag{3.4}$$

3. *Compute the bootstrap Wald test statistic, defined as,*

$$W_{T,b}^* = T(\hat{\theta}_T^* - \hat{\theta}_T)' R' (R\hat{\Sigma}_T^* R')^{-1} R(\hat{\theta}_T^* - \hat{\theta}_T), \tag{3.5}$$

where,

$$\hat{\Sigma}_T^* = \hat{J}_T^{*-1}\hat{V}_T^*\hat{J}_T^{*-1}, \quad \hat{J}_T^* = T^{-1}\sum_{t=1}^{T}\frac{\partial^2 l_t^*(\hat{\theta}_T^*)}{\partial\theta\partial\theta'}, \qquad \hat{V}_T^* = T^{-1}\sum_{t=1}^{T}\frac{\partial l_t^*(\hat{\theta}_T^*)}{\partial\theta}\frac{\partial l_t^*(\hat{\theta}_T^*)}{\partial\theta'}.$$

4. *Repeat steps 1.–3. $b = 1, \dots, B$ times. The p% critical value can then be computed as the $(1-p)$%'th quantile of the empirical cumulative distribution function of $\{W_{T,b}^*\}_{b=1}^B$.*

Notice that the bootstrap log-likelihood function in (3.4) differs from the log-likelihood function in (2.6)-(2.7) since $\Lambda_t(\theta)$ is a function of X_t and not X_t^*. The bootstrap p-value, defined as $p^* = 1 - P^*(W_{T,b}^* \leq x)$ (where $P^*(W_{T,b}^* \leq x)$ is the cumulative distribution function of $W_{T,b}^*$, conditional on the original data), can be approximated as $p_B^* = \frac{1}{B}\sum_{b=1}^{B} 1\{W_{T,b}^* > W_T\}$, noting that $p_B^* \overset{a.s.}{\to} p^*$ for $B \to \infty$, see also Cavaliere, Pedersen, and Rahbek (2018).

We now present a theorem on consistency of the bootstrap, which verifies that Algorithm 1 correctly mimics the (first order) asymptotic distribution of the fixed-design Wald test statistic under both the null and alternative.

Theorem 3.1 (Bootstrap consistency)**.** *Suppose that Assumptions 2.1-2.6 are satisfied, and in addition $E||X_t||^4 < \infty$. Then, under the null hypothesis, $H_0: \ R\theta = r$, as $T \to \infty$, $W_T^* \overset{w^*}{\to}_p \chi^2(k)$. Moreover, under the alternative, $H_A: \ R\theta \neq r$, $W_T^* \overset{w^*}{\to}_p \chi^2(k)$. That is, $W_T^* = O_p^*(1)$, and the bootstrap Wald test is consistent.*

The proof of bootstrap consistency of the Wald test follows that of Cavaliere, Pedersen, and Rahbek (2018), and can be outlined as follows. First, we show that the bootstrap estimator, $\hat{\theta}_T^*$ is consistent for θ_0, conditional on the original data. Then, we show that the bootstrap estimator, for $T \to \infty$, converges weakly in probability, that is $\sqrt{T}(\hat{\theta}_T^* - \hat{\theta}_T) \overset{w^*}{\to}_p N(0, \Sigma)$. Finally, we show that the bootstrap estimator of Σ, $\hat{\Sigma}_T^*$, is consistent under the null and alternative hypotheses.

Throughout, we use Lemma A.8, which verifies that a law of large numbers holds for the standardized residuals, such that $T^{-1}\sum_{t=1}^{T} E^*[\hat{z}_{i,t}^s]^k \overset{p}{\to} E[z_{i,t}]^k$ for $k = 1, \dots, 4$. Note also that in several lemmata, we invoke the bootstrap version of Markov's inequality,

$P^*(||X^*||^2 > \varepsilon) < E^*[||X_t^*||^4]/\varepsilon^2$, for $\varepsilon > 0$, $X^* \in \mathbb{R}$. This can potentially be relaxed by verifying the validity of a bootstrap law of large numbers using the (bootstrap) characteristic function as done in Cavaliere, Nielsen, and Rahbek (2017) (proof of Lemma B.3), which may lead to lesser moment requirements. We investigate this numerically in Section 4, where we find that X_t (and by extension X_t^*) need only have a finite first moment along with finite fourth moments of the innovations, Z_t, for the bootstrap Wald test to be consistent.

Before moving on to the numerical exercises, we present the recursive bootstrap. In the simulations below, we contrast the empirical performance of the fixed-design bootstrap to that of the recursive bootstrap, and to the asymptotic test statistic.

Algorithm 2 (Recursive bootstrap)**.** *The recursive bootstrap can be outlined as follows:*

1. *Generate the bootstrap series* $\{X_t^\star\}_{t=1}^T$ *as* $X_t^\star = \hat{V}\hat{\Lambda}_t^{1/2} Z_t^*$*, where* Z_t^* *are drawn independently with replacement from the standardized residuals,* $\{\hat{Z}_t^s\}_{t=1}^T$ *in* (3.2)*.*

2. *Estimate the parameters in* θ *using the bootstrap QMLE,*

$$\hat{\theta}_T^\star = \arg\min_{\theta\in\Theta} L_T^\star(\theta), \tag{3.6}$$

with

$$L_T^\star(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log(\det(\Lambda_t^\star(\theta))) + X_t^\star V(\theta)\Lambda_t^{\star-1}(\theta)V(\theta)' X_t^\star. \tag{3.7}$$

3. *Compute the bootstrap Wald test as,*

$$W_{T,b}^\star = T(\hat{\theta}_T^\star - \hat{\theta}_T)' R' (R\hat{\Sigma}_T^\star R')^{-1} R(\hat{\theta}_T^\star - \hat{\theta}_T), \tag{3.8}$$

where,

$$\hat{\Sigma}_T^\star = \hat{J}_T^{\star-1}\hat{V}_T^\star\hat{J}_T^{\star-1}, \quad \hat{J}_T^\star = T^{-1}\sum_{t=1}^{T}\frac{\partial^2 l_t^\star(\hat{\theta}_T^\star)}{\partial\theta\partial\theta'}, \qquad \hat{V}_T^\star = T^{-1}\sum_{t=1}^{T}\frac{\partial l_t^\star(\hat{\theta}_T^\star)}{\partial\theta}\frac{\partial l_t^\star(\hat{\theta}_T^\star)}{\partial\theta'}.$$

4. *Repeat steps 1.–3.* $b = 1,\dots,B$ *times. The* p% *critical value can then be computed as the* $1-p\%$*'th quantile of the empirical cumulative distribution function of* $\{W_{T,b}^\star\}_{b=1}^B$.

As already emphasized, the main difference between fixed-design bootstrap in Algorithm 1 and the recursive bootstrap in Algorithm 2, is that the recursive bootstrap has an additional source of variation stemming from the recursively updated eigenvalues in addition to the re-sampled innovations. That is, $\lambda_t^\star(\theta)$ is a (recursive) function of past bootstrap innovations through $X_{t-1}^\star$. It is important to note that the asymptotic properties of the recursive bootstrap in Algorithm 2 have not yet been established for multivariate GARCH models.

4 Monte Carlo Study

In this section we investigate the finite sample performance of the fixed-design Wald test statistic, which we compare to the recursive bootstrap Wald test statistic outlined in Algorithm 2 and the asymptotic test statistic in (2.13). We compare the empirical size and power of the test, i.e., the empirical rejection probabilities (ERPs) when the null hypothesis is respectively true and false.

We consider four different data-generating processes for the bivariate λ-GARCH model. Each of the four data-generating processes relax the assumptions needed to verify the consistency of the bootstrap. We find that while finite fourth order moments of the return process are sufficient (by Theorem 3.1), they do not seem necessary for consistency of the fixed-design Wald test. Rather, we conjecture that finite first order moments along with finite forth order moments of the innovations are sufficient for consistency of the fixed-design bootstrap Wald test. Furthermore, we find that the same result hold for the recursive bootstrap outlined in Algorithm 2.

For simplicity, we consider a restricted version of the model, which does not allow for spill-overs between the conditional eigenvalues. That is, A_0 and B_0 are diagonal. The data-generating process is,

$$X_t = V_0 \Lambda_{0,t}^{1/2} Z_t, \quad V_0 = \begin{pmatrix} \cos(0.50) & \sin(0.50) \\ -\sin(0.50) & \cos(0.50) \end{pmatrix}, \quad \Lambda_{0,t} = \operatorname{diag}(\lambda_{0,t}), \tag{4.1}$$

$$\lambda_{0,t} = \begin{pmatrix} \lambda_{0,1,t} \\ \lambda_{0,2,t} \end{pmatrix} = \begin{pmatrix} 0.10 + a_{0,1}\ y_{0,1,t-1}^2 + 0.80\ \lambda_{0,1,t-1} \\ 0.05 + 0.10\ y_{0,2,t-1}^2 + 0.85\ \lambda_{0,2,t-1} \end{pmatrix}. \tag{4.2}$$

The parameter vector of the data-generating process is $\theta_0 = [0.50, 0.10, 0.05, a_{0,1}, 0.10, 0.80, 0.85]$, with $t = 1, \ldots, T$. The innovations, Z_t, are drawn *iid* from either a standard bivariate normal distribution or a standardized bivariate Student's t distribution. We vary the value of $a_{0,1}$ throughout the experiments such that the process has finite fourth, second, and first order moments, while the remaining parameters are kept fixed in the data-generating process. We use Theorem 3.1 of Hetland, Pedersen, and Rahbek (2020) to ensure that $E||X_t||^k < \infty$ for $k \in (1, 2, 4)$.

We consider four different scenarios,

C_1: with $a_{0,1} = 0.15$ and Z_t *iid* $N(0, I_2)$ such that $E||X_t||^4 < \infty$.

C_2: with $a_{0,1} = 0.18$ and Z_t *iid* $N(0, I_2)$ such that $E||X_t||^2 < \infty$ and $E||X_t||^4 = \infty$.

C_3: with $a_{0,1} = 0.20$ and Z_t *iid* $N(0, I_2)$ such that $E||X_t|| < \infty$, $E||X_t||^2 = \infty$ and $E||X_t||^4 = \infty$.

C_4: with $a_{0,1} = 0.20$ and Z_t *iid* $t_v(0, I_2)$ with $v = 4.7$ such that $E||X_t|| < \infty$, $E||X_t||^2 = \infty$ and $E||X_t||^4 = \infty$. Furthermore, $E||Z_t||^4 < \infty$ while $E||Z_t||^6 = \infty$.

That is, in the first scenario, the data-generating process satisfies the sufficient and necessary conditions outlined in Theorem 3.1 for consistency of the bootstrap Wald test. In the remaining cases, we examine the performance of the bootstrap when the sufficient condition, $E||X_t||^4 < \infty$, is violated. In C_4 we consider the performance of the bootstrap when Z_t is drawn from a standardized t_v distribution with $v = 4.7$ degrees of freedom, satisfying the necessary condition of $E||Z_t||^4 < \infty$ from Assumption 2.6.

Throughout, we use $N = 1000$ Monte Carlo replications and $B = 399$ bootstrap repetitions to approximate the distribution of the Wald statistics. We consider samples sizes of $T \in (250,\ 500,\ 1000,\ 2000)$. All tests are at the nominal 10% significance level.

Table 4.1: Empirical rejection probabilities, power and size studies.

	Size			Power		
	Fixed	Recursive	Asymptotic	Fixed	Recursive	Asymptotic
T	C_1: $E\|\|X_t\|\|^4 < \infty$, $Z_t \sim N(0, I_p)$					
250	0.115	0.165	0.128	0.117	0.167	0.131
500	0.088	0.085	0.091	0.233	0.230	0.245
1000	0.087	0.078	0.078	0.525	0.522	0.517
2000	0.093	0.085	0.079	0.805	0.809	0.793
	C_2: $E\|\|X_t\|\|^2 < \infty$, $E\|\|X_t\|\|^4 = \infty$, $Z_t \sim N(0, I_p)$					
250	0.105	0.143	0.128	0.205	0.253	0.239
500	0.092	0.087	0.094	0.542	0.521	0.549
1000	0.094	0.087	0.083	0.850	0.848	0.851
2000	0.095	0.099	0.092	0.992	0.989	0.990
	C_3: $E\|\|X_t\|\| < \infty$, $E\|\|X_t\|\|^2 = \infty$, $E\|\|X_t\|\|^4 = \infty$, $Z_t \sim N(0, I_p)$					
250	0.103	0.149	0.125	0.289	0.345	0.345
500	0.099	0.091	0.104	0.727	0.691	0.736
1000	0.093	0.093	0.090	0.957	0.960	0.955
2000	0.095	0.095	0.089	1.000	1.000	1.000
	C_4: $E\|\|X_t\|\| < \infty$, $E\|\|X_t\|\|^2 = \infty$, $E\|\|X_t\|\|^4 = \infty$, $Z_t \sim t_v(0, I_p)$, $v = 4.7$					
250	0.128	0.190	0.174	0.108	0.185	0.171
500	0.138	0.134	0.165	0.254	0.265	0.357
1000	0.097	0.098	0.108	0.873	0.872	0.897
2000	0.115	0.112	0.127	0.618	0.621	0.690

"Fixed" refers to the fixed-design bootstrap, "recursive" to the recursive bootstrap and "asymptotic" to the asymptotic test. All tests are at the nominal 10% significance level.

The first three columns of Table 4.1 contain the ERPs under the null hypothesis. Here, we test if the parameter a_1 is equal to the true parameter, $a_{0,1}$, in all four scenarios, C_i,

$i = 1, \ldots, 4$. We find that the fixed-design bootstrap overall has a good coverage compared to the recursive bootstrap and the asymptotic test statistic, and indeed works very well for the short samples. Especially so when compared with the recursive bootstrap, which has remarkably bad coverage for $T = 250$ (even worse than the asymptotic test). As already noted, the moment requirement outlined in Theorem 3.1 appears to be sufficient, but not necessary for consistency of the bootstrap. Rather both the recursive and fixed-design bootstrap appear to be consistent for case C_4 where only the mean of the process is finite, $E||X_t|| < \infty$. Remarkably, the fixed-design bootstrap is by far preferred to the other tests in the scenario C_4 (with heavy-tailed innovations). We argue that this scenario is the most realistic, as the innovations from multivariate GARCH models based on financial time series are often heavy-tailed (see e.g., Bauwens, Laurent, and Rombouts, 2006, Section 3).

Columns 4-6 of Table 4.1 contain the ERPs under the alternative. In these simulations, the null hypothesis is kept fixed through the different scenarios, C_i, $i = 1, \ldots, 4$, and the tested hypothesis is $H_0 : \ a_1 = 0.10 \ (\neq a_{0,1})$. As expected, the empirical power is monotonically increasing increasing as the true $a_{0,1}$ get further away from the null hypothesis. The power of the tests is also increasing in T. We note that there are no substantial difference in power across the different tests.

Summing up, we find that the fixed-design bootstrap outperforms the recursive bootstrap and the asymptotic test for the samples of $T = 250$, while the fixed-design bootstrap, the recursive bootstrap and the asymptotic test statistic perform equally well for larger samples, $T \in (500,\ 1000,\ 2000)$. In terms of power, all three tests perform similarly across all simulations.

5 Empirical Illustration

We now consider a brief empirical illustration, based on data for Coca Cola and Pepsi (tickers "KO" and "PEP"). We consider a three-year period, from December 31st 2007 to December 31st 2010 ($T = 757$), covering the height of the financial crisis.

The log-returns are shown in Figure 5.1. There are no signs of autocorrelation in the returns, while the absolute values (not shown) of the returns are highly autocorrelated. Furthermore, the unconditional marginal densities are heavy tailed and the two time series are characterized by volatility clustering.

We are interested in testing the hypothesis that the conditional eigenvalues of the time series are integrated, $\rho(A_0+B_0) = 1$, which has implications for forecasting the conditional covariance matrix. For simplicity, we consider a diagonal version of the model, in which $A = \text{diag}(a_1, a_2)$ and $B = \text{diag}(b_1, b_2)$.

The estimated parameters from the λ-GARCH are contained in Table 5.1. We find that the process is asymptotically stationary and ergodic (albeit not significantly so,

Figure 5.1: Log-returns of Coca Cola and Pepsi

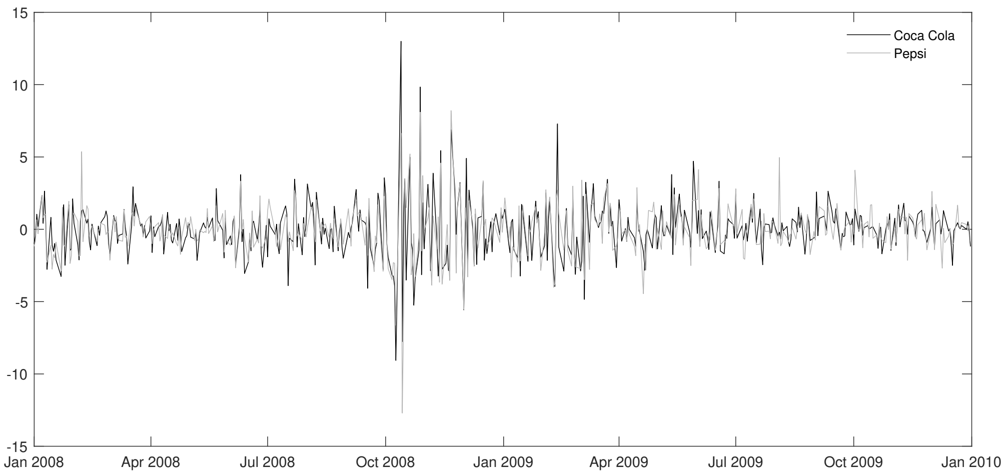


Table 5.1: Estimated parameters - λ-GARCH.

$\hat{W}$	diag($\hat{A}$)	diag($\hat{B}$)	$\hat{\phi}$	$\hat{V}$	
0.072 (0.044)	0.107 (0.051)	0.814 (0.087)	0.847 (0.047)	0.662 (0.035)	0.749 (0.031)
0.186 (0.087)	0.166 (0.041)	0.796 (0.049)		−0.749 (0.031)	0.662 (0.035)
$L_T(\hat{\theta}_T)$	−1,698.43		$\hat{\gamma}$	−0.101 (0.068)	

Standard errors are given in parenthesis below the point estimates.

possibly due to the small sample size) based on the point estimate of the top Lyapunov exponent.[1] The estimated conditional eigenvalues are given in Figure 5.2. We find that the second eigenvalue, $\hat{\lambda}_{2,t}$, explains on average 90% of the variation in the data. From the estimates of $\hat{V}$ we find that the first rotated return, $\hat{y}_{1,t} = \hat{V}_1' X_t$, corresponds to a long-short portfolio, while the second, $\hat{y}_{2,t} = \hat{V}_2' X_t$, is a long-only portfolio. The estimated residuals are given in Figure 5.3, along with histograms of the standardized residuals. We find that while the residuals are characterized by a heavy-tailed density, they do not seem to be autocorrelated in the absolute value, indicating a decent fit of the λ-GARCH model.

We are interested in testing the hypothesis that the eigenvalues are (jointly) integrated, that is,

$$H_0 : \; a_1 + b_1 = 1 \quad \vee \quad a_2 + b_2 = 1,$$

a hypothesis with $k = 2$ restrictions. The Wald test statistic is $W_T = 4.62$, with a bootstrap p-value of $p^* = 0.096$. Hence, we cannot reject that the eigenvalues are indeed

[1]In practice, direct computation of γ in (2.5) is numerically unstable, because the matrix product converges to zero exponentially fast, and we therefore employ the QR method outlined in Dieci and Van Vleck (1995) to compute the top Lyapunov exponent. See also Nielsen and Rahbek (2014).

FIGURE 5.2: Estimated eigenvalues

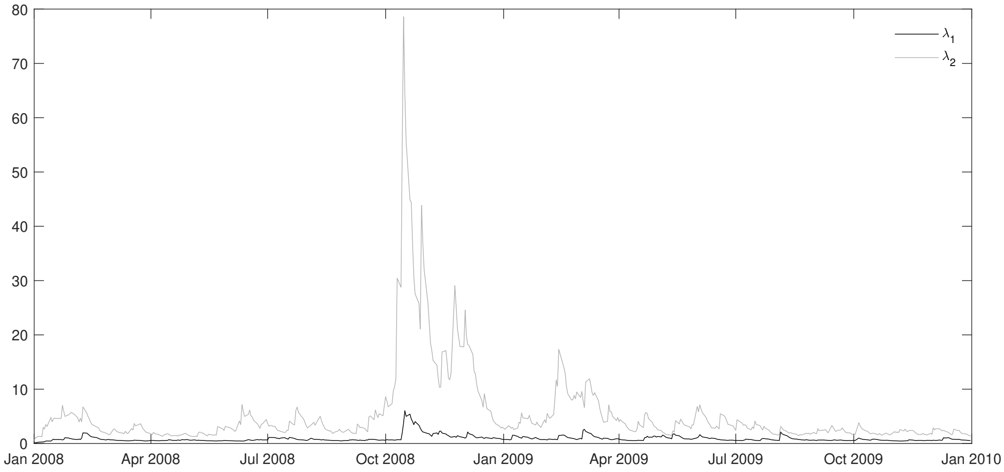


FIGURE 5.3: Estimated residuals

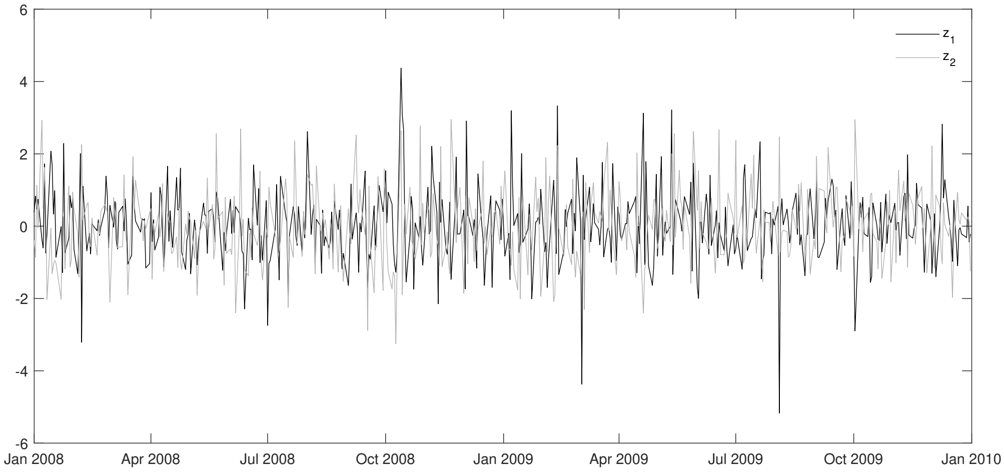
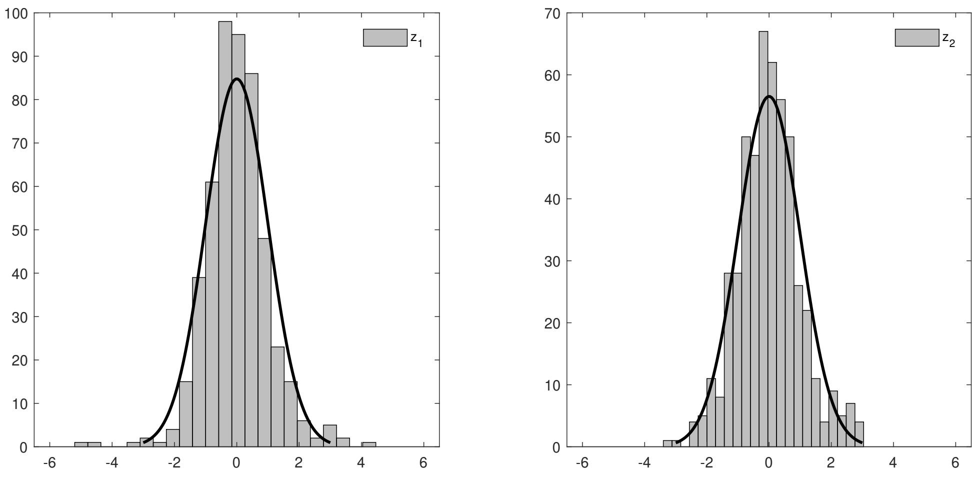


FIGURE 5.4: Histogram of standardized residuals



Note: The thick line in the histograms is a standard Gaussian reference.

integrated. When testing the auxiliary hypotheses that $a_i + b_i = 1$ for $i = 1, 2$ we also fail to reject the null that $\lambda_{i,t}$ for $i = 1, 2$ are integrated (bootstrap p-values of $p^* = 0.16$ and $p^* = 0.11$ respectively).

6 Extensions and Concluding Remarks

We extend the univariate fixed-design (or fixed volatility) bootstrap of Cavaliere, Pedersen, and Rahbek (2018) of ARCH(q)-type models to the multivariate λ-GARCH model of Hetland, Pedersen, and Rahbek (2020). We consider the fixed-design bootstrap for multivariate GARCH processes in the context of full quasi-maximum likelihood estimation, and we explore the theoretical and numerical properties this bootstrap. We show, under mild assumptions, that the bootstrap Wald test statistic is consistent, conditional on the original data. We contrast the theoretically investigated fixed-design bootstrap to that of a recursive bootstrap, and the asymptotic test statistic, and find that the fixed-design bootstrap provides excellent coverage. This is especially true for the smallest sample size of $T = 250$, where the fixed-design bootstrap dominates the recursive bootstrap and asymptotic test statistic. In sample sizes of $T \in (500, 1000, 2000)$ all three test statistics perform roughly equivalent and have a correct nominal size, and comparable power, as expected from the asymptotic theory.

We note that while the fixed-design bootstrap delivers promising results in this exposition, the theory presented in this paper only holds for hypotheses in the interior of the parameter space. That is, an interesting extension would be to extend the work of Cavaliere, Nielsen, and Rahbek (2017) and Cavaliere et al. (2020) to a multivariate setting. This would allow us test hypotheses on the boundary of the parameter space, such as a test for no-ARCH effects, or no eigenvalue spillovers. Another interesting extension would be to examine if the moment requirement can be relaxed in the theoretical exposition. We find that while finite fourth order moments of the return process is a sufficient condition for bootstrap consistency, it does not appear to be necessary based on our Monte Carlo simulations. Both of these extensions are left for future work.

APPENDIX

A Technical appendix

The technical exposition builds on Cavaliere, Pedersen, and Rahbek (2018), who demonstrate bootstrap consistency of the Wald test statistic for the univariate ARCH(q) model. We extend their results to the multivariate λ-GARCH, and most of our lemmata are the multivariate analogous to theirs. Furthermore, the “classical” asymptotic theory presented in Theorem 2.1-2.2 are proved in Hetland, Pedersen, and Rahbek (2020).

The appendix is organized as follows, Section A.1 contains the proof for Theorem 3.1, Sections A.2 and A.3 contains lemmata required to verify consistency (conditional on the original sample) and weak convergence in probability of the bootstrap QML estimator respectively. Section A.4 verifies that the bootstrap Wald test is consistent conditionally on the original data, and finally Section A.5 contain lemmata for the (bootstrap) residuals.

Let the letters K and ϕ denote generic constants, whose value can vary along the text, but always satisfy $K > 0$ and $0 < \phi < 1$. Furthermore, we let $\tilde{L}_T(\theta)$ and $\tilde{l}_t(\theta)$ denote the stationary and ergodic version of the log-likelihood function and -contribution, which are initiated in the infinite past. This is in contrast to $L_T(\theta)$ and $l_t(\theta)$, given in (2.6)-(2.7), which are initiated in a fixed initial value X_0 and Λ_0. Hence,

$$\tilde{L}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \tilde{l}_t(\theta), \tag{A.1}$$

$$\tilde{l}_t(\theta) = \log\det(\tilde{\Lambda}_t(\theta)) + X_t' V(\theta)\tilde{\Lambda}_t^{-1}(\theta)V'(\theta)X_t, \tag{A.2}$$

where $\tilde{\Lambda}_t(\theta)$ (and $\tilde{\lambda}_{i,t}(\theta)$, $i = 1,\ldots,p$) is defined analogously. With $\dot{\tilde{\Lambda}}_{i,t}(\theta) = \frac{\partial\tilde{\Lambda}_t(\theta)}{\partial\theta_i}$, $\ddot{\tilde{\Lambda}}_{i,j,t}(\theta) = \frac{\partial^2\tilde{\Lambda}_t(\theta)}{\partial\theta_i\partial\theta_j}$, and $\dddot{\tilde{\Lambda}}_{i,j,k,t}(\theta) = \frac{\partial^3\tilde{\Lambda}_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}$, we apply the following bounds from Hetland, Pedersen, and Rahbek (2020) (Lemma A.7).

$$E\left[\sup_{\theta\in\text{int}\Theta} ||\tilde{\Lambda}_t(\theta)^{-1}\dot{\tilde{\Lambda}}_{i,t}(\theta)||^r\right] < \infty, \tag{A.3}$$

$$E\left[\sup_{\theta\in\text{int}\Theta} ||\tilde{\Lambda}_t(\theta)^{-1}\ddot{\tilde{\Lambda}}_{i,j,t}(\theta)||^r\right] < \infty, \tag{A.4}$$

$$E\left[\sup_{\theta\in\text{int}\Theta} ||\tilde{\Lambda}_t(\theta)^{-1}\dddot{\tilde{\Lambda}}_{i,j,k,t}(\theta)||^r\right] < \infty \tag{A.5}$$

for $r > 0$, and for $\theta \in \text{int}\Theta$,

$$\begin{aligned}\frac{\tilde{\lambda}_{i,t}(\theta)}{\tilde{\lambda}_{j,t}(\theta)} &= \frac{w_i + \sum_{k=1}^p a_{ik}y_{k,t-1}^2(\theta) + \sum_{k=1}^p b_{ik}\tilde{\lambda}_{k,t-1}(\theta)}{w_j + \sum_{k=1}^p a_{jk}y_{k,t-1}^2(\theta) + \sum_{k=1}^p b_{jk}\tilde{\lambda}_{k,t-1}(\theta)} \\ &\leq \frac{w_i}{w_j} + \sum_{k=1}^{p}\frac{a_{ik}}{a_{jk}} + \sum_{k=1}^{p}\frac{b_{ik}}{b_{jk}} \leq K, \end{aligned}\tag{A.6}$$

for $i, j, k = 1,\ldots,d_\theta$. (A.3)-(A.6) holds under Assumptions 2.1-2.6.

Since $\tilde{\Lambda}_t(\theta)$ and $\Lambda_t(\theta)$ are defined for the same strictly stationary and ergodic sequence, $\{X_t\}_{t\in\mathbb{N}}$, it holds,

$$\sup_{\theta\in\Theta} ||\tilde{\Lambda}_t(\theta) - \Lambda_t(\theta)|| = \sup_{\theta\in\Theta} ||B^t(\tilde{\lambda}_0(\theta) - \lambda_0)|| \leq \phi^t K \xrightarrow{p} 0, \tag{A.7}$$

since $\sup_{\theta\in\Theta}|\rho(B)| < 1$ is necessary for Assumption 2.2 (Francq and Zakoïan, 2019, Corollary 10.1), see also Hetland, Pedersen, and Rahbek (2020) (proof of Theorem 3.2). Similar results also hold for the first and second derivatives of the eigenvalues,

$$\begin{aligned}\sup_{\theta\in\Theta}||\dot{\tilde{\Lambda}}_{i,t}(\theta)-\dot{\Lambda}_{i,t}(\theta)|| &\leq \sup_{\theta\in\Theta}\left|\frac{\partial B^t}{\partial\theta_i}(\tilde{\lambda}_0(\theta)-\lambda_0)\right|+\sup_{\theta\in\Theta}\left|B^t(\dot{\tilde{\lambda}}_{0,i}(\theta)-\dot{\lambda}_{0,i})\right| \\ &\leq \phi^t K \xrightarrow{p} 0, \end{aligned} \tag{A.8}$$

$$\begin{aligned}\sup_{\theta\in\Theta}||\ddot{\tilde{\Lambda}}_{i,j,t}(\theta)-\ddot{\Lambda}_{i,j,t}(\theta)|| &\leq \sup_{\theta\in\Theta}\left|\frac{\partial^2 B^t}{\partial\theta_i\partial\theta_j}\left(\tilde{\lambda}_0(\theta)-\lambda_0\right)\right|+\sup_{\theta\in\Theta}\left|B^t\frac{\partial^2\left(\tilde{\lambda}_0(\theta)-\lambda_0\right)}{\partial\theta_i\partial\theta_j}\right|+ \\ &\quad \sup_{\theta\in\Theta}\left|\frac{\partial B^t}{\partial\theta_i}\frac{\partial\left(\tilde{\lambda}_0(\theta)-\lambda_0\right)}{\partial\theta_j}\right|+\sup_{\theta\in\Theta}\left|\frac{\partial B^t}{\partial\theta_j}\frac{\partial\left(\tilde{\lambda}_0(\theta)-\lambda_0\right)}{\partial\theta_i}\right| \\ &\leq \phi^t K \xrightarrow{p} 0, \end{aligned} \tag{A.9}$$

see e.g., Hetland (2020) (proof of Lemma B.10) and Francq and Zakoïan (2012) (pp.204-206).

A.1 Proof of Theorem 3.1

Proof. The starting point of the proof is a Taylor expansion around $\hat{\theta}_T$, for which we need Theorems 2.1-2.2 along with lemma A.1.

$$0_{d_\theta\times 1}=\frac{\partial L_T^*(\hat{\theta}_T)}{\partial\theta}+\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'}\left(\hat{\theta}_T^*-\hat{\theta}_T\right)+M_T(\hat{\theta}_T^*,\hat{\theta}_T), \tag{A.10}$$

where $M_T(\cdot)$ is a remainder term and d_θ is the size of the parameter vector, $d_\theta = p(p-1)/2+p+2p^2$. By Lemmata A.3, A.4, A.6, Theorems 2.1-2.2, and the bootstrap version of Slutsky's lemma, it holds under the null (and alternative) that,

$$\sqrt{T}(\hat{\theta}_T^*-\hat{\theta}_T)=\sqrt{T}\left(\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'}\right)^{-1}\frac{\partial L_T^*(\hat{\theta}_T)}{\partial\theta}\xrightarrow{w^*}_p N(0,\Sigma). \tag{A.11}$$

with $\Sigma = J^{-1}VJ^{-1}$ where J and V defined in (A.24) in the proof of Lemma A.7. The bootstrap Wald test is,

$$W_T^* = T(\hat{\theta}_T^*-\hat{\theta}_T)'R'(R\hat{\Sigma}_T^*R')^{-1}R(\hat{\theta}_T^*-\hat{\theta}_T)\xrightarrow{w^*}_p \chi^2(k).$$

which is $\chi^2(k)$ distributed under both the null and the alternative by Lemma A.7 and the bootstrap version of Slutsky's lemma. □

A.2 Consistency of bootstrap estimator

Lemma A.1. *Under Assumptions 2.1–2.6, the bootstrap QMLE in* (2.8) *is consistent, conditional on the original data. That is, for any* $\varepsilon > 0$,

$$P^*(||\hat{\theta}_T^* - \hat{\theta}_T|| > \varepsilon) \xrightarrow{p} 0.$$

Proof. We follow Cavaliere, Pedersen, and Rahbek (2018) (Lemma A.1) and Dovonon and Gonçalves (2017) (proof of Proposition 3.1), allowing for a slight modification to account for the impact of fixed initial values. From Theorem 2.1 we know that $\theta_0 = \arg\min_{\theta\in\Theta} E[\tilde{l}_t(\theta)]$ is unique, and recall that $L_T(\theta)$ is defined in (2.6) and $\tilde{L}_T(\theta)$ in (A.1). For any $\varepsilon > 0$, such that $||\theta - \theta_0|| > \varepsilon$ there exists a $\delta = \delta(\varepsilon) > 0$ such that $E[\tilde{l}_t(\theta)] - E[\tilde{l}_t(\theta_0)] > \delta$. Hence, with $L_T(\hat{\theta}_T) \leq L_T(\theta_0)$ and $L_T^*(\hat{\theta}_T^*) \leq L_T^*(\hat{\theta}_T)$, along with the union bound,

$$\begin{aligned} P^*(||\hat{\theta}_T^* - \theta_0|| > \varepsilon) &\leq P^*(E[\tilde{l}_t(\theta)] - E[\tilde{l}_t(\theta_0)] + L_T(\hat{\theta}_T) - L_T(\hat{\theta}_T) + L_T^*(\hat{\theta}_T^*) - L_T^*(\hat{\theta}_T^*) > \delta) \\ &\leq P^*(E[\tilde{l}_t(\theta)] - E[\tilde{l}_t(\theta_0)] + L_T(\theta_0) - L_T(\hat{\theta}_T) + L_T^*(\hat{\theta}_T) - L_T^*(\hat{\theta}_T^*) > \delta) \\ &\leq P^*\left(2\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| > \delta/2\right) + P^*\left(2\sup_{\theta\in\Theta}|L_T(\theta) - L_T^*(\theta)| > \delta/2\right). \end{aligned}$$

Notice here that $P^*\left(2\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| > \delta/2\right)$ is given when we condition on the original sample, $\{X_t\}_{t=1}^T$,

$$P^*\left(2\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| > \delta/2\right) = I\left\{2\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| > \delta/2\right\}.$$

Hence, we need to verify the following two statements:

1. $\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| \xrightarrow{p} 0$.
2. $P^*\left(\sup_{\theta\in\Theta}|L_T(\theta) - L_T^*(\theta)| > \delta/2\right) \xrightarrow{p} 0$.

Starting with 1., by the triangle inequality,

$$\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - L_T(\theta)| \leq \sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - \tilde{L}_T(\theta)| + \sup_{\theta\in\Theta}|\tilde{L}_T(\theta) - L_T(\theta)|,$$

where $\sup_{\theta\in\Theta}|E[\tilde{l}_t(\theta)] - \tilde{L}_T(\theta)| \xrightarrow{p} 0$ holds trivially by an application of the uniform law of large numbers for ergodic processes, using that Θ is compact and that $E[\tilde{l}_t(\theta)]$ is continuous in θ, along with the fact that if $E||X_t||^{2+\delta} < \infty$ for $\delta > 0$, as assumed in Assumption 2.6. Second, $\sup_{\theta\in\Theta}|\tilde{L}_T(\theta) - L_T(\theta)| \xrightarrow{p} 0$ is shown to hold by utilizing (A.7) (see e.g., Hetland, Pedersen, and Rahbek, 2020, proof of Theorem 3.2, and Francq and Zakoïan, 2019, proof of Theorem 10.2). Finally, it is shown in Lemma A.2 that the statement in 2. holds.

Hence,

$$P^*(||\hat{\theta}_T^* - \hat{\theta}_T|| > \varepsilon) \xrightarrow{p} 0.$$

□

Lemma A.2. *Suppose that Assumptions 2.1–2.6 holds and in addition* $E||X_t||^4 < \infty$*, then for* $\varepsilon > 0$*,*

$$P^*\left(\sup_{\theta\in\Theta}|L_T^*(\theta) - L_T(\theta)| > \varepsilon\right) \xrightarrow{p} 0$$

Proof. Notice that,

$$\begin{aligned}\sup_{\theta\in\Theta}|L_T^*(\theta) - L_T(\theta)| &= \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\operatorname{tr}\left\{\Lambda_t^{-1}(\theta)(Y_t^*(\theta)Y_t^{*\prime}(\theta) - Y_t(\theta)Y_t'(\theta))\right\}\right| \\ &\leq \sup_{\theta\in\Theta}\left|K\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*2}(\theta) - y_{i,t}^2(\theta)\right|,\end{aligned}$$

where we use that, $E[\sup_{\theta\in\Theta}||\Lambda_t^{-1}(\theta)||] \leq K$ by compactness of Θ and strict positivity of W (see also Francq and Zakoïan (2019) proof of Theorem 10.8), and $\operatorname{tr}(Y_t^*(\theta)Y_t^{*\prime}(\theta) - Y_t(\theta)Y_t'(\theta)) = \sum_{i=1}^{p} y_{i,t}^{*2}(\theta) - y_{i,t}^2(\theta)$. By the bootstrap version of Markov's inequality,

$$\begin{aligned}&P^*\left(\sup_{\theta\in\Theta}\left|K\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*2}(\theta) - y_{i,t}^2(\theta)\right| > \varepsilon\right) \leq K\varepsilon^{-2}E^*\left[\sup_{\theta\in\Theta}\left|\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*2}(\theta) - y_{i,t}^2(\theta)\right)^2\right|\right] \\ &= K\varepsilon^{-2}\frac{1}{T}E^*\left[\sup_{\theta\in\Theta}\underbrace{\left|\frac{1}{T}(\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*4}(\theta) + y_{i,t}^4(\theta)) + \frac{1}{T}C_T(\theta)\right|}_{\triangle}\right], \qquad \text{(A.12)}\end{aligned}$$

where $C_T(\theta)$ has $T(T-1)$ terms, which contain all cross-products, all of the form,

$$\begin{gathered}y_{i,t}^2(\theta)y_{j,t}^2(\theta), \quad y_{i,t}^2(\theta)y_{i,t}^{*2}(\theta), \quad y_{i,t}^2(\theta)y_{j,t}^{*2}(\theta), \quad y_{i,t}^{*2}(\theta)y_{j,t}^{2*}(\theta), \\ y_{i,t-k}^2(\theta)y_{i,t-q}^2(\theta), \quad y_{i,t-k}^2(\theta)y_{j,t-q}^2(\theta), \quad y_{i,t-k}^{*2}(\theta)y_{i,t-q}^{*2}(\theta), \\ y_{i,t-k}^2(\theta)y_{i,t-q}^2(\theta), \quad y_{i,t-k}^{*2}(\theta)y_{i,t-q}^2(\theta), \quad y_{i,t-k}^{*2}(\theta)y_{j,t-q}^2(\theta),\end{gathered}$$

for $i \neq j$ and $k \neq q$. All the terms in $\triangle$ are defined for the same $\lambda_{t-l}(\theta)$ (a function of X_{t-1-l}) for $l = 0, \ldots, t-1$, and constitute sample averages and autocovariance-type functions. Under the assumption $E||X_t||^4 < \infty$, Lemma A.8 and the Cauchy-Schwarz inequality, all terms of $\triangle$ have a finite probability limit, uniformly in $\theta \in \Theta$.

Hence,

$$P^*\left(\sup_{\theta\in\Theta}\left|K\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p}y_{i,t}^{*2}(\theta)-y_{i,t}^{2}(\theta)\right|>\varepsilon\right)\leq$$

$$K\varepsilon^{-2}\frac{1}{T}E^*\left[\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p}(y_{i,t}^{*4}(\theta)+y_{i,t}^{4}(\theta))+\frac{1}{T}C_T(\theta)\right|\right]=o_p(1), \qquad \text{(A.13)}$$

see also Cavaliere, Pedersen, and Rahbek (2018) (proof of Lemma A.2) for a more rigorous treatment of $C_T(\theta)$.

□

A.3 Asymptotic normality of bootstrap estimator

Lemma A.3. *Suppose Assumptions 2.1–2.6 holds, and in addition* $E||X_t||^4<\infty$, *then*

$$\sqrt{T}\frac{\partial L_T^*(\hat{\theta}_T)}{\partial\theta}\overset{w^*}{\to}_p N(0,V), \qquad \text{(A.14)}$$

where $V=E\left[\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta'}\right]$ *is non-negative definite.*

Proof. The i'th element of the score vector of the bootstrap log-likelihood contribution is,

$$\begin{aligned}\frac{\partial l_t^*(\theta)}{\partial\theta_i}&=\mathrm{tr}\{\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,i}(\theta)(I_p-\Lambda_t^{-1}(\theta)Y_t^*(\theta)Y_t^{*\prime}(\theta))\}+2X_t^{*\prime}S_i(\theta)V(\theta)\Lambda_t^{-1}(\theta)Y_t^*(\theta)\\&=M_{t,i}(\theta)+N_{t,i}(\theta),\end{aligned} \qquad \text{(A.15)}$$

where $\frac{\partial Y_t^*(\theta)}{\partial\theta_i}=V'(\theta)S_i(\theta)X_t^*$ (Lemma A.2 of Hetland, Pedersen, and Rahbek, 2020) with $S_i(\theta)$ skew-symmetric ($S_i(\theta)=-S_i(\theta)$ and $\mathrm{tr}(S_i(\theta))=0$).

We show that the bootstrap score function converges weakly in probability by verifying the following regularity conditions for the Lindeberg central limit theorem for triangular arrays,

1. With $\mathcal{F}_t^*=\sigma(X_s^*,s=0,\ldots,t)$, $E^*\left[\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\middle|\mathcal{F}_{t-1}^*\right]=0$

2. $T^{-1}\sum_{t=1}^{T}E^*\left[\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta'}\middle|\mathcal{F}_{t-1}^*\right]\overset{p^*}{\to}_p V$

3. For $\upsilon\in\mathbb{R}^{d_\theta}$ and any $\varepsilon>0$, $T^{-1}\sum_{t=1}^{T}E^*\left[\left(\upsilon'\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\right)^2\mathbb{I}_{\left(\left|\upsilon'\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\right|>T^{1/2}\varepsilon\right)}\right]\overset{p}{\to}0$, where $\mathbb{I}(\cdot)$ is the indicator function.

Regarding 1., with $E^*[Z_t^*] = 0$ and $E^*[Z_t^* Z_t^{*\prime}] = I_p$ (by Lemma A.8), it follows immediately that

$$E^*\left[\frac{\partial l_t^*(\hat{\theta}_T)}{\partial \theta_i}\middle|\mathcal{F}_{t-1}^*\right] = E^*\Big[\text{tr}\{\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,i}(\hat{\theta}_T)(I_p - \Lambda_t^{-1}(\hat{\theta}_T)\Lambda_t^{1/2}(\hat{\theta}_T)Z_t^* Z_t^{*\prime}\Lambda_t^{1/2}(\hat{\theta}_T))\} + 2\text{tr}\{S_i(\hat{\theta}_T)V(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)V'(\hat{\theta}_T)X_t^* X_t^{*\prime}\}\Big|\mathcal{F}_{t-1}^*\Big] = 0.$$

Next, turning to 2., we note that it suffices to show that $T^{-1}\sum_{t=1}^T E^*\left[\left(\frac{\partial l_t^*(\hat{\theta}_T)}{\partial \theta_i}\right)^2\middle|\mathcal{F}_{t-1}^*\right] \xrightarrow{p} K < \infty$ for $i = 1, \ldots, d_\theta$. In light of (A.15), along with the Cauchy-Schwarz inequality, the variance of the score vector is finite if $\frac{1}{T}\sum_{t=1}^T E^*[M_{t,i}(\hat{\theta}_T)^2] \xrightarrow{p} K < \infty$ and $\frac{1}{T}\sum_{t=1}^T E^*[N_{t,i}(\hat{\theta}_T)^2] \xrightarrow{p} K < \infty$.

Starting with $\frac{1}{T}\sum_{t=1}^T E[M_{t,i}^2(\hat{\theta}_T)]$,

$$\frac{1}{T}\sum_{t=1}^T E^*\left[M_{t,i}^2(\hat{\theta}_T)\middle|\mathcal{F}_{t-1}^*\right] =$$
$$\frac{1}{T}\sum_{t=1}^T E^*\left[\text{tr}^2\{\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,i}(\hat{\theta}_T)(I_p - Z_t^* Z_t^{*\prime})\}\middle|\mathcal{F}_{t-1}^*\right] =$$
$$\frac{1}{T}\sum_{t=1}^T\sum_{k=1}^p \frac{\dot{\lambda}_{k,t,i}^2(\hat{\theta}_T)}{\lambda_t^2(\hat{\theta}_T)}E^*[z_{k,t}^{*4} - 1] \xrightarrow{p^*}_p \sum_{k=1}^p E\left[\frac{\dot{\tilde{\lambda}}_{k,t,i}^2(\theta_0)}{\tilde{\lambda}_t^2(\theta_0)}\right]E[z_{k,t}^4 - 1] < \infty,$$

by (A.3), (A.7)-(A.8) and Lemma A.8, along with the uniform law of large numbers for stationary and ergodic sequences, along with the fact that $\hat{\theta}_T \xrightarrow{p} \theta_0$ and $E\left[\frac{\dot{\tilde{\lambda}}_{k,t,i}^2(\theta_0)}{\tilde{\lambda}_t^2(\theta_0)}\right]$ is continuous at θ_0.

Turning to $\frac{1}{T}\sum_{t=1}^T E^*[N_{t,i}^2(\hat{\theta}_T)]$,

$$\frac{1}{T}\sum_{t=1}^T E^*\left[N_{t,i}^2(\hat{\theta}_T)\middle|\mathcal{F}_{t-1}^*\right] = \frac{4}{T}\sum_{t=1}^T$$
$$E^*\Big[\text{tr}\{V'(\hat{\theta}_T)S_i(\hat{\theta}_T)V(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T)V'(\hat{\theta}_T)S_i(\hat{\theta}_T)V(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T)\}\Big|\mathcal{F}_{t-1}^*\Big] \leq$$
$$\frac{1}{T}\sum_{t=1}^T KE^*\left[||\Lambda_t^{-1}(\hat{\theta}_T)Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T)||^2\middle|\mathcal{F}_{t-1}^*\right] = \frac{K}{T}\sum_{t=1}^T E^*[Y_t^{*\prime}(\hat{\theta}_T)Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T)\Lambda_t^{-2}Y_t^{*\prime}(\hat{\theta}_T)|\mathcal{F}_{t-1}^*] =$$
$$\frac{1}{T}\sum_{t=1}^T KE^*\left[\left(\sum_{i=1}^p y_{i,t}^{*2}(\hat{\theta}_T)\right)\left(\sum_{i=1}^p y_{i,t}^{*2}(\hat{\theta}_T)/\lambda_{i,t}^2(\hat{\theta}_T)\right)\middle|\mathcal{F}_{t-1}^*\right], \tag{A.16}$$

where we utilize that both $\sup_{\theta\in\Theta}||V(\theta)|| < K$ and $\sup_{\theta\in\Theta}||S_i(\theta)|| < K$ for $i = 1, \ldots, d_\theta$, as both matrices consist solely of trigonometrical functions. (A.16) has terms of the form,

$$\frac{1}{T}\sum_{t=1}^{T} E^*\left[\frac{y_{i,t}^{*2}(\hat{\theta}_T)y_{j,t}^{*2}(\hat{\theta}_T)}{\lambda_{j,t}^2(\hat{\theta}_T)}\middle|\mathcal{F}_{t-1}^*\right] = \frac{1}{T}\sum_{t=1}^{T} E^*[z_{i,t}^{*2}z_{j,t}^{*2}]\frac{\lambda_{i,t}(\hat{\theta}_T)}{\lambda_{j,t}(\hat{\theta}_T)} \xrightarrow{p^*}_p E[z_{i,t}^2 z_{j,t}^2]E\left[\frac{\tilde{\lambda}_{i,t}(\theta_0)}{\tilde{\lambda}_{j,t}(\theta_0)}\right] < \infty,$$

by (A.6)-(A.7), Lemma A.8 and the uniform law of large numbers for stationary and ergodic sequences. The cross product, $\frac{1}{T}\sum_{t=1}^{T} E^*\left[M_{t,i}(\hat{\theta}_T)N_{t,i}(\hat{\theta}_T)\middle|\mathcal{F}_{t-1}^*\right]$ converges in probability, conditional on the original sample, by the Cauchy-Schwarz inequality. The matrix V is non-negative definite by construction.

Turning to the Lindeberg condition in 3., we follow Francq, Horvath, and Zakoïan (2014) (proof of Theorem 4.1). That is, since the bootstrap sample, conditional on the original data, is independent with finite fourth order moments, we have that,

$$E^*\left[\left(\upsilon'\frac{l_t^*(\hat{\theta}_T)}{\partial\theta}\right)^2\right] \xrightarrow{p} K < \infty \quad \text{for} \quad i = 1, \ldots, d_\theta.$$

Second, for $T \to \infty$, the indicator function tends to zero,

$$\mathbb{I}\left(\left|\upsilon'\frac{l_t^*(\hat{\theta}_T)}{\partial\theta}\right| \geq \varepsilon\sqrt{T}\right) \xrightarrow{p} 0.$$

Hence, each term of the Lindeberg condition tends to zero,

$$E^*\left[\left(\upsilon'\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\right)^2 \mathbb{I}_{\left(\left|\upsilon'\frac{\partial l_t^*(\hat{\theta}_T)}{\partial\theta}\right| > \sqrt{T}\varepsilon\right)}\right] \xrightarrow{p} 0. \tag{A.17}$$

By Cesáros lemma, the average also tends to zero almost surely, and the Lindeberg condition holds.

□

Lemma A.4. *Under Assumptions 2.1–2.6 along with $E||X_t||^4 < \infty$, for $\varepsilon > 0$, there exists a $\delta > 0$ such that,*

$$P\left(P^*\left(\left|\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right]\right|\right| > \delta\right) > \varepsilon\right) \to 0.$$

Proof. By the triangle inequality,

$$\left|\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right]\right|\right| \leq$$
$$\left|\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'} - \frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta\partial\theta'}\right|\right| + \left|\left|\frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta\partial\theta'} - \frac{\partial^2 \tilde{L}_T(\hat{\theta}_T)}{\partial\theta\partial\theta'}\right|\right| + \left|\left|\frac{\partial^2 \tilde{L}_T(\hat{\theta}_T)}{\partial\theta\partial\theta'} - E\left[\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right]\right|\right|.$$

The first term tends to zero in probability by Lemma A.5 below. The second term concerns the fixed initial values and tends to zero by (A.7)-(A.9), see also Francq and Zakoïan (2012) (p. 204-206) and Hetland (2020) (proof of Lemma B.10). The third term tends to zero in probability by standard arguments (compact Θ, the uniform law of large numbers, continuity at θ_0 and consistency of $\hat{\theta}_T$). □

Lemma A.5. *Under Assumptions 2.1–2.6 along with $E||X_t||^4 < \infty$, for any $\varepsilon > 0$ there exists a $\delta > 0$ such that*

$$P\left(P^*\left(\left\|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta\partial\theta'} - \frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta\partial\theta'}\right\| > \delta\right) > \varepsilon\right) \to 0.$$

Proof. It suffices to show that the result holds element-wise, i.e.,

$$P\left(P^*\left(\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\right| > \delta\right) > \varepsilon\right) \to 0, \tag{A.18}$$

for $i, j = 1, \ldots, d_\theta$. First, from Lemma A.4 of Hetland, Pedersen, and Rahbek (2020), the (i,t)'th element of the Hessian matrix is (suppressing the dependency on θ),

$$\begin{aligned}\frac{\partial^2 L_T(\theta)}{\partial\theta_i\partial\theta_j} = \frac{1}{T}\sum_{t=1}^{T}\Bigg(& -\operatorname{tr}\left(\Lambda_t^{-1}\dot{\Lambda}_{t,j}\Lambda_t^{-1}\dot{\Lambda}_{t,i}\right) + \operatorname{tr}\left(\Lambda_t^{-1}\ddot{\Lambda}_{t,i,j}\right) + \operatorname{tr}\left(\Lambda_t^{-1}\dot{\Lambda}_{t,j}\Lambda_t^{-1}\dot{\Lambda}_{t,i}\Lambda_t^{-1}Y_tY_t'\right)\\
& - \operatorname{tr}\left(\Lambda_t^{-1}\ddot{\Lambda}_{t,i,j}\Lambda_t^{-1}Y_tY_t'\right) + \operatorname{tr}\left(\Lambda_t^{-1}\dot{\Lambda}_{t,i}\Lambda_t^{-1}\dot{\Lambda}_{t,j}\Lambda_t^{-1}Y_tY_t'\right) - 2\operatorname{tr}\left(\tilde{S}_j'\Lambda_t^{-1}\dot{\Lambda}_{t,i}\Lambda_t^{-1}Y_tY_t'\right)\\
& + 2\operatorname{tr}\left((\dot{S}_{i,j} + S_iS_j)\Omega_t^{-1}X_tX_t'\right) + 2\operatorname{tr}\left(V'\left(\dot{S}_{i,j} + S_iS_j\right)V\Lambda_t^{-1}Y_tY_t'\right) - 2\operatorname{tr}\left(\tilde{S}_i\Lambda_t^{-1}\dot{\Lambda}_{t,j}\Lambda_t^{-1}Y_tY_t'\right)\\
& + 2\operatorname{tr}\left(\tilde{S}_i'\Lambda_t^{-1}\tilde{S}_jY_tY_t'\right)\Bigg),\end{aligned} \tag{A.19}$$

where $\tilde{S}_i$ is a skew-symmetric matrix, $\tilde{S}_i' = V'S_iV = -V'S_i'V = -\tilde{S}_i$ and $\dot{S}_{i,j} = \frac{\partial S_i}{\partial\theta_j}$.

Hence,

$$\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\right| =$$

$$\begin{aligned}
&\Bigg|\frac{1}{T}\sum_{t=1}^{T}\Bigg(\text{tr}\left(\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,j}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,i}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&- \text{tr}\left(\Lambda_t^{-1}(\hat{\theta}_T)\ddot{\Lambda}_{t,i,j}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&+ \text{tr}\left(\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,i}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,j}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&- 2\text{tr}\left(\tilde{S}_j'(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,i}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&+ 2\text{tr}\left((\dot{S}_{i,j}(\hat{\theta}_T) + S_i(\hat{\theta}_T)S_j(\hat{\theta}_T))\Omega_t^{-1}(\hat{\theta}_T)(X_t^*X_t^{*\prime} - X_tX_t')\right)\\
&+ 2\text{tr}\left(V'(\hat{\theta}_T)\left(\dot{S}_{i,j}(\hat{\theta}_T) + S_i(\hat{\theta}_T)S_j(\hat{\theta}_T)\right)V(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&- 2\text{tr}\left(\tilde{S}_i(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)\dot{\Lambda}_{t,j}(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\\
&+ 2\text{tr}\left(\tilde{S}_i'(\hat{\theta}_T)\Lambda_t^{-1}(\hat{\theta}_T)\tilde{S}_j(\hat{\theta}_T)(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))\right)\Bigg)\Bigg|. \qquad \text{(A.20)}
\end{aligned}$$

Note that all terms of (A.20) essentially are of the form $K\text{tr}(Y_t^*(\hat{\theta}_T)Y_t^{*\prime}(\hat{\theta}_T) - Y_t(\hat{\theta}_T)Y_t'(\hat{\theta}_T))$, and recall that $E[\sup_{\theta\in\Theta}||\Lambda_t^{-1}(\theta)||] \leq K$ by compactness of Θ and strict positivity of W (see also Francq and Zakoïan (2019) proof of Theorem 10.8). Furthermore, $\sup_{\theta\in\Theta}||V(\theta)|| < K$ and $\sup_{\theta\in\Theta}||S_i(\theta)|| < K$ for $i = 1,\ldots,d_\theta$, as V and S_i consist solely of trigonometrical functions. This, in addition to (A.3)-(A.5), allows us to evaluate up,

$$\left|\frac{\partial^2 L_T^*(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j}\right| \leq \left|K\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*2}(\hat{\theta}_T) - y_{i,t}^2(\hat{\theta}_T)\right|. \qquad \text{(A.21)}$$

Notice that (A.21), is the same as (A.13) from the proof of Lemma A.2, which is shown to be $o_p(1)$, conditional on the original sample, under the assumption $E||X_t||^4 < \infty$ and using the Cauchy-Schwarz inequality. This allows us to conclude that

$$P^*\left(\left|K\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{p} y_{i,t}^{*2}(\hat{\theta}_T) - y_{i,t}^2(\hat{\theta}_T)\right| > \delta\right) \xrightarrow{p} 0, \qquad \text{(A.22)}$$

and hence (A.18) holds for $i,j = 1,\ldots,d_\theta$.

□

Lemma A.6. *Suppose Assumptions 2.1–2.6 hold. Then,*

$$\sup_{\theta\in\Theta}\left|\frac{\partial^3 L_T^*(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leq C_T^*, \text{ where } C_T^* = O_p^*(1) \text{ in probability,}$$

for $i,j,k = 1,\ldots,d_\theta$.

Proof. The entire expression of the array of third order derivative is given in Hetland, Pedersen, and Rahbek (2020) Lemma A.6, and all terms can be bounded stochastically.

Specifically, all terms can be shown to be bounded by the same method as applied to the terms #3 and #5 (pp. 38-39), which contain,

$$\frac{1}{T}\sum_{t=1}^{T}\text{tr}\{\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,k}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,j}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,i}(\theta)V'(\theta)V(\hat{\theta}_T)\Lambda_t^{1/2}(\hat{\theta}_T)Z_t^*Z_t^{*\prime}\Lambda_t^{1/2}(\hat{\theta}_T)V'(\hat{\theta}_T)V(\theta)\} =$$
$$\frac{1}{T}\sum_{t=1}^{T}\text{vec}(V'(\theta)V(\hat{\theta}_T))'(\Lambda_t^{1/2}(\hat{\theta}_T)Z_t^*Z_t^{*\prime}\Lambda_t^{1/2}(\hat{\theta}_T)\otimes\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,k}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,j}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,i}(\theta))\text{vec}(V'(\hat{\theta}_T)V(\theta)),$$

where $\sup_{\theta\in\Theta}|\text{vec}(V'(\hat{\theta}_T)V(\theta))| < K$, as it consists of rotations of trigonometric functions. Next, $\Lambda_t^{1/2}(\hat{\theta}_T)Z_t^*Z_t^{*\prime}\Lambda_t^{1/2}(\hat{\theta}_T)\otimes\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,k}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,j}(\theta)\Lambda_t^{-1}(\theta)\dot{\Lambda}_{t,i}(\theta)$ consists of $p\times p$ blocks, each of which are diagonal, with elements,

$$\lambda_{g,t}^{1/2}(\hat{\theta}_T)z_{g,t}^*z_{h,t}^*\lambda_{h,t}^{1/2}(\hat{\theta}_T)\frac{\dot{\lambda}_{s,t,i}(\theta)\dot{\lambda}_{s,t,j}(\theta)\dot{\lambda}_{s,t,k}(\theta)}{\lambda_{s,t}^4(\theta)}$$

for $g,h,s=1,\ldots,p$. Here $\sup_{\theta\in\Theta}|\dot{\lambda}_{s,t,i}(\theta)\dot{\lambda}_{s,t,j}(\theta)\dot{\lambda}_{s,t,k}(\theta)/\lambda_{s,t}^3(\theta)| \leq K$ by (A.3) for $T\to\infty$, but such a property does not hold for $\lambda_{g,t}^{1/2}(\hat{\theta}_T)z_{g,t}^*z_{h,t}^*\lambda_{h,t}^{1/2}(\hat{\theta}_T)/\lambda_{s,t}(\theta)$ for $g\neq h\neq s$ without assuming $E||X_t||^{2+\delta}$, $\delta>0$. This is because the numerator and denominator are evaluated in θ and $\hat{\theta}_T$ respectively. Hence, by Assumption 2.1–2.6,

$$\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\lambda_{g,t}^{1/2}(\hat{\theta}_T)z_{g,t}^*z_{h,t}^*\lambda_{h,t}^{1/2}(\hat{\theta}_T)/\lambda_{s,t}(\theta)| \xrightarrow{p} K<\infty. \tag{A.23}$$

for $g,h,s=1,\ldots,p$. This can be applied to all terms of the array of third order derivatives, such that $\sup_{\theta\in\Theta}\left|\frac{\partial^3 L_T^*(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| \leq C_T^*$ for $i,j,k=1,\ldots,d_\theta$. □

A.4 Validity of bootstrap Wald test

Lemma A.7. *Suppose that Assumptions 2.1–2.6 along with $E||X_t||^4<\infty$ holds, then*

$$\hat{\Sigma}_T^* \xrightarrow{p^*}_p \Sigma.$$

Proof. Recall that $\Sigma = J^{-1}VJ^{-1}$, with

$$J = E\left[\frac{\partial^2\tilde{l}_t(\theta_0)}{\partial\theta\partial\theta}\right] \text{ and } V = E\left[\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta'}\right], \tag{A.24}$$

with $\tilde{l}_t(\theta)$ defined in (A.2). Recall that, $\hat{J}_T \xrightarrow{p} J$ and $\hat{V}_T \xrightarrow{p} V$ by Theorem 2.2. We now show that $\hat{J}_T^* \xrightarrow{p^*} J$ and $\hat{V}_T^* \xrightarrow{p^*} V$. Notice that it is sufficient to show that for any $i,j=1,\ldots,d_\theta$ $|J_{ij,T}^*-J_{ij}| \xrightarrow{p^*}_p 0$ and $|V_{ij,T}^*-V_{ij}| \xrightarrow{p^*}_p 0$.

Hence, for any $i, j = 1, \ldots, d_\theta$,

$$\left|\hat{J}^*_{ij,T} - J_{ij}\right| = \left|\frac{\partial^2 L^*_t(\hat{\theta}^*_T)}{\partial\theta_i\partial\theta_j} - J_{ij}\right| \leq \left|\frac{\partial^2 L_T(\hat{\theta}_T)}{\partial\theta_i\partial\theta_j} - J_{ij}\right| + \sup_{\theta\in\Theta}\left|\left|\frac{\partial^3 L^*_T(\theta)}{\partial\theta_i\partial\theta_j\partial\theta}\right|\right| \, ||\hat{\theta}^*_T - \hat{\theta}_T|| = o^*_p(1), \tag{A.25}$$

by Lemmata A.1, A.4 and A.6.

Next, with $V^*_T = T^{-1}\sum_{t=1}^T(\partial l^*_t(\hat{\theta}_T)/\partial\theta)(\partial l^*_t(\hat{\theta}_T)/\partial\theta')$ consider

$$||\hat{V}^*_T - V|| \leq ||\hat{V}^*_T - V^*_T|| + ||V^*_T - \hat{V}_T|| + ||\hat{V}_T - V||. \tag{A.26}$$

In a fashion similar to the proof of Lemma A.5, each term can be shown to converge to zero in probability conditional on the original sample. That is, $||\hat{V}^*_T - V^*_T|| \xrightarrow{p^*}_p 0$ by $E||X_t||^4 < \infty$ along with Lemma A.1, $||V^*_T - \hat{V}_T|| \xrightarrow{p^*}_p 0$ by Lemma A.8 and $E||X^4_t|| < \infty$. Finally, $||\hat{V}_T - V|| \xrightarrow{p} 0$ by Theorem 2.1 (consistency). Hence, $\hat{\Sigma}^*_T \xrightarrow{p^*}_p \Sigma$. □

A.5 Lemmata for the bootstrap residuals

Lemma A.8. *Under Assumptions 2.1–2.6 along with $E||X_t||^4 < \infty$,*

$$E^*\left[Z^*_t\right] \xrightarrow{p} E\left[Z_t\right], \qquad E^*\left[Z^*_t Z^{*\prime}_t\right] \xrightarrow{p} E\left[Z_t Z'_t\right], \qquad E^*\left[z^{*4}_{i,t}\right] \xrightarrow{p} E\left[z^4_{i,t}\right],$$

for $i = 1, \ldots, p$.

Proof. First, with $\Omega_Z = \frac{1}{T}\sum_{t=1}^T\left(\hat{Z}_t - \bar{Z}\right)\left(\hat{Z}_t - \bar{Z}\right)'$, and $\bar{Z} = \frac{1}{T}\sum_{t=1}^T \hat{Z}_t$,

$$E^*[Z^*_t] = \frac{1}{T}\sum_{t=1}^T \hat{Z}^s_t = \Omega_Z^{-1/2}\left(\frac{1}{T}\sum_{t=1}^T\left(\hat{Z}_t - \bar{Z}\right)\right) = 0 = E[Z_t],$$
$$E^*[Z^*_t(Z^*_t)'] = \frac{1}{T}\sum_{t=1}^T \hat{Z}^s_t\hat{Z}^{s\prime}_t = \Omega_Z^{-1/2}\frac{1}{T}\sum_{t=1}^T\left(\hat{Z}_t - \bar{Z}\right)\left(\hat{Z}_t - \bar{Z}\right)'\Omega_Z^{-1/2} = I_p = E[Z_t Z'_t].$$

Next, notice that the i'th standardized residual is,

$$\hat{z}^s_{i,t} = \sum_{j=1}^p[\Omega_z^{-1/2}]_{ij}(\hat{z}_{j,t} - \bar{z}_j),$$

where $[\Omega_z^{-1/2}]_{ij}$ is the (i,j)'th element of $\Omega_z^{-1/2}$. The expectation, $E^*[z^{*4}_{i,t}]$, conditional on the original data, is therefore

$$E^*[z^{*4}_{i,t}] = \frac{1}{T}\sum_{t=1}^T(\hat{z}^s_{i,t})^4 = \frac{1}{T}\sum_{t=1}^T\left(\sum_{j=1}^p[\Omega_z^{-1/2}]_{ij}(\hat{z}_{j,t} - \bar{z}_j)\right)^4.$$

Since $\Omega_z \xrightarrow{p} I_p$, it must also hold that $\Omega_z^{-1/2} \xrightarrow{p} I_p$. Hence, for $T \to \infty$ all terms of $E^*[z_{i,t}^{*4}]$, except $[\Omega_z^{-1/2}]_{ii}^4(\hat{z}_{i,t} - \bar{z}_i)^4$, converge to zero in probability,

$$\begin{aligned}
E^*[z_{i,t}^{*4}] &= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{p}[\Omega_z^{-1/2}]_{ii}^4(\hat{z}_{i,t} - \bar{z}_i)\right)^4 + o_p(1) \\
&= [\Omega_z^{-1/2}]_{ii}^4\left(\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^4 + \bar{z}_i^4 + 6\frac{1}{T}\sum_{t=1}^{T} z_i^2\bar{z}_i^2 + 4\frac{1}{T}\sum_{t=1}^{T} z_{i,t}\bar{z}_i^3 + 4\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^3\bar{z}_i\right) + o_p(1) \\
&= [\Omega_z^{-1/2}]_{ii}^4\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^4 + o_p(1) \xrightarrow{p} E[z_{i,t}^4].
\end{aligned}$$

by repeated use of Lemma A.9, along with $E[Z] = 0$ and $E[Z_tZ_t'] = I_p$. □

Lemma A.9. *Under Assumptions 2.1–2.6 along with $E||X_t||^4 < \infty$, for $k = 1, 2, 3, 4$, it holds that*

$$\frac{1}{T}\sum_{t=1}^{T}\hat{z}_{i,t}^k \xrightarrow{p} E[z_{i,t}^k], \quad i = 1, \dots, p.$$

Proof. With $\hat{z}_{i,t} = z_{i,t}(\hat{\theta}_T) = V_i'(\hat{\theta}_T)X_t/\lambda_{i,t}(\hat{\theta}_T)$, where V_i is the i'th column of V, we apply the mean-value theorem around θ_0

$$\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^k(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}^k + \sum_{j=1}^{d_\theta}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial z_{i,t}^k(\tilde{\theta})}{\partial\theta_j}(\hat{\theta}_j - \theta_{0,j}), \tag{A.27}$$

where $z_{i,t} = V_i'(\theta_0)X_t/\lambda_{i,t}(\theta_0)$ and $\tilde{\theta}$ is on the line between $\hat{\theta}_T$ and θ_0 and

$$\frac{\partial z_{i,t}^k(\tilde{\theta})}{\partial\theta_j} = k\left(\frac{1}{\lambda_{i,t}(\tilde{\theta})}\frac{\partial V_i'(\tilde{\theta})}{\partial\theta_j}X_t - \frac{V_i'(\tilde{\theta})X_t}{\lambda_{i,t}(\tilde{\theta})}\frac{\partial\lambda_{i,t}(\tilde{\theta})/\partial\theta_j}{\lambda_{i,t}(\tilde{\theta})}\right)^{k-1}. \tag{A.28}$$

Notice here that by (A.8), along with $E||X_t||^k < \infty$ (for $k = 1, 2, 3, 4$), implies that a uniform law of large numbers hold,

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial z_{i,t}^k(\theta)}{\partial\theta_j}\right| \xrightarrow{p} E\left[\sup_{\theta\in\Theta}\left|\frac{\partial z_{i,t}^k(\theta)}{\partial\theta_j}\right|\right] < \infty.$$

This, along with consistency of the estimator, $||\hat{\theta}_T - \theta_0|| = o_p(1)$, implies that the second term of (A.27) is $o_p(1)$. Furthermore, $\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^k \xrightarrow{p} E[z_{i,t}^k]$ by the law of large numbers for *iid* processes, under the assumption $E[z_{i,t}^k] < \infty$ implied by the moment assumption for the vector of returns, $E||X_t||^k < \infty$. Hence,

$$\frac{1}{T}\sum_{t=1}^{T} z_{i,t}^k(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T} z_{i,t}^k + o_p(1) \xrightarrow{p} E[z_{i,t}^k].$$

□

Bibliography

Alexander, C. and A. Chibumba (1997). "Multivariate orthogonal factor GARCH". *University of Sussex Discussion Papers in Mathematics.*

Anderson, T.W. (1963). "Asymptotic Theory for Principal Component Analysis". *The Annals of Mathematical Statistics* 34 (1), pp. 122–148.

Andrews, D.W.K. (2001). "Testing when a parameter is on the boundary of the maintained hypothesis". *Econometrica* 69 (3), pp. 683–734.

Aït-Sahalia, Y. and D. Xiu (2019). "Principal Component Analysis of High-Frequency Data". *Journal of the American Statistical Association* 114 (525), pp. 287–303.

Avarucci, M., E. Beutner, and P. Zaffaroni (2013). "On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models". *Econometric Theory* 29 (3), pp. 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 (5B), pp. 2730–2759.

Bauwens, L., S. Laurent, and J.V.K. Rombouts (2006). "Multivariate GARCH models: a survey". *Journal of Applied Econometrics* 21 (1), pp. 79–109.

Berkes, I., L. Horváth, and P. Kokoszka (2003). "GARCH processes: Structure and Estimation". *Bernoulli* 9 (2), pp. 201–227.

Beutner, E., A. Heinemann, and S. Smeekes (2020). *A Residual Bootstrap for Conditional Value-at-Risk*. Working Paper.

Bollerslev, T. (1990). "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model". *The Review of Economics and Statistics* 72 (3), pp. 498–505.

Boswijk, H.P. and R. van der Weide (2011). "Method of Moments Estimation of GO-GARCH Models". *Journal of Econometrics* 163 (1), pp. 118–126.

Brown, B.M. (1971). “Martingale Central Limit Theorems”. *The Annals of Mathematical Statistics* 42 (1), pp. 59–66.

Cavaliere, G., H.B. Nielsen, R.S. Pedersen, and A. Rahbek (2020). “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 (4), pp. 513–534.

Cavaliere, G., R.S. Pedersen, and A. Rahbek (2018). “The fixed volatility bootstrap for a class of ARCH(q) models”. *Journal of Time Series Analysis* 39 (6), pp. 920–941.

Cavaliere, G., A. Rahbek, and A.M.R. Taylor (2012). “Bootstrap determination of the co-integration rank in vector autoregressive models”. *Econometrica* 80 (4), pp. 1721–1740.

Christoffersen, P. (1998). “Evaluating Interval Forecasts”. *International Economic Review* 39 (4), pp. 841–862.

— (2009). “Value–at–risk models”. *Handbook of Financial Time Series.* Springer, pp. 753–766.

Comte, F. and O. Lieberman (2003). “Asymptotic theory for multivariate GARCH processes”. *Journal of Multivariate Analysis* 84 (1), pp. 61–84.

Conrad, C. and M. Karanasos (2010). “Negative volatility spillovers in the unrestricted ECCC-GARCH model”. *Econometric Theory* 26 (3), pp. 838–862.

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 (4), pp. 552–563.

— (2013). “Generalized autoregressive score models with applications”. *Journal of Applied Econometrics* 28 (5), pp. 777–795.

Dieci, L. and E.S. Van Vleck (1995). “Computation of a few Lyapunov exponents for continuous and discrete dynamical systems”. *Applied Numerical Mathematics* 17 (3), pp. 275 –291.

Dovonon, P. and S. Gonçalves (2017). “Bootstrapping the GMM overidentification test under first-order underidentification”. *Journal of Econometrics* 201 (1), pp. 43 –71.

Dovonon, P. and E. Renault (2013). "Testing for common conditionally heteroskedastic factors". *Econometrica* 81 (6), pp. 2561–2586.

Eaton, M.L. and D.E. Tyler (1991). "On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix". *The Annals of Statistics* 19 (1), pp. 260–271.

Engle, R.F. (2002). "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models". *Journal of Business & Economic Statistics* 20 (3), pp. 339–350.

Engle, R.F. and K.F. Kroner (1995). "Multivariate simultaneous generalized ARCH". *Econometric Theory* 11 (1), pp. 122–150.

Engle, R.F., O. Ledoit, and M. Wolf (2019). "Large Dynamic Covariance Matrices". *Journal of Business & Economic Statistics* 37 (2), pp. 363–375.

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 (4), pp. 679–702.

Francq, C., L. Horvath, and J.M. Zakoïan (2011). "Merits and drawbacks of variance targeting in GARCH models". *Journal of Financial Econometrics* 9 (4), pp. 619–656.

— (2014). "Variance targeting estimation of multivariate GARCH models". *Journal of Financial Econometrics* 14 (2), pp. 353–382.

Francq, C. and J.M. Zakoïan (2004). "Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes". *Bernoulli* 10 (4), pp. 605–637.

— (2009). "Testing the nullity of GARCH coefficients: correction of the standard tests and relative efficiency comparisons". *Journal of the American Statistical Association* 104 (485), pp. 313–324.

— (2012). "QML estimation of a class of multivariate asymmetric GARCH models". *Econometric Theory* 28 (1), pp. 179–206.

— (2016). "Estimating multivariate volatility models equation by equation". *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 78 (3), pp. 613–635.

— (2019). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Francq, C. and J.M. Zakoïan (2020). “Virtual Historical Simulation for estimating the conditional VaR of large portfolios”. *Journal of Econometrics* 217 (2), pp. 356–380.

Hafner, C.M. and A. Preminger (2009a). “On asymptotic theory for multivariate GARCH models”. *Journal of Multivariate Analysis* 100 (9), pp. 2044–2054.

— (2009b). “Asymptotic theory for a factor GARCH model”. *Econometric Theory* 25 (2), pp. 336–363.

Hall, P. and Q. Yao (2003). “Inference in ARCH and GARCH models with heavy–tailed errors”. *Econometrica* 71 (1), pp. 285–317.

Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). “Asymptotic inference on cointegrating rank in partial systems”. *Journal of Business & Economic Statistics* 16 (4), pp. 388–399.

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.

Hetland, S. (2020). *Spectral Targeting Estimation of Dynamic Conditional Eigenvalue GARCH Models*. Working Paper.

Hetland, S., R.S. Pedersen, and A. Rahbek (2020). *Dynamic Conditional Eigenvalue GARCH*. Working Paper.

Hidalgo, J. and P. Zaffaroni (2007). “A goodness-of-fit test for ARCH(∞) models”. *Journal of Econometrics* 141 (2), pp. 973–1013.

Jacod, J. and P. Protter (2012). *Probability Essentials*. Springer Science and Business Media.

Jeantheau, T. (1998). “Strong consistency of estimators for multivariate ARCH models”. *Econometric Theory* 14 (1), pp. 70–86.

Jensen, S.T. and A. Rahbek (2004). “Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case”. *Econometrica* 72 (2), pp. 641–646.

Jeong, M. (2017). “Residual-based GARCH bootstrap and second order asymptotic refinement”. *Econometric Theory* 33 (3), pp. 779–790.

Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer.

Lanne, M. and P. Saikkonen (2007). "A multivariate generalized orthogonal factor GARCH model". *Journal of Business & Economic Statistics* 25 (1), pp. 61–75.

Ledoit, O. and M. Wolf (2004). "A well-conditioned estimator for large-dimensional covariance matrices". *Journal of Multivariate Analysis* 88 (2), pp. 365–411.

— (2012). "Nonlinear shrinkage estimation of large-dimensional covariance matrices". *The Annals of Statistics* 40 (2), pp. 1024–1060.

Ling, S. and M. McAleer (2003). "Asymptotic theory for a vector ARMA-GARCH model". *Econometric Theory* 19 (2), pp. 280–310.

Magnus, J.R. (1985). "On differentiating eigenvalues and eigenvectors". *Econometric Theory* 1 (2), pp. 179–191.

Newey, W.K. and D. McFadden (1994). "Large sample estimation and hypothesis testing". *Handbook of Econometrics* 4, pp. 2111–2245.

Nielsen, H.B. and A. Rahbek (2014). "Unit root vector autoregression with volatility induced stationarity". *Journal of Empirical Finance* 29, pp. 144 –167.

Noureldin, D., N. Shephard, and K. Sheppard (2014). "Multivariate rotated ARCH models". *Journal of Econometrics* 179 (1), pp. 16–30.

Pascual, L., J. Romo, and E. Ruiz (2006). "Bootstrap prediction for returns and volatilities in garch models". *Computational Statistics & Data Analysis* 50 (9), pp. 2293–2312.

Pedersen, R.S. (2016). "Targeting estimation of CCC-GARCH models with infinite fourth moments". *Econometric Theory* 32 (2), pp. 498–531.

— (2017). "Inference and testing on the boundary in extended constant conditional correlation GARCH models". *Journal of Econometrics* 196 (1), pp. 25–36.

Pedersen, R.S. and A. Rahbek (2014). "Multivariate variance targeting in the BEKK–GARCH model". *The Econometrics Journal* 17 (1), pp. 24–55.

— (2019). "Testing GARCH-X type models". *Econometric Theory* 35 (5), pp. 1012–1047.

Pinheiro, J.C. and D.M. Bates (1996). "Unconstrained parametrizations for variance-covariance matrices". *Statistics and Computing* 6 (3), pp. 289–296.

Shimizu, K. (2010). *Bootstrapping stationary ARMA-GARCH models*. Vieweg+Teubner Research.

Silvennoinen, A. and T. Teräsvirta (2009). "Multivariate GARCH models". *Handbook of Financial Time Series*. Springer, pp. 201–229.

Weide, R. van der (2002). "GO-GARCH: a multivariate generalized orthogonal GARCH model". *Journal of Applied Econometrics* 17 (5), pp. 549–564.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.