DYNAMIC SCORE DRIVEN INDEPENDENT COMPONENT ANALYSIS

Christian M. Hafner, Helmut Herwartz

REPRINT | 2022 / 10

UCLouvain

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

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Dynamic score driven independent component analysis

Christian M. Hafner*　　Helmut Herwartz†

February 24, 2022

Abstract

A model for dynamic independent component analysis is introduced where the dynamics are driven by the score of the pseudo likelihood with respect to the rotation angle of model innovations. While conditional second moments are invariant with respect to rotations, higher conditional moments are not, which may have important implications for applications. The pseudo maximum likelihood estimator of the model is shown to be consistent and asymptotically normally distributed. A simulation study reports good finite sample properties of the estimator, including the case of a mis-specification of the innovation density. In an application to a bivariate exchange rate series of the Euro and the British Pound against the US Dollar, it is shown that the model-implied conditional portfolio kurtosis largely aligns with narratives on financial stress as a result of the global financial crisis in 2008, the European sovereign debt crisis (2010-2013) and early rumors signalling the UK to leave the European Union (2017). These insights are consistent with a recently proposed model that associates portfolio kurtosis with a geopolitical risk factor.

Keywords: structural vector autoregressions, multivariate GARCH, portfolio selection, risk management

---

*Corresponding author, Louvain Institute of Data Analysis and Modeling in Economics and Statistics (LIDAM), and ISBA, Université catholique de Louvain, Belgium. e-mail `christian.hafner@uclouvain.be`

†Department of Economics, University of Göttingen, Germany. e-mail `hherwartz@uni-goettingen.de`.

1 Introduction

Independent component analysis is a powerful method with widespread applications in economics, finance, management, econometrics, and statistics. For example, one important field in econometrics is the identification of (hidden) structural relations in vector autoregressive models (VARs). It is well known at least since Lancaster (1954) and Comon (1994) that observed linear combinations of independent random variables allow to recover uniquely the independent components if at most one of their distributions is Gaussian. Moneta et al. (2013) used this result to assign loss measures to recursive (i.e. lower triangular) structural relations, thereby promoting data-based variable ordering in structural VARs (SVARs). Going beyond the framework of recursive models, Lanne et al. (2017) suggest a maximum likelihood approach to SVAR identification that rests upon explicit assumptions about the distributional features of underlying stochastic model components, i.e., the structural shocks. Addressing likelihood mis-specification as a typical problem in maximum likelihood estimation, Gouriéroux et al. (2017) provide a theoretical framework that establishes consistency and asymptotic normality of a pseudo maximum likelihood (PML) estimator. With few exceptions such as Baumeister and Peersman (2013), the literature on SVARs in general and the identification of independent components in particular builds upon the assumption that the structural relations of interest are time invariant. This paper contributes to the scarce literature on dynamic ICA by introducing a flexible model that is motivated by recent advances in time series econometrics. An overview of approaches to SVAR identification is given by Kilian and Lütkepohl (2017).

In this paper we introduce a new dynamic model for independent component analysis where the dynamics are driven by the score of the pseudo likelihood with respect to the rotation angle of structural innovations. Score driven dynamic models have been extensively used and discussed in the econometrics literature since their introduction by Creal et al. (2011) and Harvey (2013), and depending on their origin are either called dynamic conditional score (DCS) or generalized autoregressive score (GAS) models. While score driven rotations of independent components have scope for the identification of low frequency SVARs in longitudinal samples, we consider the orthogonalized innovations as the stochastic components driving multivariate volatility models in GARCH form, i.e. multivariate GARCH (MGARCH).

In MGARCH modelling, (co-)variances of vector valued speculative returns are quantified by means of conditioning on a filtration that collects the history of the return process. Yet, the family of MGARCH specifications comprises a rich set of alternative ways to relate the conditional covariance to the filtration under scrutiny (see e.g. Bauwens et al. (2006) for an overview of this literature). While the dynamic covariance profile implied by MGARCH structures can be estimated consistently by means of quasi ML (QML) estimation, the pattern of how stochastic model components enter the observable speculative returns poses an identification problem.

On the basis of ad-hoc assumptions, prominent suggestions to solve this identification problem have been made, for instance, by van der Weide (2002) (generalized orthogonal GARCH), Weber (2010) (structural conditional correlation model), Hu and Tsay (2014) (principal volatility component analysis) and Matteson and Tsay (2011) (dynamic orthogonal components). Building upon data-characteristics, Hafner and Herwartz (2006) used the identification of independent components to define unique volatility impulse response functions. Recently, Hafner et al. (2020) have suggested to identify independent components in non-Gaussian MGARCH models by means of a static rotation of ad-hoc identification schemes such as spectral or Cholesky decompositions. While all these identification approaches leave second order features such as portfolio variances unaffected, rotations of independent components under non-Gaussianity are key in shaping higher order properties of portfolio returns. The novelty of our approach is the introduction of a natural updating mechanism for the dynamic rotations, extending the existing literature of ICA to a genuine dynamic framework.

The PML estimator of our dynamic rotation model is shown, under weak regularity conditions, to be consistent and asymptotically normally distributed. We provide an extensive simulation study in Appendix D of the supplementary online document which confirms our theoretical results and demonstrates the good finite sample performance of the PML estimator. Estimation of its finite sample covariance is noisy and converges only slowly to its asymptotic limit. As a practical recommendation we propose a bootstrap method to obtain a reliable means for finite sample inference.

We consider the dynamic independent components to lay the ground for an improved understanding of the (higher order) stochastic properties of vector valued speculative returns. Score driven orthogonalizations are convenient to define time varying risk measures such as the conditional skewness and kurtosis, Value-at-Risk and expected shortfall. Put differently, the suggested identified MGARCH model allows for a more complete understanding of distributional characteristics of speculative (portfolio) returns which are essential for risk modelling. We apply the dynamic rotation model to a bivariate foreign exchange (FX) rate series of the Euro and the British Pound against the US Dollar. Orthogonalized residuals are obtained by fitting a multivariate GARCH model. The conditional kurtosis estimates of an equally weighted portfolio implied by the dynamic rotation model are highly correlated with events at international financial markets. In particular, excess portfolio kurtosis appears in the context of (i) the global financial crisis of 2008, (ii) the European sovereign debt crisis (2010-2013) and (iii) early rumors about the Brexit (2017).

This is consistent with the recently proposed model of Engle and Martins (2020) that links the portfolio kurtosis to a global geopolitical risk factor. We find this link indeed using our dynamic rotation model, but not for models based on static rotations or ad-hoc decompositions, which highlights the importance of a correct model specification to recover dynamic independent components.

The remainder of the paper has the following structure. Sections 2 and 3 introduce the model and pseudo ML estimation, respectively. Section 4 consists of an empirical analysis of a bivariate FX system. Section 5 comments on the generalisation of the bivariate model towards higher dimensions. Section 6 summarizes and concludes. Figures are delegated to Appendix A, and tables to Appendix B. Proofs of the theorems are collected in Appendix C, and a Monte-Carlo experiment to analyse the finite sample properties of PML estimation in Appendix D. Appendices C and D are available in a separate document that is available online.

2 The score driven dynamic rotation model

Let $\{e_t\}, t \in \mathbb{Z}$ be a centered and standardized bivariate time series process, i.e. we assume that the mean of e_t conditional on past observations is zero, and that the conditional variance-covariance matrix of e_t is given by the identity matrix. We think of e_t as the standardized residual series of a model for the conditional mean and variance-covariance matrix, such as for example a VAR-GARCH model. The bivariate case is considered for notational and computational simplicity, but Section 5 provides basic expressions for the generalization to higher dimensions. We now consider the model

$$e_t = R(\delta_t)\xi_t, \tag{2.1}$$

where $\xi_t \sim i.i.d.(0, I_2)$, non-Gaussian, are the innovations that are assumed to be componentwise independent. The matrix $R(\delta_t)$ is a rotation matrix of the form

$$R(\delta_t) = \begin{pmatrix} \cos(\delta_t) & -\sin(\delta_t) \\ \sin(\delta_t) & \cos(\delta_t) \end{pmatrix}$$

with rotation angle δ_t specified as a score driven autoregressive process of order one, i.e.,

$$\delta_t = \omega(1-\beta) + \beta\delta_{t-1} + \kappa u_{t-1}, \quad \omega \in [0, \pi), |\beta| < 1, \kappa \geq 0 \tag{2.2}$$

with location parameter ω, persistence parameter β, and updating or forcing parameter κ. The forcing variable u_t will be defined as the score of the likelihood with respect to δ_t, discussed in more detail below. Since u_t is a martingale difference sequence, stationarity of δ_t is ensured by the restriction $|\beta| < 1$. The unconditional mean of δ_t is given by the location parameter ω, which is restricted to $[0, \pi)$ rather than $[-\pi, \pi)$, because otherwise it would not be identifiable under symmetric distributions of ξ_t.

In distributed lag form, we can write model (2.2) equivalently as

$$\delta_t = \omega + \sum_{j=1}^{\infty} \psi_j u_{t-j}, \tag{2.3}$$

with lag coefficients $\psi_j = \kappa\beta^{j-1}$, or in lag polynomial form $\delta_t = \omega + \psi(L)u_t$ with lag operator L and $\psi(L) = \psi_1 L + \psi_2 L^2 + \ldots$.

We collect the parameters of the model in the vector $\theta = (\omega, \beta, \kappa)'$, and denote the true parameter by θ_0. Let $\xi_{it}(\theta) := R_i'(\delta_t(\theta))e_t$, be the pseudo innovations for a given parameter θ, where R_i' denotes the i-th row of R'. Also, let $\delta_t = \delta_t(\theta)$ and $\delta_{0t} = \delta_t(\theta_0)$ for simplicity of notation. The relations between the pseudo and true innovations are given by

$$\xi_{1t}(\theta) = \cos(\Delta_t)\xi_{1t} + \sin(\Delta_t)\xi_{2t} \tag{2.4}$$

$$\xi_{2t}(\theta) = -\sin(\Delta_t)\xi_{1t} + \cos(\Delta_t)\xi_{2t}, \tag{2.5}$$

where $\Delta_t := \delta_t - \delta_{0t}$.[1] Of course, at the true parameter, $\Delta_t = 0$ and hence $\xi_{it}(\theta_0) = \xi_{it}$, $i = 1, 2$. For later reference, note also that, using elementary formulas,

$$\xi^2_{1t}(\theta) = \cos^2(\Delta_t)\xi^2_{1t} + \sin^2(\Delta_t)\xi^2_{2t} + \sin(2\Delta_t)\xi_{1t}\xi_{2t} \tag{2.6}$$

$$\xi^2_{2t}(\theta) = \sin^2(\Delta_t)\xi^2_{1t} + \cos^2(\Delta_t)\xi^2_{2t} - \sin(2\Delta_t)\xi_{1t}\xi_{2t} \tag{2.7}$$

and that $\xi^2_{1t}(\theta) + \xi^2_{2t}(\theta) = \xi^2_{1t} + \xi^2_{1t}$.

The innovation term u_t in (2.2) is specified in the spirit of dynamic conditional score models as in Harvey (2013) and Creal et al. (2011). Hence, it will depend on the particular distribution of ξ_{it} used for estimation by pseudo maximum likelihood. Here, u_t is the derivative of the log likelihood function with respect to the rotation angle δ_t. More precisely,

$$u_t = \sum_{i=1}^{2} \frac{\partial \log g_i(\xi_{it}(\theta))}{\partial \xi_{it}} \frac{\partial \xi_{it}(\theta)}{\partial \delta_t},$$

where $g_i(\cdot)$ are pseudo densities of ξ_{it}.

We will use the following assumptions.

(A1) The random vector ξ_t is i.i.d. with mean zero, variance unity, and independent components ξ_{1t} and ξ_{2t}.

(A2) The score u_t has a non-degenerate distribution with $\mathrm{E}|u_t|^s < \infty$ for some $s > 0$.

[1] This uses elementary formulas such as $\cos(a)\cos(b) + \sin(a)\sin(b) = \cos(a - b)$ and $-\sin(a)\cos(b) + \cos(a)\sin(b) = -\sin(a - b)$.

(A3) At most one of the distributions of ξ_{1t} and ξ_{2t} is Gaussian.

Assumption (A3) is the crucial identification assumption in independent component analysis without imposing further restrictions, see e.g. Gouriéroux et al. (2017) and Lanne et al. (2017). Assumption (A2) ensures that δ_t is strictly stationary and ergodic, since at the true parameter, u_t is i.i.d. with a non-degenerate distribution and finite fractional moments, and hence, δ_t is a stationary autoregressive process of order one with i.i.d. increments. Furthermore, Lemma 1 in Appendix C shows that under Assumptions (A1)-(A3), the model (2.1) is identifiable up to a permutation of the columns of $R(\delta_t)$. In the next section, we will make stronger assumptions about the distributions of ξ_{1t} and ξ_{2t} that ensure unique identifiability.

3 Estimation and inference

In this section, we will first present the estimation method by pseudo maximum likelihood, and then discuss its asymptotic properties in terms of consistency and asymptotic normality.

3.1 Pseudo maximum likelihood estimation

Denote the parameter vector by $\theta = (\omega, \beta, \kappa)'$. The log-likelihood function for a sample of T observations $e_1, \ldots, e_T$ is given by

$$\log L(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{2} \log g_i(\xi_{it}(\delta_t), \theta) =: \sum_{t=1}^{T} l_t(\theta), \tag{3.1}$$

where $g_i(\cdot)$ are pseudo densities of the components ξ_{it}. The PML estimator is defined as $\widehat{\theta} = \arg\max_\theta \log L(\theta)$.

We use a standardized Student's t-distribution with log-density given by

$$\log g_i(\xi_{it}(\delta_t)) = \log \Gamma\left(\frac{\nu_i + 1}{2}\right) - \log \Gamma\left(\frac{\nu_i}{2}\right) - \frac{1}{2}\log(\pi(\nu_i - 2)) - \frac{\nu_i + 1}{2} \log\left(1 + \frac{\xi_{it}(\delta_t)^2}{\nu_i - 2}\right),$$

with degrees of freedom coefficients $\nu_i > 2, i = 1, 2$.

The score with respect to δ_t is

$$\frac{\partial \log L(\theta)}{\partial \delta_t} = \sum_{t=1}^{T} \sum_{i=1}^{2} \frac{\partial \log g_i(\xi_{it}(\delta_t), \theta)}{\partial \xi_{it}} \frac{\partial \xi_{it}}{\partial \delta_t}$$

with

$$\frac{\partial \log g_i(\xi_{it}(\delta_t))}{\partial \xi_{it}} = -\frac{(\nu_i + 1)\xi_{it}(\delta_t)}{\nu_i - 2 + \xi_{it}(\delta_t)^2}.$$

We have $\frac{\partial \xi_{1t}}{\partial \delta_t} = \xi_{2t}$ and $\frac{\partial \xi_{2t}}{\partial \delta_t} = -\xi_{1t}$, so that the score is given by

$$u_t := \frac{\partial l_t}{\partial \delta_t} = \frac{(\nu_2 + 1)\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_2 - 2 + \xi_{2t}(\delta_t)^2} - \frac{(\nu_1 + 1)\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_1 - 2 + \xi_{1t}(\delta_t)^2}. \tag{3.2}$$

Note that

$$\frac{\partial u_t}{\partial \theta} = \frac{\partial u_t}{\partial \delta_t} \frac{\partial \delta_t}{\partial \theta},$$

where (dependence on δ_t is suppressed for simplicity)

$$\frac{\partial u_t}{\partial \delta_t} = \frac{(\nu_2 + 1)\left\{(\nu_2 - 2)(\xi_{2t}^2 - \xi_{1t}^2) + \xi_{2t}^4 + \xi_{1t}^2\xi_{2t}^2\right\}}{\left\{\nu_2 - 2 + \xi_{2t}^2\right\}^2} + \frac{(\nu_1 + 1)\left\{(\nu_1 - 2)(\xi_{1t}^2 - \xi_{2t}^2) + \xi_{1t}^4 + \xi_{1t}^2\xi_{2t}^2\right\}}{\left\{\nu_1 - 2 + \xi_{1t}^2\right\}^2} \tag{3.3}$$

and

$$\frac{\partial \delta_t}{\partial \theta} = A z_{t-1} + x_{t-1} \frac{\partial \delta_{t-1}}{\partial \theta} \tag{3.4}$$

with the notation

$$A = \begin{pmatrix} 1-\beta & 0 & 0 \\ -\omega & 1 & 0 \\ 0 & 0 & 1, \end{pmatrix}, \quad z_t = \begin{pmatrix} 1 \\ \delta_t \\ u_t \end{pmatrix}, \quad \text{and } x_t = \beta + \kappa \frac{\partial u_t}{\partial \delta_t}. \tag{3.5}$$

Finally,

$$\frac{\partial l_t}{\partial \theta} = \frac{\partial l_t}{\partial \delta_t} \frac{\partial \delta_t}{\partial \theta}, \tag{3.6}$$

where $\frac{\partial l_t}{\partial \delta_t}$ is given in (3.2) and $\frac{\partial \delta_t}{\partial \theta} = (\frac{\partial \delta_t}{\partial \omega}, \frac{\partial \delta_t}{\partial \beta}, \frac{\partial \delta_t}{\partial \kappa})'$ in (3.4).

3.2 Consistency

First we give our additional assumptions. We denote the true parameter by θ_0.

(A4) The parameter vector $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^3$ such that $|\beta| \leq \bar{\beta} < 1$ and $\omega \in [0, \pi)$.

(A5) The pseudo distributions g_i of ξ_{it} are different for $i = 1, 2$.

(A6) $\mathrm{E}|\log g_i(\xi_{it})| < \infty$ and $\nu_i \geq 2 + \epsilon$, $\epsilon > 0$, for $i = 1, 2$.

(A7) The (true) distribution of ξ_t is such that for all $c > 0$, $a \in (0, 1)$ and $b \in (-1, 1)$,

$$\mathrm{E}\log(c + \xi_{it}^2) - \mathrm{E}\log\left(c + a\xi_{it}^2 + (1-a)\xi_{jt}^2 + b\xi_{it}\xi_{jt}\right) \leq 0.$$

Assumptions (A4) and (A5) are standard to ensure stationarity, ergodicity and identifiability of the model. (A5) avoids un-identifiability due to swapping the components of ξ_t and columns of $R(\delta_t)$, where $\omega \pm \pi/2$ would give the same likelihood value if (A5) did not hold. (A6) is a mild existence of moments assumption for the pseudo-densities used for estimation. (A7) is used to ensure a unique maximum of the likelihood function at the true parameter. It is also a mild assumption that can be verified for a variety of distributions, including the Student-t distributions that we use in this paper. We can use these assumptions to establish the following strong consistency result.

Theorem 1. *Under Assumptions (A1)-(A7),* $\hat{\theta} \to \theta_0 \quad$ *a.s.*

Proof: see Appendix C.

3.3 Asymptotic normality

To establish asymptotic normality of the PML estimator, we make the following additional assumptions.

(A8) The parameter θ_0 lies in the interior of Θ.

(A9) $\mathrm{E}||\xi_t||^4 < \infty$.

(A10) $|\beta + \kappa \mathrm{E}(\partial u_t/\partial \delta_t)| < 1$.

(A11) $|\beta^2 + 2\kappa\beta \mathrm{E}(\partial u_t/\partial \delta_t)| + \kappa^2 \mathrm{E}[(\partial u_t/\partial \delta_t)^2] < 1$.

Assumptions (A8) and (A9) are standard assumptions in time series models to establish asymptotic normality of PML-type estimators. Assumptions (A10) and (A11) are typical for score driven models, essentially ensuring that the score and its first derivative have finite second order moments, see e.g. Chapter 2 of Harvey (2013). Note that (A10) is implied by (A11) using Jensen's inequality, but for some results in the proofs only (A10) will be necessary. We can use these assumptions to establish the following result.

Theorem 2. *Under Assumptions (A1)-(A11),*

$$\sqrt{T}(\hat{\theta} - \theta_0) \to_d N(0, V(\theta_0)), \tag{3.7}$$

where $V(\theta) = J(\theta)^{-1} I(\theta) J(\theta)^{-1}$ *and* $I(\theta) = E\left[\frac{\partial l_t}{\partial \theta}\frac{\partial l_t}{\partial \theta'}\right]$, $J(\theta) = -E\left[\frac{\partial^2 l_t}{\partial \theta \partial \theta'}\right]$. *An explicit expression for the asymptotic covariance matrix is given by*

$$V(\theta) = \frac{(1-b)\sigma_u^2}{\tau^2}\begin{pmatrix} \frac{(1-\beta)^2(1+a)}{1-a} & \frac{ac\kappa(1-\beta)}{(1-a)(1-a\beta)} & \frac{c(1-\beta)}{1-a} \\ \frac{ac\kappa(1-\beta)}{(1-a)(1-a\beta)} & \frac{\kappa^2\sigma_u^2(1+a\beta)}{(1-\beta^2)(1-a\beta)} & \frac{a\kappa\sigma_u^2}{1-a\beta} \\ \frac{c(1-\beta)}{1-a} & \frac{a\kappa\sigma_u^2}{1-a\beta} & \sigma_u^2 \end{pmatrix}^{-1}, \tag{3.8}$$

where $c = \kappa E(u_t \frac{\partial u_t}{\partial \theta})$, $\tau = -E\frac{\partial u_t}{\partial \delta_t}$, $\sigma_u^2 = E(u_t^2)$, $a = \beta + \kappa E \partial u_t/\partial \delta_t$ *and* $b = E(\beta + \kappa E \partial u_t/\partial \delta_t)^2$.

Proof: see Appendix C.

Note that $-J(\theta) = -\tau \mathrm{E}\frac{\partial \delta_t}{\partial \theta}\frac{\partial \delta_t}{\partial \theta'}$ and $I(\theta) = \sigma_u^2 \mathrm{E}\frac{\partial \delta_t}{\partial \theta}\frac{\partial \delta_t}{\partial \theta'}$. If the pseudo density functions $g_i(\cdot), i = 1, 2$, correspond to the true densities of ξ_{1t} and ξ_{2t}, then by the information equality, $\tau = \sigma_u^2$, and hence, $I = J$. In that case the asymptotic covariance matrix reduces to the inverse of the information matrix, I^{-1}.

An estimator of V can be based on the sample analogues of I and J, $\widehat{I}$ and $\widehat{J}$, respectively, evaluated at $\hat{\theta}$, i.e., $\widehat{I} = T^{-1}\sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta})}{\partial \theta}\frac{\partial l_t(\hat{\theta})}{\partial \theta'}$ and $\widehat{J} = T^{-1}\sum_{t=1}^{T} \frac{\partial^2 l_t(\hat{\theta})}{\partial \theta \partial \theta'}$. Alternatively, we can use a plug-in estimator of V using the expression (3.8), i.e. $\widehat{V} := V(\hat{\theta})$, where the moments c, τ, σ_u^2, a and b are replaced by sample moments. The Monte Carlo section in the supplementary appendix investigates the finite sample performance of the estimators $\hat{\theta}$ and $\widehat{V}$.

4 Empirical application

In this section we provide an empirical illustration to highlight the potential of score driven rotations for understanding the higher order characteristics of time series data. We first subject a bivariate FX rate series to an extraction of orthogonalized model residuals as implied by an estimated multivariate GARCH model. Subsequently, we extract independent components by means of score driven rotations of orthogonalized model residuals, and discuss diagnostic features of the model. Finally, we illustrate how the estimated model can be used (i) to evaluate higher order characteristics of portfolio returns, and (ii) to quantify portfolio risks.

4.1 The data and an ad-hoc orthogonalization

We analyse a bivariate series of two FX rates against the US Dollar: The price of the US Dollar in terms of the Euro (r_{1t}) and the British Pound (r_{2t}). The sample comprises 4956 daily price quotes and covers the period from Jan 3rd, 2000 until Dec 31st, 2018. The data are drawn from Datastream (codes S00057 (EUR2USD) and Y79935 (GBP2USD)).[2] We denote the time series of centered log differences (or returns) as $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$.

Prior to dynamic independent component analysis we need to take conditional heteroskedasticity and correlation into account.[3] For the extraction of orthogonalized model residuals e_t from ε_t we employ a bivariate GARCH model in the parsimonious (unrestricted) BEKK(1,1) form (Engle and Kroner (1995), see Bauwens et al. (2006) for a review of multivariate GARCH models). [4] Denoting the information set available up to time $t-1$ by $\mathcal{F}_{t-1}$, and $H_t := \mathrm{Var}(\varepsilon_t|\mathcal{F}_{t-1})$, we employ the specification

$$H_t \quad = \quad CC' + A'\varepsilon_{t-1}\varepsilon_{t-1}'A + G'H_{t-1}G, \tag{4.1}$$

where A, C and G are 2×2 parameter matrices. Moreover, C is lower triangular. The resulting estimates (joint with QML t-ratios) are

$$C = \begin{pmatrix} \text{5.8E-04} & 0 \\ (1.57) & (-) \\ \text{7.7E-04} & \text{4.0E-04} \\ (1.22) & (6.54) \end{pmatrix}, A = \begin{pmatrix} 0.174 & 0.013 \\ (4.40) & (0.21) \\ 0.003 & 0.210 \\ (0.062) & (2.63) \end{pmatrix}, G = \begin{pmatrix} 0.984 & -0.001 \\ (122.6) & (-0.09) \\ -0.008 & 0.962 \\ (-0.41) & (28.9) \end{pmatrix}.$$

Subsequent to estimating the BEKK model parameters and the implied covariance path $\{H_t\}_{t=1}^T$, we determine orthogonalized model residuals $e_t = H_t^{-1/2}\varepsilon_t$, where

[2]Both FX series are closing spot rates fixed in London at 4pm GMT, source: https://www.refinitiv.com/en/financial-data/financial-benchmarks/wm-reuters-fx-benchmarks. These rates are determined over a one-minute fix period, from 30 seconds before to 30 seconds after the time of the fix (4:00 pm in London). During this one-minute window, bid and offer rates from the order matching system and actual trades executed are captured. The median bid and offer are calculated using valid rates over the fix period, and the mid-rate is then calculated from them.

[3]As usual for this kind of data, the dynamic properties of the mean returns are much less informative than the second order properties, and, hence, we assume time invariance of the conditional mean. Accordingly, centered log FX rate changes are subjected to covariance modelling.

[4]BEKK models of higher order turned out to unnecessary. We tested the autocorrelation of the squares and cross-products of standardized residuals using a multivariate Portmanteau test with 15 lags, but with a p-value of 0.09 remaining autocorrelation was insignificant.

$H_t^{1/2}$ is the symmetric square root matrix of the conditional covariance matrix H_t, obtained by spectral decomposition of H_t. The number of observations available for e_t is $T = 4955$.

4.2 Independent component analysis

For PML estimation of the dynamic rotation model we take advantage of the outcomes of the Monte-Carlo analysis in Appendix D. In particular, we (i) determine a global PML estimator as the best outcome from four alternative initializations of the BHHH algorithm,[5] and (ii) condition the estimation of model parameters on a rich set of alternative combinations of degrees of freedom parameters. To be specific, we take as the PML estimator the best performing outcome obtained from 72 scenarios (each implemented with four alternative starting values), where the presumed degrees of freedom are taken from the set $\mathcal{N}^{(0)} = \{\nu_1^{(0)}, \nu_2^{(0)} | \nu_1^{(0)} \in \{3, 4, \ldots, 11\}, \nu_2^{(0)} \in \{3, 4, \ldots, 11\}, \nu_2^{(0)} \neq \nu_1^{(0)}\}$.

Before estimating the dynamic rotation model, we determine log-likelihood statistics for two more restrictive model specifications. First, we use the space of degrees of freedom $\mathcal{N}^{(0)}$ to determine a distributional model for unrotated orthogonalized residuals (i.e. $\delta_t = 0$ for all t) which are best in-line with the independence assumption. Combining the degrees of freedom $\nu_1^{(0)} = 9$ and $\nu_2^{(0)} = 7$ yields a log-likelihood value of -13919.65. Second, we use a grid of 100 points on the interval $[0, \pi)$ to determine the best static rotation, i.e. $\delta_t = \delta$ for all t, where δ is constant. It turns out that a static rotation offers a sizeable improvement of the unrotated model in terms of respective log-likelihood values. Conditional on the degrees of freedom $\nu_1^{(0)} = 9$ and $\nu_2^{(0)} = 7$, the log-likelihood increases to -13909.42 if $\delta = 0.08\pi \approx 0.25\,(= \hat{\omega})$. Hence, in comparison with the unrotated model the static rotation improves the

[5] We set $\theta^{(0)} = (1.5, 0.95, 0.02)'$, $(2.25, 0.98, 0.01)'$, $(1.13, 0.70, 0.15)'$ and $(1.35, 0.90, 0.10)'$ alternatively.

log-likelihood by approximately 10.2, delivering a likelihood ratio (LR) statistic of 20.5. This statistic which would be highly significant at conventional levels when compared with critical values of a $\chi^2(1)$ limit distribution, provides strong evidence against the unrotated model irrespective of the unknown distribution.

For the model with static rotation we obtain a sample average of the derivative of the scores $\partial u_t/\partial\delta$ of -0.061 and a corresponding t-statistic of -2.46. An LM-test for serial correlation of $\partial u_t/\partial\delta$, i.e., a Portmanteau-statistic with five lags, gives a test statistic of 9.68 which is significant at the 10% level (p-value of 0.085). Hence, we reject the null hypothesis of constant rotation angles δ_t. Estimating the model with score-driven dynamic rotations provides a further improvement of the log-likelihood to obtain -13895.05 for the choice of $\nu_1^{(0)} = 9$ and $\nu_2^{(0)} = 6$. The estimated model parameters joint with PML standard errors are given in Table B.1.

Insert Table B.1 around here

Both documented asymptotic standard errors using either $\widehat{I}^{-1}$ or $\widehat{V}$ are small, but may underestimate the finite sample standard errors as shown in the simulation exercise in the supplementary appendix. As a complement, we therefore employ bootstrap estimation of model parameters and their variances, which are also reported in Table B.1. Bootstrap samples are obtained by considering the PML estimates as the true parameters. Bootstrap samples $\{e_t^*\}_{t=1}^T$ are determined with bootstrap innovations $\{\xi_t^* = (\xi_{1t}^*, \xi_{2t}^*)'\}_{t=1}^T$ and according to the model structure of (2.1) and (2.2). Building on the presumed independence, samples of the elements ξ_{it}^*, $i = 1, 2$, are drawn with replacement from the marginal distributions of the corresponding sample quantities ξ_{it}. After sampling $\{e_t^*\}_{t=1}^T$, these data are subjected to the same PML estimation steps as the original sample of orthogonalized model residuals. The number of bootstrap replications is 199. Note that the bootstrap means of all model parameters are close to the PML estimates. The documented bootstrap standard error for κ is close to its data-based PML counterparts. This provides further support of our claim that κ is estimated with sufficient precision and can be regarded as significantly larger than zero.

Figure A.1 (upper panel) shows the time series of estimated rotation angles $\hat{\delta}_t$ joint with the time invariant rotation angle implied by the static model. Owing to an estimate of β close to unity, dynamic rotations exhibit a high degree of persistence, similar to classical volatility or correlation models. Moreover, the responses of δ_t to news or innovations in the score are relatively strong, given a sizeable estimate of $\kappa \approx 0.10$, which might even be slightly underestimated given the simulation results documented in Table D.2.

Insert Figure A.1 around here

For the series of estimated rotation angles $\{\hat{\delta}_t\}$ we can determine its stationary distribution by means of nonparametric methods such as kernel smoothing. A natural candidate for a parametric distribution is the von Mises distribution, whose density is given by

$$f(\delta|\omega, \upsilon) = \frac{\exp(\upsilon \cos(\delta - \omega))}{2\pi I_0(\upsilon)},$$

where ω is the mean of δ, and υ is a concentration parameter. The function I_0 is the modified Bessel function of order 0. In our case, the location parameter is estimated by $\hat{\omega}$, while υ could be estimated by maximizing the concentrated likelihood function, up to an additive constant with respect to υ, i.e.

$$L(\upsilon|\hat{\omega}) = \upsilon \sum_{t=1}^{T} \cos(\delta_t - \hat{\omega}) - T \log(I_0(\upsilon)). \tag{4.2}$$

This gives an estimate of $\hat{\upsilon} = 18.8$. One generally considers the Gaussian distribution as a good approximation for the von Mises distribution for $\upsilon > 2$, which is the case here. Figure A.3 shows the good approximation of the nonparametric kernel density estimator by the fitted von Mises distribution.

Insert Figure A.3 around here

4.3 Conditional kurtosis and portfolio selection

Time dependence of the optimal rotations is of interest in its own, but has also important consequences for higher order characteristics of return processes, for example the conditional fourth order moments. While conditional second order moments are invariant with respect to rotations, we emphasize that conditional fourth order moments are not. However, they may play an important role in applications, for example higher order approximations of Value-at-Risk using Edgeworth or Cornish-Fisher type expansions. Another application is portfolio selection, as we discuss in the following. For high-dimensional stock portfolios, DeMiguel et al. (2009) have shown that it is difficult to outperform an equally weighted portfolio out-of-sample in terms of Sharpe ratios because of estimation errors. For exchange rates, however, Ackermann et al. (2017) argue that mean-variance portfolio optimization is preferable due to the availability of interest rate spreads that provide a predictable component of exchange rate returns.

For the sake of generality, consider a portfolio constructed from N assets,

$$\tau_t = \sum_{i=1}^{N} w_{it}\varepsilon_{it},\ t = 1, 2, \ldots, T, \tag{4.3}$$

where w_{it} are $\mathcal{F}_{t-1}$-adapted processes such that $\sum_{i=1}^{N} w_{it} = 1$. Suppose that the conditional mean of ε_t is zero, and that the conditional covariance matrix is given by H_t. Standard portfolio selection à la Markowitz (1952) relies on the first and second moment structure of the universe of investment assets. At least since Chan et al. (1999) and Jagannathan and Ma (2003) it is well known that the global minimum variance (GMV) portfolio is performing well out-of-sample not only in terms of volatility, but also taking average returns into account, e.g. in terms of a Sharpe ratio. In a dynamic context, the conditional variance of a portfolio with weight vector $w_t = (w_{1t}, w_{2t}, \dots, w_{Nt})'$ at time t is given by $\sigma_t^2 := w_t' H_t w_t$, and the GMV portfolio weights are

$$w_t^{GMV} = \frac{H_t^{-1}\iota}{\iota' H_t^{-1}\iota}, \tag{4.4}$$

where $\iota_N := (1, 1, \dots, 1)'$ is an N-dimensional vector of ones.

Recently, the consideration of higher moments such as portfolio skewness and kurtosis has become a focus of interest. For example, in the model of Engle and Martins (2020) a portfolio that reduces exposure to a global geopolitical risk factor would reduce the portfolio kurtosis compared with a portfolio that only considers a variance constraint. In the light of the previous arguments, one might be interested in combining portfolio selection based on volatility minimization with a conditional kurtosis criterion. In the one extreme case where only volatility is minimized, we obtain GMV portfolios as in (4.4) that are invariant with respect to δ_t. In the other extreme case where only the portfolio kurtosis $K_t(\delta_t, w)$ is minimized at time t, conditional on past information, the weights of the global minimum kurtosis (GMK) portfolio would be

$$w_t^{GMK} = \arg\min_w K_t(\delta_t, w).$$

While analytical solutions for w_t^{GMK} are not available, efficient numerical algorithms have been proposed, such as the difference of convex curves (DC) algorithm of Pham dinh and Niu (2011) and the branch and bound algorithm of Barkhagen et al. (2019). Because $K_t(\delta_t, w)$ depends on δ_t, the GMK-optimal weights are not invariant with respect to the rotation angles at time t. The particular form of this functional dependence will be derived below.

In order to balance volatility- and kurtosis-minimal portfolios, we could interpolate the GMV and GMK portfolios using $\alpha \in [0,1]$ as $\alpha w_t^{GMV} + (1-\alpha) w_t^{GMK}$. Smaller values of α would put more weight on the kurtosis and hence reduce exposure to tail risk and geopolitical risk factors in the sense of Engle and Martins (2020). The choice of α might be motivated by particular utility functions such as CARA or CRRA.[6]

To compute the conditional kurtosis of portfolio returns, denote the N-vector of kurtosis coefficients of ξ_t by $m^{(4)}$. Under the assumption of mutually independent components of ξ_t let

$$\Psi := \mathrm{E}[(\xi_t \xi_t') \otimes (\xi_t \xi_t')] = 2 D_N D_N^+ + \mathrm{vec}(I_N)\mathrm{vec}(I_N)' + \mathrm{diag}(\mathrm{vec}(\mathrm{diag}(m^{(4)} - 3\iota_N))), \tag{4.5}$$

where D_N is the duplication matrix, and D_N^+ its generalized inverse. In our bivariate exchange rate application, this matrix reduces to

$$\Psi = \begin{pmatrix} m_1^{(4)} & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & m_2^{(4)} \end{pmatrix}. \tag{4.6}$$

Then, by Theorem 2 of Hafner et al. (2020), the conditional kurtosis of portfolio returns is a non-trivial function of δ_t and given by $K_t(\delta_t, w_t) = \varsigma_t(\delta_t, w_t)/\sigma_t^4$, where

$$\varsigma_t(\delta_t, w_t) := E[\tau_t^4 | \mathcal{F}_{t-1}] = \mathrm{vec}(w_t w_t')'(A_t \otimes A_t)\Psi(A_t' \otimes A_t')\mathrm{vec}(w_t w_t')], \tag{4.7}$$

$A_t = H_t^{1/2} R(\delta_t)$, and Ψ is defined in (4.5).

Obviously, the conditional kurtosis depends on the fourth order moment structure of ξ_t given by Ψ. Considering the dynamic rotation model as the true data

[6] A thorough out-of-sample performance analysis of these selection strategies is beyond the scope of this paper but merits further research.

generating process, the only unknowns in this matrix Ψ are the marginal kurtosis coefficients collected in $m^{(4)}$, which can be estimated by the sample kurtosis coefficients of the components of ξ_t. To illustrate the time-varying behavior of the conditional kurtosis, we show in Figure A.1 the conditional kurtosis of an equal weight portfolio, for both the static and dynamic rotation model. Descriptive statistics characterizing the model implied kurtosis estimates of equal weight portfolios are provided in Table B.2. On average, portfolio kurtosis is larger if evaluated by means of the dynamic model. Moreover, the documented standard deviations show that score driven rotation shows up in much more pronounced time variation of portfolio kurtosis.

Insert Table B.2 around here

Using the dynamic rotation model to determine the conditional portfolio kurtosis provides a sharpened view at specific time episodes shaping higher order portfolio risks. For instance, we observe local peaks of portfolio kurtosis in 2008 and 2010 that are well in line with the global financial crisis and the European sovereign debt crisis, respectively. Since both events (or periods) can be seen as important historic marks of financial stress, it is intuitive to expect that these developments have impacted on the dynamic properties of two of the most liquid and important exchange rates. Interestingly, the empirical kurtosis statistics issued by the static rotation model clearly lack the sensitivity of the dynamic rotation model. From the respective display in the second panel of Figure A.1 we observe only mild increases of respective kurtosis estimates.[7] Further local peaks of kurtosis statistics obtained from the dynamic rotation model occur on November 6, 2013 ($K_t = 5.41$) and July

[7]Empirical kurtosis statistics peak on Dec 30th, 2008 ($K_t = 5.94$) and March 2nd, 2010 ($K_t = 5.66$). While the former date is close to the end of a four month period starting with the default of Lehman Brothers on Sept 15th, 2008, the latter date is close to the agreement of the Euro Group and the IMF on a first aid package for Greece on March 25th, 2010.

21, 2017 ($K_t = 5.42$). While the former date could be considered a residual of the European sovereign debt crisis, the latter can be associated with early rumors concerning the advent of the Brexit.[8]

Insert Table B.3 around here

We consider further examples of time invariant portfolio compositions with four alternative weights of EUR/USD $w_1 \in \{0.10, 0.25, 0.50, 0.75\}$, and $w_2 = 1 - w_1$. Tables B.2 and B.3 provide descriptive statistics for portfolio kurtosis and portfolio returns, respectively. With increasing weights w_1, the stochastic features of portfolio returns are less shaped by model innovations ξ_{2t} which are more leptokurtic than ξ_{1t}. For the model with dynamic conditional correlations, accordingly, average portfolio kurtosis is overall diminishing in the portfolio shares w_1 (e.g. in terms of the maximum kurtosis). Interestingly, almost all documented quantiles of empirical kurtosis statistics are larger for the dynamic model in comparison with the model employing a static rotation. Descriptive statistics summarizing the stationary distribution of returns of alternative portfolio compositions in Table B.3 highlight that in comparison with excess kurtosis, skewness of portfolio returns is much less pronounced for our FX portfolio. In absolute value, the largest skewness estimate of 0.47 obtains for portfolios that put strong weight on GBP/USD ($w_1 = 0.1$). Over four alternative portfolio compositions, the unconditional kurtosis of portfolio returns ranges from 6.11 to 13.0. Not surprisingly, it is increasing in the weight of GBP/USD, since ξ_{2t}, which are more leptokurtic than ξ_{1t}, load heavier on GBP/USD than on EUR/USD. As a result, the unconditional distribution of portfolio returns deviates strongly from the normal distribution. With increasing kurtosis, analytical approximations

[8] In March 2013 Euro group countries and the IMF launched an aid package for Cyprus. On July 21st, 2017 (!) Brian Moynihan, chief executive of the Bank of America announced to move the Bank's European headquarter from London to Dublin and referred to the UK's potential leave of the European Union to explain this strategic decision.

of established risk measures such Value-at-Risk or expected shortfall are likely to provide non-trivial improvements of normal quantiles or conditional expectations.

Insert Figure A.2 around here

Figure A.2 shows portfolio weights assigned to EUR/USD, denoted w_{1t}, according to the alternative minimizations of conditional portfolio variances on the one hand and conditional portfolio kurtosis on the other hand. Interestingly, while assigning small weights to EUR/USD is effective in reducing portfolio variance, such portfolios are subject to excess kurtosis in comparison with minimum kurtosis portfolios implied by the dynamic rotation model. To compose portfolios with smallest conditional kurtosis, it is effective to put higher weight on EUR/USD. Throughout, kurtosis minimizing portfolio weights of EUR/USD are varying around 75% and beyond an equal weight portfolio share of 50%. During periods that can be related to the global financial crisis, the European sovereign debt crisis and the Brexit campaign, assessing higher weights to EUR/USD allows to control both conditional portfolio variance and kurtosis at most time instances. Interestingly, using the static rotation model for minimization of portfolio kurtosis implies a time profile of portfolio weights which is very similar to minimum variance portfolio weights.

5 Generalization to higher dimensions

The generalization of the bivariate model outlined in Sections 2 and 3 to the case of N dimensions is straightforward. In model (2.1), $\{e_t\}$ is now an N-variate time series process, $\xi_t \sim i.i.d.(0, I_N)$, non-Gaussian, componentwise independent, and the rotation matrix is given by

$$R(\delta_t) = \prod_{j=1}^{N^*} R_j(\delta_{jt}), \quad \delta_t = (\delta_{1t}, \dots, \delta_{N^*t})',$$

where $N^* := N(N-1)/2$, and R_j are $(N \times N)$ matrices of the form

$$R_j(\delta_{jt}) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\delta_{jt}) & \cdots & -\sin(\delta_{jt}) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & \sin(\delta_{jt}) & \cdots & \cos(\delta_{jt}) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1. \end{pmatrix}$$

Each sub-matrix $R_j(\delta_{jt})$ is operating a rotation with angle δ_{jt} in the plane spanned by two components of the N-dimensional space, leaving the other components unchanged. The dynamic rotation angle δ_t is now an N^*-vector with components

$$\delta_{it} = \omega_i(1-\beta_i) + \beta_i \delta_{i,t-1} + \kappa_i u_{i,t-1}, \quad i = 1, \ldots, N^*,$$

where we assume separability in the sense that each rotation angle has its own dynamics, and there is no causality or spillover between different angles. Of course this could be generalized at the expense of more complicated estimation and inference, especially in high dimensions. As it stands, the number of parameters is $3N^* = O(N^2)$, and this will obviously be an issue in high dimensions. However, various restrictions are possible, such as, for example, setting the parameters characterizing the dynamics of δ_t, i.e. κ_i and β_i, to be identical across different planes. Another possibility is to define strategies for building groups for which the parameters are identical. We leave this important topic for future research.

In the general case, the log likelihood function takes the form

$$\log L(\theta) = \sum_{t=1}^{T} \sum_{i=1}^{N} \log g_i(\xi_{it}(\delta_{it}), \theta) =: \sum_{t=1}^{T} l_t(\theta)$$

and the score with respect to δ_t is

$$\frac{\partial \log L(\theta)}{\partial \delta_t} = \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\partial \log g_i(\xi_{it}(\delta_t), \theta)}{\partial \xi_{it}} \frac{\partial \xi_{it}}{\partial \delta_t}.$$

If a standardized Student-t distribution is used for the pseudo densities, then as before in the bivariate case,

$$\frac{\partial \log g_i(\xi_{it}(\delta_t))}{\partial \xi_{it}} = -\frac{(\nu_i + 1)\xi_{it}(\delta)}{\nu_i - 2 + \xi_{it}(\delta_t)^2},$$

while

$$\frac{\partial \xi_{it}}{\partial \delta_t} = \frac{\partial R'_{(i)}(\delta_t)}{\partial \delta_t} e_t,$$

where $R'_{(i)}(\delta_t)$ denotes the i-th row of $R'(\delta_t)$. To compute $\partial R'_{(i)}(\delta_t)/\partial \delta_t$, note that $R'(\delta_t) = \prod_{j=1}^{N^*} R'_{N^*-j+1}(\delta_{N^*-j+1,t})$, and hence

$$\frac{\partial \mathrm{vec}(R')}{\partial \delta_{kt}} = \begin{cases} \left(I_N \otimes (\prod_{j=2}^{N^*} R'_{N^*-j+2})\right) \frac{\partial \mathrm{vec}(R'_k)}{\partial \delta_{kt}}, & k = 1 \\ \left((\prod_{j=1}^{k-1} R_j) \otimes (\prod_{j=k+1}^{N^*} R'_{N^*-j+k+1})\right) \frac{\partial \mathrm{vec}(R'_k)}{\partial \delta_{kt}}, & k = 2, \ldots, N^* - 1 \\ \left((\prod_{j=1}^{N^*-1} R_j) \otimes I_N\right) \frac{\partial \mathrm{vec}(R'_k)}{\partial \delta_{kt}}, & k = N^* \end{cases}$$

and

$$\frac{\partial R'_k}{\partial \delta_{kt}} = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & -\sin(\delta_{kt}) & \cdots & \cos(\delta_{kt}) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\cos(\delta_{kt}) & \cdots & -\sin(\delta_{kt}) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}.$$

Thus, the score is now an N^*-vector given by

$$u_t := \frac{\partial l_t}{\partial \delta_t} = \sum_{i=1}^{N} \frac{(\nu_i + 1)\xi_{it}(\delta_t)}{\nu_i - 2 + \xi_{it}(\delta_t)^2} \frac{\partial \xi_{it}}{\partial \delta_t} \tag{5.1}$$

Furthermore,

$$\frac{\partial u_t}{\partial \theta'} = \frac{\partial u_t}{\partial \delta_t'} \frac{\partial \delta_t}{\partial \theta'},$$

where (dependence on δ_t is suppressed for simplicity)

$$\frac{\partial u_t}{\partial \delta_t'} = \sum_{i=1}^{N} \frac{\nu_i + 1}{(\nu_i - 2 + \xi_{it}^2)^2} \left\{ (\nu_i - 2) \frac{\partial \xi_{it}}{\partial \delta_t} \frac{\partial \xi_{it}}{\partial \delta_t'} + \xi_{it}(\nu_i - 2 + \xi_{it}^2) \frac{\partial^2 \xi_{it}}{\partial \delta_t \partial \delta_t'} \right\}. \tag{5.2}$$

We decompose the parameter vector as $\theta = (\theta_1', \theta_2', \ldots, \theta_{N^*}')'$, letting $\theta_i = (\omega_i, \beta_i, \kappa_i)'$. Due to the separability we obtain

$$\frac{\partial \delta_{it}}{\partial \theta_j} = \begin{cases} A_{(i)} z_{i,t-1} + x_{i,t-1} \frac{\partial \delta_{i,t-1}}{\partial \theta_i}, & i = j \\ 0, & i \neq j \end{cases} \tag{5.3}$$

with the notation

$$A_{(i)} = \begin{pmatrix} 1 - \beta_i & 0 & 0 \\ -\omega_i & 1 & 0 \\ 0 & 0 & 1, \end{pmatrix}, \quad z_{it} = \begin{pmatrix} 1 \\ \delta_{it} \\ u_{it} \end{pmatrix}, \quad \text{and } x_{it} = \beta_i + \kappa_i \frac{\partial u_{it}}{\partial \delta_{it}}. \tag{5.4}$$

Finally, the score with respect to the parameter vector is given by

$$\frac{\partial l_t}{\partial \theta} = \frac{\partial l_t}{\partial \delta_t} \frac{\partial \delta_t}{\partial \theta}, \ a.s.,$$

where $\frac{\partial l_t}{\partial \delta_t}$ is given in (5.1) and $\frac{\partial \delta_t}{\partial \theta}$ has a blockdiagonal structure given by (5.3).

6 Conclusions and outlook

We have introduced a dynamic independent component model whose dynamics are driven by the conditional score of the pseudo likelihood with respect to the rotation angle of model innovations. The pseudo maximum likelihood (PML) estimator is shown to be consistent and asymptotically normally distributed, which is confirmed in an extensive simulation study that also documents its good finite sample properties (see online Appendix D).

We employ the dynamic rotation model to assess the kurtosis of an equal weight portfolio consisting of two FX rates (i.e. the price of the USD in Euro and British Pound Sterling) in a time-varying manner. Unlike kurtosis assessment by means of static rotations (or, ad-hoc covariance decompositions), the dynamic model issues kurtosis estimates that are highly sensitive to important historical events recorded at international financial markets, i.e. (i) the global financial crisis of 2008, (ii) the European sovereign debt crisis (2010-2013) and (iii) the onset of the UK to leave the European Union (2017). This is consistent with the model of Engle and Martins (2020) which links portfolio kurtosis to a geopolitical risk factor. Our analysis shows that it is important to have a well-specified model to extract independent components in a dynamic framework, as otherwise this link with geopolitical risk would be lost. Portfolio selection strategies that combine, for example, variance and kurtosis minimization, therefore have to rely on a dynamic model that allows to estimate both conditional variance and kurtosis in a consistent way, which is what our model does.

In our application we have concentrated on portfolio selection, but there are many other potential applications, for example in risk management. Risk measures such as Value-at-Risk and expected shortfall can be estimated using higher order expansions such as Edgeworth or Cornish-Fisher of corresponding Gaussian quantities. This has been recognized by the European supervisory authorities EBA, EIOPA and ESMA, which have established regulatory technical standards for packaged retail and insurance-based investment products (see ESA 2016) that include the use of skewness and kurtosis via Cornish-Fisher expansions to evaluate market risk measures.

Another potential application is the construction of factor risk parity portfolios as proposed e.g. by Meucci (2009) and Roncalli and Weisang (2016), where the asset contributions to factor risks are diversified. Uncorrelated factors are typically obtained from principal components, but Lassance et al. (2019) emphasize the deficiencies of this approach and instead propose, in a static framework, to use factors based on independent components. As a generalization of this approach, our model could be applied to construct dynamic independent component risk parity portfolios which, given the results of Lassance et al. (2019), are very promising in terms of Sharpe ratio performance.

As confirmed by our simulation study, PML estimation delivers consistent but inefficient estimates in the case where the true distribution does not correspond to the pseudo distribution used for estimation. In our case, the difference occurred in the degrees-of-freedom parameters of the Student-t distributions, related to the thickness of the tails of the corresponding distributions. We have not analyzed other possible scenarios such as a skewness different from zero, or multi-modal distributions. It might be, for example, that the efficiency loss using PML assuming a symmetric distribution is larger when the true distribution is skewed, in which case other potential estimators might be considered, which merits further research.

Our empirical illustration is throughout restricted to the bivariate case and pursued in an ex-post fashion. While we present a generalization of the model to higher dimensions, the implementation might face algorithmic and computational challenges. Addressing these challenges promises valuable insights for the practical and (pseudo) ex-ante conduct of portfolio allocation and risk management. A large-scale out-of-sample analysis in high dimensions was beyond the scope of this paper but is certainly of utmost interest for financial applications, which we also leave for future research.

Acknowledgements

Much of the research documented in this paper has been done while the second author was visiting the Department of Economics and the Center of Environmental and Natural Resource Economics (CENRE) of the University of Chile, Santiago de Chile. The hospitality and productive research environment of both institutions are cordially acknowledged. The first author is grateful for financial support of the research grant ARC 18-23/089 of the Belgian Federal Science Policy Office. The second author gratefully acknowledges financial support by the Deutsche Forschungsgemeinschaft (HE2188/14-1). We cordially acknowledge helpful comments received from an anonymous reviewer.

Appendices

A Figures

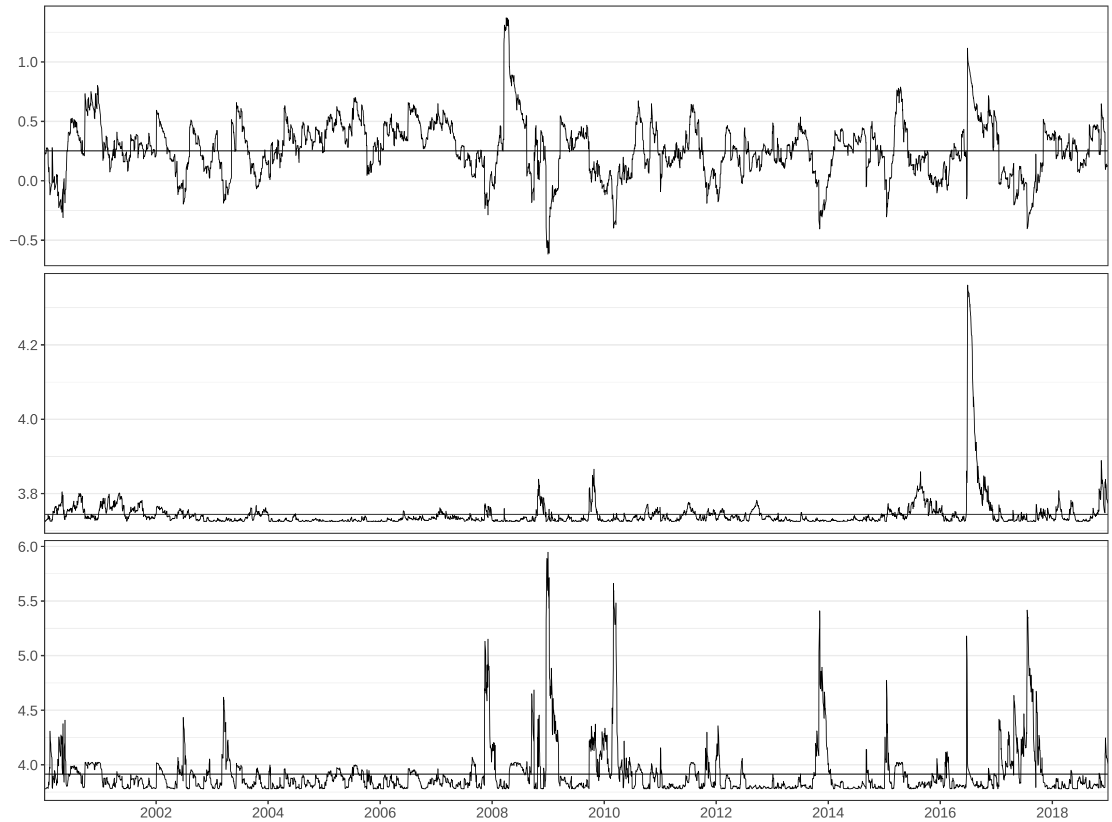

Figure A.1: Time varying rotation angles ($\hat{\delta}_t$, upper panel), model implied kurtosis estimates for equal weight portfolio returns (panel 2: static rotation model; panel 3: dynamic rotation model). The horizontal lines indicate sample averages.

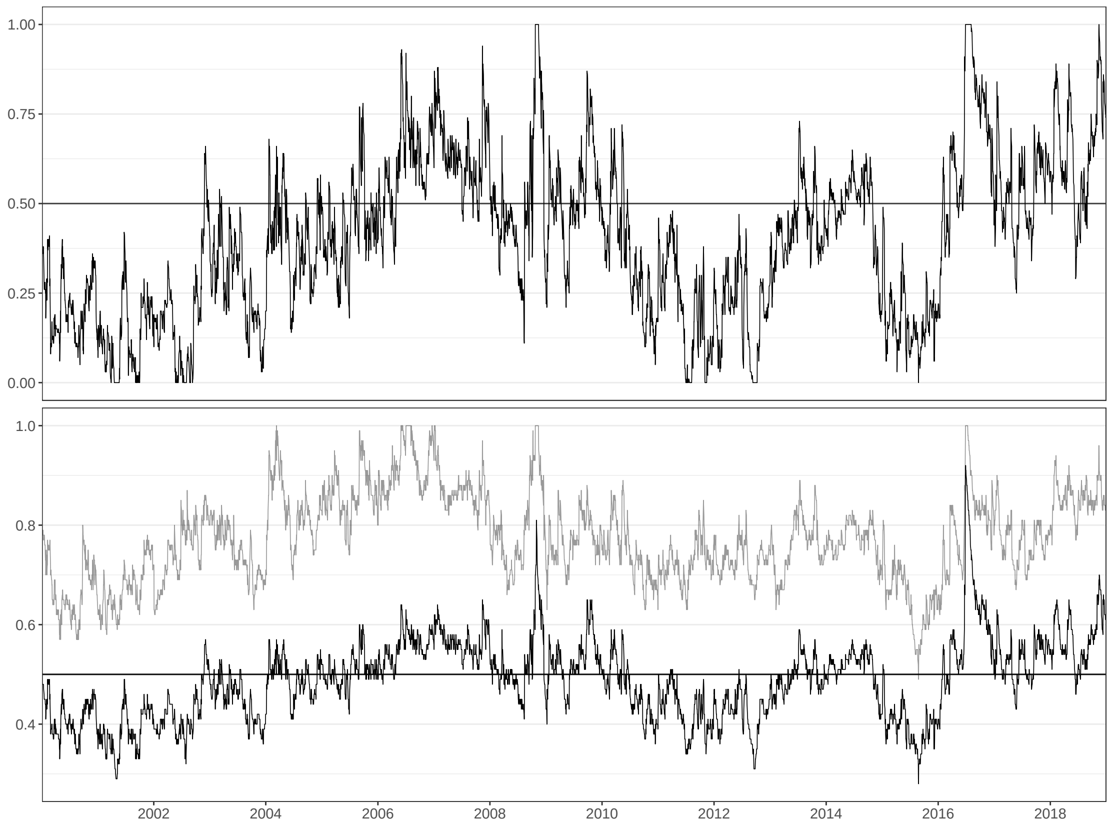

Figure A.2: Time-varying portfolio weights (w_{1t}) as implied by the minimum variance portfolio (upper panel) and minimum kurtosis portfolios (lower panel). Kurtosis minimizing portfolio weights are shown for static (black) and the dynamic rotation model (grey). The horizontal lines indicate equal weight portfolio compositions ($w_{1t} = 0.5$).

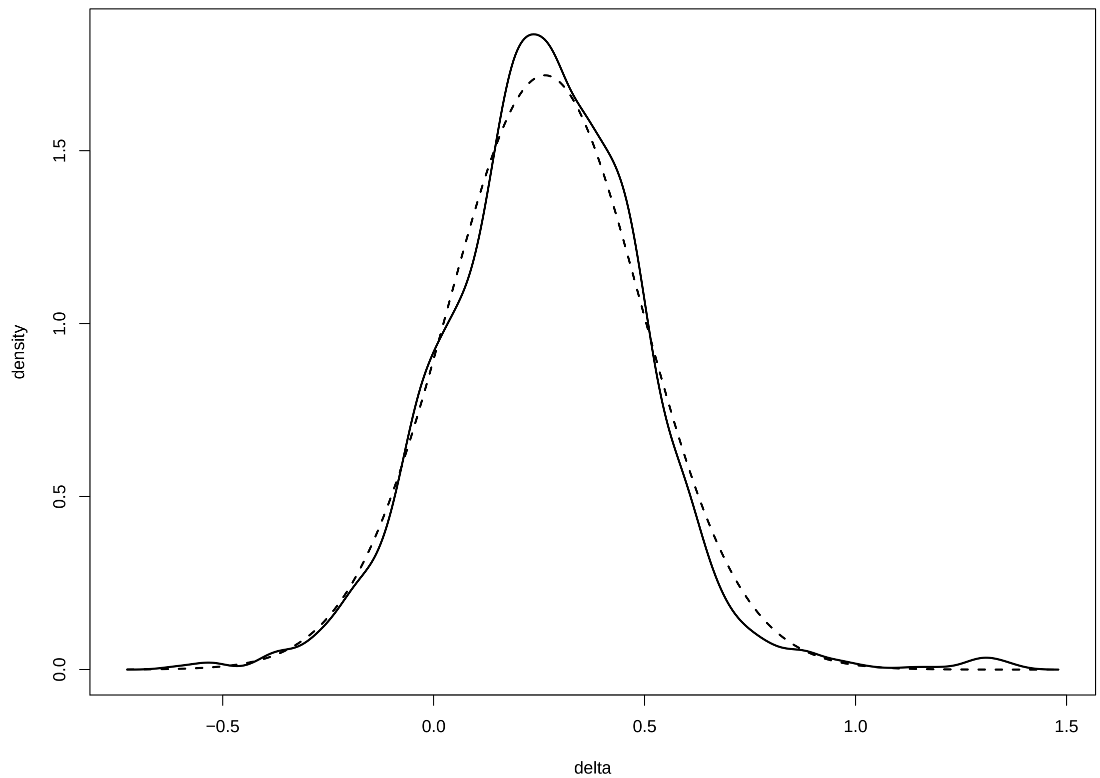

Figure A.3: Estimate of the density of the stationary distribution of δ_t. Solid line: nonparametric kernel estimate with Silverman's rule of thumb, dashed: density of a von Mises distribution with mean $\hat{\omega} = 0.26$ and scale parameter $\hat{\upsilon} = 18.8$ estimated by maximizing the concentrated likelihood in (4.2).

B Tables

	ω	β	κ
PML estimator	0.252	0.980	0.101
SE based on $\widehat{I}^{-1}$	0.0408	0.0059	0.0146
SE based on $\widehat{V}$	0.0401	0.0050	0.0155
bootstrap mean	0.252	0.979	0.101
bootstrap SE	0.046	0.007	0.0162

Table B.1: PML estimation results, where $\widehat{I}^{-1}$ is the inverse of an estimator of the information matrix I, and $\widehat{V}$ is an estimator of the asymptotic covariance matrix $J^{-1}IJ^{-1}$.

	mean	std	min	Q25	Q50	Q75	max	mean	std	min	Q25	Q50	Q75	max
model	$w_1 = 0.10$							$w_1 = 0.25$						
δ	3.74	0.046	3.73	3.73	3.73	3.75	4.36	3.84	0.047	3.73	3.80	3.84	3.87	3.95
δ_t	3.91	0.229	3.78	3.79	3.84	3.94	5.95	3.91	0.109	3.78	3.82	3.92	3.99	5.18
	$w_1 = 0.50$							$w_1 = 0.75$						
δ	4.22	0.208	3.81	4.07	4.18	4.33	5.03	3.92	0.146	3.73	3.83	3.89	3.98	4.79
δ_t	4.44	0.629	3.78	3.93	4.23	4.80	6.29	4.16	0.488	3.78	3.83	3.94	4.33	6.29

Table B.2: Descriptive statistics of model implied kurtosis estimates K_t for alternative portfolio weights w_1 and ($w_2 = 1 - w_1$). Row-wise: Constant rotation model ($\delta_t = \delta$); dynamic rotation model ($\delta_t = \delta$). The respective approximations of Ψ (see (4.6)) comprise $m_1^{(4)} = 3.96$, $m_2^{(4)} = 5.99$ and $m_1^{(4)} = 4.02$, $m_2^{(4)} = 6.29$ for the static and dynamic rotation model, respectively. Documented statistics include: average kurtosis (mean) and standard deviation (std) and minimum, 25% (Q25), 50% (Q50, median), 75% (Q75) quantiles, and minimum (min) and maximum (max).

w_1	sd	$m^{(3)}$	$m^{(4)}$	skew	kurt
0.10	5.72E-3	8.84E-8	1.39E-8	0.473	13.0
0.25	5.52E-3	4.74E-8	1.00E-8	0.282	10.8
0.50	5.42E-3	1.57E-9	6.66E-9	0.010	7.70
0.75	5.62E-3	-2.21E-8	6.10E-9	-0.124	6.11

Table B.3: Descriptive statistics for portfolio returns with alternative weights $w_1 = 0.10, 0.25, 0.50$ and $w_1 = 0.75$. Owing to data centering mean portfolio returns are zero throughout.

References

Ackermann, F., W. Pohl, and K. Schmedders (2017). "Optimal and naive diversification in currency markets". *Management Science* 63.10, pp. 3347–3360.

Barkhagen, M., B. Fleming, S. G. Quiles, J. Gondzio, J. Kalcsics, J. Kroeske, S. Sabanis, and A. Staal (2019). *Optimising portfolio diversification and dimensionality.*

Baumeister, C. and G. Peersman (2013). "Time-varying effects of oil supply shocks on the US economy". *American Economic Journal: Macroeconomics* 5.4, pp. 1–28.

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

Chan, L. K., J. Karceski, and J. Lakonishok (1999). "On portfolio optimization: Forecasting covariances and choosing the risk model". *The Review of Financial Studies* 12.5, pp. 937–974.

Comon, P. (1994). "Independent component analysis, A new concept?" *Signal Processing* 36.3, pp. 287–314.

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.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *The Review of Financial Studies* 22.5, pp. 1915–1953.

Engle, R. F. and K. F. Kroner (1995). "Multivariate simultaneous generalized ARCH". *Econometric Theory* 11, pp. 122–150.

Engle, R. F. and S. Martins (2020). *Measuring and hedging geopolitical risk.* Tech. rep. NYU Stern School of Business.

ESA (2016). *Final draft regulatory technical standards.* Working Paper. Joint Committee of the European Supervisory Authorities ESMA, EBA, and EIOPA.

Gouriéroux, C, A Monfort, and J. Renne (2017). "Statistical inference for independent component analysis: Application to structural VAR models". *Journal of Econometrics* 196, pp. 111–126.

Hafner, C. M. and H. Herwartz (2006). "Volatility impulse responses for multivariate GARCH models: An exchange rate illustration". *Journal of International Money and Finance* 25, pp. 719–740.

Hafner, C. M., H. Herwartz, and S. Maxand (2020). "Identification of structural multivariate GARCH models". *Journal of Econometrics.*

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Vol. 52. Cambridge University Press.

Hu, Y.-P. and R. Tsay (2014). "Principal component volatility analysis". *Journal of Business & Economic Statistics* 32.2, pp. 153–164.

Jagannathan, R. and T. Ma (2003). "Risk reduction in large portfolios: Why imposing the wrong constraints helps". *The Journal of Finance* 58.4, pp. 1651–1683.

Kilian, L. and H. Lütkepohl (2017). *Structural Vector Autoregressive Analysis*. Cambridge University Press.

Lancaster, H. O. (1954). "Traces and cumulants of quadratic forms in normal variables". *Journal of the Royal Statistical Society. Series B (Methodological)* 16.2, pp. 247–254.

Lanne, M., M. Meitz, and P. Saikkonen (2017). "Identification and estimation of non-Gaussian structural vector autoregressions". *Journal of Econometrics* 196, pp. 288–304.

Lassance, N., V. DeMiguel, and F. D. Vrins (2019). "Optimal portfolio diversification via independent component analysis". *Available at SSRN 3285156*.

Markowitz, H. (1952). "Portfolio Selection". *The Journal of Finance* 7.1, pp. 77–91.

Matteson, D. and R. Tsay (2011). "Dynamic orthogonal components for multivariate time series". *Journal of the American Statistical Association* 106.496, pp. 1450–1463.

Meucci, A. (2009). "Managing diversification". *Risk* 22.5, pp. 74–79.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013). "Causal inference by independent component analysis: Theory and applications". *Oxford Bulletin of Economics and Statistics* 75 (5), pp. 705–730.

Pham dinh, T. and Y.-s. Niu (Dec. 2011). "An efficient DC programming approach for portfolio decision with higher moments". *Computational Optimization and Applications* 50.3, pp. 525–554.

Roncalli, T. and G. Weisang (2016). "Risk parity portfolios with risk factors". *Quantitative Finance* 16.3, pp. 377–388.

van der Weide, R. (2002). "GO-GARCH: A multivariate generalized orthogonal GARCH model". *Journal of Applied Econometrics* 17, pp. 549–564.

Weber, E. (2010). "Structural conditional correlation". *Journal of Financial Econometrics* 8.3, pp. 392–407.

Dynamic score driven independent component analysis
Supplementary Appendix

Christian M. Hafner* Helmut Herwartz†

February 24, 2022

Appendices

C Proofs of the theorems

Let $\tilde{\delta}_t$ denote the process (2.3) where the starting value $\tilde{\delta}_0$ is drawn from the stationary distribution of $\tilde{\delta}_t$, and let $\tilde{L}(\theta)$ and $\tilde{l}_t(\theta)$ be defined accordingly. In practice these are not feasible and a fixed starting value δ_0 has to be used. One of the challenges is to show that this does not affect the asymptotic properties of the estimator. The true parameter is denoted θ_0, and we will use the notation $\delta_t = \delta_t(\theta)$ and $\delta_{0t} = \delta_t(\theta_0)$. Furthermore, let $C > 0$ denote a generic absolute constant that may take different values in different equations.

Lemma 1. *Under Assumptions (A1)-(A3), the model (2.1) is identifiable up to a permutation of the columns of $R(\delta_t)$. Under Assumptions (A1)-(A3) and (A5), it is*

*Corresponding author, Louvain Institute of Data Analysis and Modeling in Economics and Statistics (LIDAM), and ISBA, Université catholique de Louvain, Belgium. e-mail christian.hafner@uclouvain.be

†Department of Economics, University of Göttingen, Germany. e-mail hherwartz@uni-goettingen.de.

uniquely identifiable.

Proof: First, from (2.4) and (2.5), $\xi_{1t}(\theta)$ and $\xi_{2t}(\theta)$ are linear combinations of the true innovations, ξ_{1t} and ξ_{2t} with, by construction, identical second moment structure. The distributions of $\xi_t(\theta)$ and ξ_t would be identical under Gaussianity, which is excluded by Assumption (A3). As shown by Lancaster (1954), under the requirement of independence and "non-trivial" rotations, the only situation where non-identifiability occurs is the case of a normal distribution. As for 'trivial' rotations: Rotations by 180°, which correspond to the sign changes $\xi_t(\delta_{0t}-\pi) = -\xi_t$, are excluded by the restriction $|\delta_t - \delta_{0t}| < \pi$. Rotations by 90° correspond to a change in the order of the columns of $R(\delta_t)$: $\xi_{1t}(\delta_{0t} - \pi/2) = \xi_{2t}$ and $\xi_{2t}(\delta_{0t} - \pi/2) = \xi_{1t}$. This is not excluded by our restrictions and so $R(\delta_t)$ is identified only up to a change of columns under Assumptions (A1)-(A3). Adding Assumption (A5), the distributions of ξ_{1t} and ξ_{2t} are different, so that $R(\delta_t)$ is uniquely identified.

It remains to show that $\delta_t(\theta) = \delta_t(\theta_0)$ a.s. for all t implies that $\theta = \theta_0$. Let $\delta_t(\theta) = \omega + \psi(L)u_t$ and suppose that $\delta_t(\theta) = \delta_t(\theta_0)$ a.s. for all t. Then $\delta_t(\theta) - \delta_t(\theta_0) = 0 \Rightarrow \omega - \omega_0 = \left[\psi(L) - \psi_0(L)\right] u_t$ a.s. for all t. If $\psi(L) \neq \psi_0(L)$, the right hand side consists of some linear combination of u_{t-j}, $j > 0$, which has a nondegenerate distribution by Assumption (A2) and, hence, is not constant. Therefore, $\psi(L) = \psi_0(L)$ and $\omega = \omega_0$. $\square$

Proof of Theorem 1: We show strong consistency by contradiction. Suppose that $\hat{\theta}_T \not\to \theta_0$ a.s., so that for some arbitrary $\eta > 0$ the set $\Omega = \{\omega : \lim_{T\to\infty} \|\hat{\theta}_T - \theta_0\| \geq \eta\}$ has a positive probability. Since the set $\Lambda = \Theta \cap \{\theta : \|\theta - \theta_0\| \geq \eta\}$ is compact, we can find for each $\omega \in \Omega$ a convergent subsequence $\hat{\theta}_{T_i} \to \tilde{\theta} \in \Lambda$. From the definition of the QMLE,

$$\limsup_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\theta_0) \le \limsup_{i\to\infty}\sup_{\theta\in\Lambda}\frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\theta) = \limsup_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i} l_t(\hat\theta_{T_i}).$$

Lemma 3 implies that this also holds for the functions $\tilde{l}_t(\cdot)$, i.e.

$$\limsup_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\theta_0) \le \limsup_{i\to\infty}\sup_{\theta\in\Lambda}\frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\theta) = \limsup_{i\to\infty}\frac{1}{T_i}\sum_{t=1}^{T_i} \tilde{l}_t(\hat\theta_{T_i}).$$

From Lemmas 2(i) and 2(iii), $\mathrm{E}\tilde{l}_t(\theta_0) \le \mathrm{E}\sup_{\theta\in\Lambda}\tilde{l}_t(\theta)$, which contradicts Lemma 2(ii). Since $\eta > 0$ is arbitrary, the strong consistency follows. $\square$

Lemma 2: Under Assumptions (A1)-(A7),

i) $\frac{1}{T}\sum\limits_{t=1}^{T}\tilde{l}_t(\theta_0) \to \mathrm{E}\tilde{l}_t(\theta_0),\ a.s.$

ii) $\mathrm{E}\tilde{l}_t(\theta_0) \ge \mathrm{E}\tilde{l}_t(\theta)$ with equality if and only if $\theta = \theta_0$.

iii) For any compact set $\Lambda \subseteq \Theta$, $\limsup_{T\to\infty}\sup_{\theta\in\Lambda}\frac{1}{T}\sum\limits_{t=1}^{T}\tilde{l}_t(\theta) \le \mathrm{E}\sup_{\theta\in\Lambda}\tilde{l}_t(\theta)$ a.s.

Proof:

i) Since $\mathrm{E}\tilde{l}_t(\theta_0) = \mathrm{E}\sum_i \log g_i(\xi_{it})$ and $\mathrm{E}|\log g_i(\xi_{it})|$ is bounded by Assumption (A4), the desired result follows from the ergodic theorem.

ii) Note that from (2.7), $\xi_{1t}^2(\delta_t) = a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}$ and $\xi_{2t}^2(\delta_t) = (1-a_t)\xi_{1t}^2 + a_t\xi_{2t}^2 - b_t\xi_{1t}\xi_{2t}$, where $a_t = \cos^2(\Delta_t)$ and $b_t = \sin(2\Delta_t)$. Hence,

$$\begin{aligned}
\mathrm{E}\tilde{l}_t(\theta) - \mathrm{E}\tilde{l}_t(\theta_0) &= \frac{\nu_1+1}{2}\mathrm{E}\log(\nu_1 - 2 + \xi_{1t}^2) + \frac{\nu_2+1}{2}\mathrm{E}\log(\nu_2 - 2 + \xi_{2t}^2) \\
&- \frac{\nu_1+1}{2}\mathrm{E}\log(\nu_1 - 2 + a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}) \\
&- \frac{\nu_2+1}{2}\mathrm{E}\log(\nu_2 - 2 + (1-a_t)\xi_{1t}^2 + a_t\xi_{2t}^2 - b_t\xi_{1t}\xi_{2t})
\end{aligned}$$

It suffices to show that

$$\mathrm{E}\log(\nu_1 - 2 + \xi_{1t}^2) - \mathrm{E}\log(\nu_1 - 2 + a_t\xi_{it}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}) \leq 0. \quad \text{(C.1)}$$

Conditioning on $\mathcal{F}_{t-1}$ and denoting $\mathrm{E}_t(\cdot) = \mathrm{E}[\cdot|\mathcal{F}_{t-1}]$, we have due to the law of iterated expectation

$$\mathrm{E}[\log(\nu_1 - 2 + \xi_{1t}^2) - \mathrm{E}_t\log(\nu_1 - 2 + a_t\xi_{it}^2 + (1-a_t)\xi_{2t}^2) + b_t\xi_{1t}\xi_{2t}] \leq 0,$$

where $a_t \in [0,1)$ and $b_t \in (-1,1)$ are $\mathcal{F}_{t-1}$-measurable. Therefore, Assumption (A7) applies and (C.1) holds.

iii) For any compact subset $\Lambda \subseteq \Theta$ we have that

$$\limsup_{T\to\infty} \sup_{\bar\theta\in\Lambda} \frac{1}{T}\sum_{t=1}^{T} l_t(\bar\theta) \leq \limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sup_{\bar\theta\in\Lambda} l_t(\bar\theta).$$

Now from Lemma 2(i) we have that $\mathrm{E}\sup_{\bar\theta\in\Lambda}\tilde{l}_t(\bar\theta) \in \mathbb{R}\cup\{+\infty\}$ and since $\{\sup_{\bar\theta\in\Lambda} l_t(\bar\theta)\}$ is a stationary and ergodic process we have that (see e.g. Pfanzagl (1969))

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \sup_{\bar\theta\in N} l_t(\bar\theta) \leq \mathrm{E}\sup_{\bar\theta\in N} l_t(\bar\theta),$$

and the desired result follows. □

Lemma 3: Under Assumptions (A1)-(A11) we have that

$$\limsup_{T\to\infty} \sup_{\theta\in\Theta} \left| \tfrac{1}{T}\sum_{t=1}^{T} l_t(\theta) - \tfrac{1}{T}\sum_{t=1}^{T} \tilde{l}_t(\theta) \right| = 0 \;\; a.s.$$

Proof: By Cesaro's mean theorem it is sufficient to check that $\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|$ tends to zero a.s. If $\mathrm{E}\sup_{\theta\in\Theta}|l_t(\theta) - \tilde{l}_t(\theta)|^r$ is bounded by a summable sequence in t

for some $r>0$, by the generalized Chebyshev inequality, $\sum_{t=1}^{\infty} \mathrm{P} \sup_{\theta \in \Theta} |l_t(\theta) - \tilde{l}_t(\theta)| > \eta < \infty$, for all $\eta > 0$. Then the desired result follows by the Borel-Cantelli lemma.

Let $a_t = \cos^2(\delta_t - \delta_{0t})$, $\tilde{a}_t = \cos^2(\tilde{\delta}_t - \tilde{\delta}_{0t})$, $b_t = \sin(2\Delta_t)$ and $\tilde{b}_t = \sin(2\tilde{\Delta}_t)$. Furthermore, let $\tilde{\xi}_{it} = \xi_{it}(\tilde{\delta}_{0t})$ and note that $\xi_t - \tilde{\xi}_t = (R'(\delta_{0t})R(\tilde{\delta}_{0t}) - I_2)\tilde{\xi}_t$, where $R'(\delta_{0t})R(\tilde{\delta}_{0t})$ is a matrix of bounded and continuously differentiable functions, so that there exists a constant $C>0$ such that

$$\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}| \|\tilde{\xi}_t\|. \tag{C.2}$$

Then,

$$\begin{aligned} l_t(\theta) - \tilde{l}_t(\theta) &= \frac{\nu_1+1}{2} \log(\nu_1 - 2 + \tilde{a}_t \tilde{\xi}_{1t}^2 + (1-\tilde{a}_t)\tilde{\xi}_{2t}^2) + b_t \tilde{\xi}_{1t}\tilde{\xi}_{2t} \\ &+ \frac{\nu_2+1}{2} \log(\nu_2 - 2 + (1-\tilde{a}_t) \tilde{\xi}_{1t}^2 + \tilde{a}_t\tilde{\xi}_{2t}^2) - b_t \tilde{\xi}_{1t}\tilde{\xi}_{2t} \\ &- \frac{\nu_1+1}{2} \log(\nu_1 - 2 + a_t \xi_{1t}^2 + (1-a_t)\xi_{2t}^2) + b_t \xi_{1t}\xi_{2t} \\ &- \frac{\nu_2+1}{2} \log(\nu_2 - 2 + (1-a_t) \xi_{1t}^2 + a_t\xi_{2t}^2) - b_t \xi_{1t}\xi_{2t} \end{aligned}$$

It suffices to show that, for some $r>0$ and $\rho \in (0,1)$,

$$\mathrm{E} \sup_{\theta \in \Theta} \left| \log(h(\delta_t)) - \log(h(\tilde{\delta}_t)) \right|^r = O(\rho^t) \tag{C.3}$$

where $h(\delta_t) := \nu_1 - 2 + a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t}$. Note that $a_t\xi_{1t}^2 + (1-a_t)\xi_{2t}^2 + b_t\xi_{1t}\xi_{2t} \geq 0$ a.s., and that by Assumption (A6), $h(\delta_t)$ is bounded below by an absolute constant $\epsilon > 0$. Moreover, a_t and b_t are bounded, continuously differentiable functions of $\Delta_t = \delta_t - \delta_{0t}$. Hence, $\log(h(\delta_t))$ is a Lipschitz-continuous function, since it has a bounded first derivative. An application of the mean value theorem implies that there exists a constant $C>0$ such that

$$\left| \log(h(\delta_t)) - \log(h(\tilde{\delta}_t)) \right| \leq C \left| h(\delta_t) - h(\tilde{\delta}_t) \right| \tag{C.4}$$

Because a_t and b_t are bounded functions, we further have

$$\left|h(\delta_t) - h(\tilde{\delta}_t)\right| \leq C\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}|\|\tilde{\xi}_t\| \tag{C.5}$$

using (C.2). Therefore,

$$\sup_{\theta\in\Theta}\left|\log(h(\delta_t)) - \log(h(\tilde{\delta}_t))\right|^r \leq C\|\tilde{\xi}_t\|^r|\delta_{0t} - \tilde{\delta}_{0t}|^r \tag{C.6}$$

For $0 < r < 1$, $\mathrm{E}[\|\tilde{\xi}_t\|^r] \leq 2^r$ by Assumption (A1) and Jensen's inequality.

To establish (C.3) it remains to show that $\mathrm{E}|\delta_{0t} - \tilde{\delta}_{0t}|^r = O(\rho^t)$. By the definition of δ_t in (2.3), we have

$$\delta_t - \tilde{\delta}_t = \beta(\delta_{t-1} - \tilde{\delta}_{t-1}) + \kappa(u_{t-1} - \tilde{u}_{t-1}) \tag{C.7}$$

By the mean value theorem, $u_t - \tilde{u}_t = \frac{\partial u_t}{\partial \delta_t}(\bar{\delta}_t)(\delta_t - \tilde{\delta}_t)$ where $\bar{\delta}_t$ is on the line segment joining δ_t and $\tilde{\delta}_t$, so that

$$\delta_t - \tilde{\delta}_t = \prod_{i=1}^{t-1} x_{t-i}(\bar{\delta}_{t-i})(\delta_0 - \tilde{\delta}_0),$$

where $x_t = \beta + \kappa\partial u_t/\partial\delta_t(\bar{\delta}_t)$. Note that $|x_t|$ is a stationary ergodic sequence of non-negative random variables with $\mathrm{E}\log|x_t| < 0$, which is implied by Assumption (A11). Thus, by Lemma 2.4 of Straumann and Mikosch (2006), $\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| \to_{e.a.s.} 0$, meaning that there exists a $\gamma > 1$ s.t. $\gamma^t\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| \to_{a.s.} 0$. This implies that $\mathrm{E}\prod_{i=1}^{t-1}|x_{t-i}(\bar{\delta}_{t-i})| = O(\rho^t)$, and

$$\mathrm{E}\left|\delta_t - \tilde{\delta}_t\right|^r = \mathrm{E}\prod_{i=1}^{t-1}\left|x_{t-i}(\bar{\delta}_{t-i})\right|^r \mathrm{E}|\delta_0 - \tilde{\delta}_0|^r = O(\rho^t) \tag{C.8}$$

for some $r > 0$. Hence, (C.3) follows by (C.8) and (C.6). □

Proof of Theorem 2: Let $N(\theta_0)$ be an arbitrarily small compact set around θ_0. By strong consistency, we have that for sufficiently large T, $\hat{\theta}_T \in$ int $N(\theta_0)$ a.s., hence the mean-value expansion of the score vector around θ_0 gives

$$0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\hat{\theta}_T)}{\partial\theta} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\right) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} \tag{C.9}$$

$$+ \left[\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} + J\right) + \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\tilde{\theta}_T)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right) - J\right]\sqrt{T}(\hat{\theta}_T - \theta_0)$$

where $\tilde{\theta}_T$ is on the line segment between $\hat{\theta}_T$ and θ_0, and recall that $J = -\mathrm{E}\left(\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}\right)$.

The first term in (C.9) converges to zero in probability by Lemma 7. Next, we show that the second term, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}$, obeys a CLT. As $\partial\tilde{l}_t(\theta_0)/\partial\theta$ is the product of an i.i.d. term, u_t, and $\partial\tilde{\delta}_t(\theta_0)/\partial\theta$, it is strictly stationary and ergodic if $\partial\tilde{\delta}_t(\theta_0)/\partial\theta$ is strictly stationary and ergodic, for which Assumption (A11) is sufficient since $\mathrm{E}(x_t^2) < 1$ implies $\mathrm{E}\log|x_t| < 0$ by Jensen's inequality. We have that $\mathrm{E}\left[\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}|\mathcal{F}_{t-1}\right] = 0$, where $\mathcal{F}_t$ is the σ-field generated by the current and past values of e_t, i.e. $\mathcal{F}_t \equiv \sigma(e_t, e_{t-1}, \ldots)$. Thus, the score is also a martingale difference sequence. Lemma 4 shows that $\mathrm{E}\left\|\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}'\right\|$ is finite, so that we can apply the CLT of Scott (1973) and the Cramer-Wold device to establish the asymptotic normality of the score function.

Next, we show that the first term inside the square parenthesis in (C.9) converges a.s. to zero. The second derivative is given by

$$\frac{\partial^2 \tilde{l}_t}{\partial\theta\partial\theta'} = \frac{\partial^2 \delta_t}{\partial\theta\partial\theta'}u_t + \frac{\partial u_t}{\partial\delta_t}\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}.$$

The score u_t and its derivative, $\frac{\partial u_t}{\partial\delta_t}$, are independent of δ_t and its derivatives. Furthermore, $\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}$ is stationary and ergodic by Corollary 6 of Harvey (2013), and $\frac{\partial^2\delta_t}{\partial\theta\partial\theta'}$ is stationary and ergodic by Corollary 7 of Harvey (2013). This implies by the ergodic theorem that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \rightarrow_{a.s.} \mathrm{E}(u_t)\mathrm{E}\frac{\partial^2\delta_t}{\partial\theta\partial\theta'} + \mathrm{E}\frac{\partial u_t}{\partial\delta_t}\mathrm{E}\frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'} \tag{C.10}$$

The expectation $\mathrm{E}\frac{\partial^2 \delta_t}{\partial\theta\partial\theta'}$ is finite if $\mathrm{E}[\frac{\partial^2 u_t}{\partial \delta_t^2}] < \infty$, which holds by Lemma 6. Hence, the first term in (C.10) is zero because $\mathrm{E}(u_t) = 0$. The second term is equal to $-J$, which shows that the first term inside the square parenthesis in (C.9) converges a.s. to zero. In addition, it follows by standard arguments that $-\mathrm{E}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'}$ is positive definite.

Lemmas 8 and 9 imply that the second term inside the square parenthesis in (C.9) is $o_{a.s.}(1)$. Since we have shown that the second term in (C.9) obeys a CLT, asymptotic normality follows from Slutsky's theorem. Finally, following Theorem 1 and Appendix A of Harvey (2013) to calculate the expectation of the outer product of $\partial\delta_t/\partial\theta$, we obtain the expression for the asymptotic covariance matrix in (3.8). $\square$

Lemma 4: Under Assumptions (A1)-(A11), $\mathrm{E}\left\| \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta}' \right\| < \infty$.

Proof: Note that

$$\frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta'} = u_t^2 \frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta} \frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta'}$$

As u_t is i.i.d. and $\partial\tilde{\delta}_t/\partial\theta$ is $\mathcal{F}_{t-1}$-measurable, by the law of iterated expectations,

$$\mathrm{E}\left\| \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta'} \right\| = \mathrm{E}(u_t^2)\mathrm{E}\left\| \frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta} \frac{\partial \tilde{\delta}_t(\theta_0)}{\partial\theta'} \right\| \tag{C.11}$$

Let $\bar{\nu} = \max(\nu_1, \nu_2) < \infty$ and $\underline{\nu} = \min(\nu_1, \nu_2)$. By Assumption (A6), $\underline{\nu} > 2$. By the definition of u_t in (3.2),

$$u_t^2 \leq (\underline{\nu} - 2)^{-2}(\bar{\nu} + 1)^2 \xi_{1t}^2 \xi_{2t}^2, \quad a.s.$$

and hence, using Assumption (A1),

$$E(u_t^2) \leq (\underline{\nu} - 2)^{-2}(\bar{\nu} + 1)^2 < \infty.$$

Furthermore,

$$\mathrm{E}\left\|\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta}\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta'}\right\| \leq \mathrm{E}\left\|\frac{\partial\tilde{\delta}_t(\theta_0)}{\partial\theta}\right\|^2 < \infty$$

where the first inequality is an application of the Cauchy-Schwarz inequality, and the second follows by Lemma 10 of Harvey (2013) and Assumption (A11). □

Lemma 5: Under Assumption (A1), $\mathrm{E}\left|\frac{\partial u_t}{\partial\delta_t}\right| < \infty$.

Proof: Write $\frac{\partial u_t}{\partial\delta_t}$ in (3.3) as $\frac{\partial u_t}{\partial\delta_t} = (\nu_2+1)I_1 + (\nu_1+1)I_2$, where

$$I_1 = \frac{(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} \tag{C.12}$$

$$I_2 = \frac{(\nu_1-2)(\xi_{1t}^2-\xi_{2t}^2)+\xi_{1t}^4+\xi_{1t}^2\xi_{2t}^2}{\left\{\nu_1-2+\xi_{1t}(\delta_t)^2\right\}^2} \tag{C.13}$$

By the triangle inequality $\mathrm{E}\left|\frac{\partial u_t}{\partial\delta_t}\right| \leq (\nu_2+1)\mathrm{E}|I_1| + (\nu_1+1)\mathrm{E}|I_2|$. It suffices to consider the first term, I_1. By symmetry, the result then applies also to the second term. Note that

$$\begin{aligned}\frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} &\leq \frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} + \frac{\xi_{2t}^4}{\left\{\nu_2-2+\xi_{2t}(\delta_t)^2\right\}^2} \\ &\leq \frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{\left\{\nu_2-2\right\}^2} + 1 \ a.s. \qquad \text{(C.14)}\end{aligned}$$

and due to the finite second moments of ξ_t by Assumption (A1),

$$\begin{aligned}\mathrm{E}|I_1| &\leq \mathrm{E}\frac{\left|(\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right|}{(\nu_2-2)^2}+1 &\text{(C.15)}\\ &\leq \mathrm{E}\frac{\left|\xi_{2t}^2-\xi_{1t}^2\right|}{\nu_2-2}+\mathrm{E}\frac{\xi_{1t}^2\xi_{2t}^2}{(\nu_2-2)^2}+1 &\text{(C.16)}\\ &\leq \frac{2}{\nu_2-2}+\frac{1}{(\nu_2-2)^2}+1<\infty &\text{(C.17)}\end{aligned}$$

□

Lemma 6: Under Assumption (A9), $\mathrm{E}\left|\frac{\partial^2 u_t}{\partial \delta_t^2}\right| < \infty$.

Proof: We have

$$\frac{\partial^2 u_t}{\partial \delta_t^2} = (\nu_2+1)(T_{11}+T_{12}) + (\nu_1+1)(T_{21}+T_{22}),$$

where

$$\begin{aligned}
T_{11} &= \frac{-2\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_2-2)\right)}{\left(\nu_2-2+\xi_{2t}^2\right)^2} && \text{(C.18)}\\
T_{12} &= \frac{-4\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{2t}^4+\xi_{1t}^2\xi_{2t}^2\right)}{\left(\nu_2-2+\xi_{2t}^2\right)^3} && \text{(C.19)}\\
T_{21} &= \frac{2\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_1-2)\right)}{\left(\nu_1-2+\xi_{1t}^2\right)^2} && \text{(C.20)}\\
T_{22} &= \frac{4\xi_{1t}\xi_{2t}\left((\nu_1-2)(\xi_{1t}^2-\xi_{2t}^2)+\xi_{1t}^4+\xi_{1t}^2\xi_{2t}^2\right)}{\left(\nu_2-2+\xi_{1t}^2\right)^3} && \text{(C.21)}
\end{aligned}$$

By the c_r-inequality we have $\mathrm{E}\left|\frac{\partial^2 u_t}{\partial \delta_t^2}\right| \leq (\nu_2+1)(\mathrm{E}|T_{11}|+\mathrm{E}|T_{12}|)+(\nu_1+1)(\mathrm{E}|T_{21}|+\mathrm{E}|T_{22}|)$. It suffices to show that $\mathrm{E}|T_{11}|+\mathrm{E}|T_{12}|<\infty$. By symmetry, the result then also applies to $\mathrm{E}|T_{21}|+\mathrm{E}|T_{22}|$. Note that $|T_{11}| \leq |2(\nu_2-2)^{-1}\xi_{2t}\xi_{1t}\left(\xi_{1t}^2+\xi_{2t}^2+2(\nu_2-2)\right)$ a.s., and hence, as a direct consequence of Assumption (A9), $\mathrm{E}|T_{11}|<\infty$. The term $|T_{12}|$ can be bounded almost surely by

$$\begin{aligned}
|T_{12}| &\leq \frac{4|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)|}{\left(\nu_2-2+\xi_{2t}^2\right)^3} + \frac{4|\xi_{1t}\xi_{2t}^5|}{\left(\nu_2-2+\xi_{2t}^2\right)^3}\\
&\leq 4(\nu_2-2)^{-3}|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)| + \frac{4(\nu_2-2)^{-1}|\xi_{1t}\xi_{2t}^5|}{\left(\nu_2-2+\xi_{2t}^2\right)^2}\\
&\leq 4(\nu_2-2)^{-3}|\xi_{1t}\xi_{2t}\left((\nu_2-2)(\xi_{2t}^2-\xi_{1t}^2)+\xi_{1t}^2\xi_{2t}^2\right)| + 4(\nu_2-2)^{-1}|\xi_{1t}\xi_{2t}|,\ a.s.
\end{aligned}$$

and by Assumption (A9) both terms on the right hand side have finite expectation, and therefore $\mathrm{E}|T_{12}|<\infty$. $\square$

Lemma 7: Under Assumptions (A1)-(A11), $T^{-1/2}\sum_{t=1}^{T} \frac{\partial l_t(\theta_0)}{\partial\theta} - \frac{\partial \tilde{l}_t(\theta_0)}{\partial\theta} = o_{\mathrm{P}}(1)$.

Proof: By the generalized Chebyshev inequality and the c_r inequality we have for $\varepsilon > 0$ and $r > 0$,

$$\mathrm{P}\left(\left\|T^{-1/2}\sum_{t=1}^{T}\frac{\partial l_t(\theta_0)}{\partial\theta_i}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta_i}\right\|^r \geq \varepsilon\right) \leq T^{-r/2}\sum_{t=1}^{T}\mathrm{E}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\|^r.$$

Hence, it is sufficient to show that for some $r > 0$, there exists a $\rho \in (0, 1)$ such that,

$$\mathrm{E}\left\|\frac{\partial l_t(\theta_0)}{\partial\theta}-\frac{\partial\tilde{l}_t(\theta_0)}{\partial\theta}\right\|^r = O(t\rho^t). \tag{C.22}$$

Given the score function in (3.6), we have that (the argument θ_0 is omitted for simplicity),

$$\frac{\partial l_t}{\partial\theta}-\frac{\partial\tilde{l}_t}{\partial\theta} = (u_t - \tilde{u}_t)\frac{\partial\delta_t}{\partial\theta} + \tilde{u}_t(\frac{\partial\delta_t}{\partial\theta}-\frac{\partial\tilde{\delta}_t}{\partial\theta}),$$

and by the c_r inequality and the independence of u_t and $\frac{\partial\delta_t}{\partial\theta}$,

$$\mathrm{E}\left\|\frac{\partial l_t}{\partial\theta}-\frac{\partial\tilde{l}_t}{\partial\theta}\right\|^r \leq c_r\mathrm{E}|u_t-\tilde{u}_t|^r\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta}\|^r + c_r\mathrm{E}|\tilde{u}_t|^r\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta}-\frac{\partial\tilde{\delta}_t}{\partial\theta}\|^r$$

Now, $\mathrm{E}|\tilde{u}_t|^r < \infty$ by Assumption (A2), and $\mathrm{E}\|\frac{\partial\delta_t}{\partial\theta}\|^r < \infty$ for $0 < r \leq 2$ by Lemma 10 of Harvey (2013). To analyze $u_t - \tilde{u}_t$ it suffices to consider

$$\frac{\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_2 - 2 + \xi_{2t}(\delta_t)^2} - \frac{\xi_{1t}(\tilde{\delta}_t)\xi_{2t}(\tilde{\delta}_t)}{\nu_2 - 2 + \xi_{2t}(\tilde{\delta}_t)^2}.$$

The function $f(x, y) = xy/(\nu_2-2+y^2)$ is Lipschitz continuous as it is both Lipschitz continuous relative to x and to y. Therefore, there exists $C > 0$ such that

$$f(\xi_{1t}, \xi_{2t}) - f(\tilde{\xi}_{1t}, \tilde{\xi}_{2t}) \leq C\|\xi_t - \tilde{\xi}_t\|$$

and by (C.2), $\|\xi_t - \tilde{\xi}_t\| \leq C|\delta_{0t} - \tilde{\delta}_{0t}|\|\tilde{\xi}_t\|$. By symmetry, the same applies to the term

$$\frac{\xi_{1t}(\delta_t)\xi_{2t}(\delta_t)}{\nu_1 - 2 + \xi_{1t}(\delta_t)^2} - \frac{\xi_{1t}(\tilde{\delta}_t)\xi_{2t}(\tilde{\delta}_t)}{\nu_1 - 2 + \xi_{1t}(\tilde{\delta}_t)^2}.$$

Hence, $\mathrm{E}|u_t - \tilde{u}_t|^r \leq C\mathrm{E}|\delta_t - \tilde{\delta}_t|^r \mathrm{E}\|\tilde{\xi}_t\|^r = O(\rho^t)$, using (C.8) and $\mathrm{E}\|\tilde{\xi}_t\|^r < \infty$.

Furthermore, note that

$$\begin{aligned}\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta} &= A(z_{t-1} - \tilde{z}_{t-1}) + \frac{\partial \tilde{\delta}_{t-1}}{\partial \theta}(x_{t-1} - \tilde{x}_{t-1}) + x_{t-1}\left(\frac{\partial \delta_{t-1}}{\partial \theta} - \frac{\partial \tilde{\delta}_{t-1}}{\partial \theta}\right) \\ &= \sum_{j=1}^{t}\prod_{i=1}^{j-1} x_{t-i}\left\{A(z_{t-j} - \tilde{z}_{t-j}) + \frac{\partial \tilde{\delta}_{t-j}}{\partial \theta}(x_{t-j} - \tilde{x}_{t-j})\right\} + \prod_{i=1}^{t} x_{t-i}\frac{\partial \tilde{\delta}_0}{\partial \theta}\end{aligned}$$

with the convention $\prod_{i=1}^{0} = 1$. Now, $\mathrm{E}|z_t - \tilde{z}_t|^r = O(\rho^t)$ follows from the above results for $\mathrm{E}|u_t - \tilde{u}_t|^r$ and $\mathrm{E}|\delta_t - \tilde{\delta}_t|^r$. Then, $x_t - \tilde{x}_t = \kappa(\partial u_t/\partial \delta_t - \partial \tilde{u}_t/\partial \delta_t)$, and the fact that $\mathrm{E}|\partial u_t/\partial \delta_t - \partial \tilde{u}_t/\partial \delta_t|^r = O(\rho^t)$ follows by Lipschitz continuity of $\partial u_t/\partial \delta_t$ and the same arguments used to establish $\mathrm{E}|u_t - \tilde{u}_t|^r = O(\rho^t)$ above. Due to the independence of $\{x_t\}$, we obtain

$$\mathrm{E}|\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta}|^r \leq C\sum_{j=1}^{t}\prod_{i=1}^{j-1}\mathrm{E}|x_{t-i}|^r \mathrm{E}|\varpi_{t-j}|^r + C\prod_{i=1}^{t}\mathrm{E}|x_{t-i}|^r \mathrm{E}|\frac{\partial \tilde{\delta}_0}{\partial \theta}|^r$$

where $\varpi_t = A(z_t - \tilde{z}_t) + \frac{\partial \tilde{\delta}_t}{\partial \theta}(x_t - \tilde{x}_t)$ with $\mathrm{E}|\varpi_t|^r = O(\rho^t)$. Using Assumption (A11) and Jensen's inequality, $\mathrm{E}|x_{t-i}|^r \leq b < 1$, for $0 < r \leq 2$. Hence,

$$\mathrm{E}\|\frac{\partial \delta_t}{\partial \theta} - \frac{\partial \tilde{\delta}_t}{\partial \theta}\|^r \leq C\sum_{j=1}^{t} b^{j-1}\mathrm{E}|\varpi_{t-j}|^r + O(b^t) = O(tb^{j-1})O(\rho^{t-j}) + O(b^t) = O(t\bar{\rho}^t)$$

where $\bar{\rho} = \max(b, \rho)$. This implies that

$$\mathrm{E}\left\|\frac{\partial l_t}{\partial \theta} - \frac{\partial \tilde{l}_t}{\partial \theta}\right\|^r = O(t\bar{\rho}^t),$$

which is sufficient to establish the result.

$\square$

Lemma 8: Under Assumptions (A1)-(A11), there exists a neighborhood $N(\theta_0) \subseteq \Theta$ around θ_0 such that

$$\sup_{\theta \in N(\theta_0)}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial \theta \partial \theta'}\right\| = o_{a.s.}(1)$$

Proof: We have

$$\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} = (\tilde{u}_t - u_t)\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} + u_t\left(\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right) + \left(\frac{\partial\tilde{u}_t}{\partial\delta_t} - \frac{\partial u_t}{\partial\delta_t}\right)\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} + \frac{\partial u_t}{\partial\delta_t}\left(\frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{\delta}_t}{\partial\theta'} - \frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}\right)$$

In the proof of Lemma 7 we have shown that $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\tilde{u}_t - u_t|^r = O(\rho^t)$ and $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\frac{\partial\tilde{u}_t}{\partial\delta_t} - \frac{\partial u_t}{\partial\delta_t}|^r = O(\rho^t)$, while $\mathrm{E}\sup_{\theta\in N(\theta_0)}\|\frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{\delta}_t}{\partial\theta'} - \frac{\partial\delta_t}{\partial\theta}\frac{\partial\delta_t}{\partial\theta'}\|^r = O(t\rho^t)$. It remains to consider the second term.

$$\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} = \frac{\partial Az_{t-1}}{\partial\theta'} + \frac{\partial\delta_{t-1}}{\partial\theta}\frac{\partial x_{t-1}}{\partial\theta'} + x_{t-1}\frac{\partial^2\tilde{\delta}_{t-1}}{\partial\theta\partial\theta'}$$

Following the same arguments as in the proof of Lemma 7, we can show that

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial Az_{t-1}}{\partial\theta'} - \frac{\partial Az_{t-1}}{\partial\theta'}\right\|^r = O(t\rho^t)$$

and

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial\delta_t}{\partial\theta}\frac{\partial x_t}{\partial\theta'} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{x}_t}{\partial\theta'}\right\|^r = O(t\rho^t)$$

Let $\zeta_t = \frac{\partial Az_t}{\partial\theta'} - \frac{\partial Az_t}{\partial\theta'} + \frac{\partial\delta_t}{\partial\theta}\frac{\partial x_t}{\partial\theta'} - \frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{x}_t}{\partial\theta'} + \frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'}(\tilde{x}_t - x_t)$. Then,

$$\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'} = \zeta_{t-1} + x_{t-1}\left(\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right) \tag{C.23}$$

$$= \sum_{j=1}^{t}\prod_{i=1}^{j-1}x_{t-i}\zeta_{t-j} + \prod_{i=1}^{t}x_{t-i}\frac{\partial^2\tilde{\delta}_0^*}{\partial\theta\partial\theta'} \tag{C.24}$$

and since $\mathrm{E}\sup_{\theta\in N(\theta_0)}|\zeta_t|^r = O(t\rho^t)$ and $\mathrm{E}\sup_{\theta\in N(\theta_0)}\prod_{i=1}^{j-1}|x_{t-i}|^r = O(b^{j-1})$, we have

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial^2\tilde{\delta}_t}{\partial\theta\partial\theta'} - \frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\right\| = O(tb^{j-1})O(t\rho^{t-j}) + O(b^t) = O(t^2\bar{\rho}^t),$$

This implies that

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| = O(t^2\bar{\rho}^t),$$

which is a summable sequence in t. Hence, the result follows by an application of the generalized Chebychev inequality and the Borel-Cantelli lemma.

$\square$

Lemma 9: Under Assumptions (A1–(A11),

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\tilde{\theta}_T)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \right\| \to 0 \;\; a.s.$$

where $\tilde{\theta}_T$ is on the line segment between $\hat{\theta}_T$ and θ_0.

Proof: Applying the mean-value theorem for the second order derivative function about θ_0, gives

$$\left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \tilde{l}_t(\theta_0)}{\partial\theta\partial\theta'} \right\| \leq \sup_{\theta\in N(\theta_0)} \left\{ \left\| \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta'}\left(\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right)\right\| \cdot \left\|\tilde{\theta}_T-\theta_0\right\| \right\} \tag{C.25}$$

where $N(\theta_0)\subseteq\Theta$ is a small neighborhood around θ_0.

It suffices to check that

$$\mathrm{E}\sup_{\theta\in N(\theta_0)}\left\|\frac{\partial \mathrm{vec}}{\partial\theta'}\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'}\right\| < \infty, \tag{C.26}$$

where, almost surely,

$$\begin{aligned}\frac{\partial \mathrm{vec}}{\partial\theta'}\frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta\partial\theta'} &= u_t\frac{\partial \mathrm{vec}}{\partial\theta'}\frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'} + \frac{\partial u_t}{\partial\delta_t}\mathrm{vec}\left(\frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'}\right)\frac{\partial\delta_t}{\partial\theta'} \\ &+ \frac{\partial^2 u_t}{\partial\delta_t^2}\mathrm{vec}\left(\frac{\partial\tilde{\delta}_t}{\partial\theta}\frac{\partial\tilde{\delta}_t}{\partial\theta'}\right)\frac{\partial\tilde{\delta}_t}{\partial\theta'} + \frac{\partial u_t}{\partial\delta_t}(I_m+K_{mm})(\frac{\partial\delta_t}{\partial\theta}\otimes I_m)\frac{\partial^2\delta_t}{\partial\theta\partial\theta'}\end{aligned}$$

and where K_{mm} is the commutation matrix and $m=\dim(\theta)$.

The second, third and fourth terms can be uniformly bounded similar to Lemma 7. For the first term, note that

$$\frac{\partial \mathrm{vec}}{\partial\theta'}\frac{\partial^2 \tilde{\delta}_t(\theta)}{\partial\theta\partial\theta'} = \frac{\partial \mathrm{vec}}{\partial\theta'}\left(\frac{\partial A z_{t-1}}{\partial\theta'} + \frac{\partial\delta_{t-1}}{\partial\theta}\frac{\partial x_{t-1}}{\partial\theta'} + x_{t-1}\frac{\partial^2\tilde{\delta}_{t-1}}{\partial\theta\partial\theta'}\right)$$

A typical element of this matrix, denoting the i-th row of A as A_i, is given by

$$\begin{aligned}
\frac{\partial^3 \tilde{\delta}_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} &= \frac{\partial A_i}{\partial\theta_j}\frac{\partial z_{t-1}}{\partial\theta_k} + \frac{\partial A_i}{\partial\theta_k}\frac{\partial z_{t-1}}{\partial\theta_j} + A_i \frac{\partial^2 z_{t-1}}{\partial\theta_j \partial\theta_k} \\
&+ \frac{\partial^2 \delta_{t-1}}{\partial\theta_i \partial\theta_k}\frac{\partial x_{t-1}}{\partial\theta_j} + \frac{\partial \delta_{t-1}}{\partial\theta_i}\frac{\partial^2 x_{t-1}}{\partial\theta_j \partial\theta_k} + \frac{\partial x_{t-1}}{\partial\theta_k}\frac{\partial^2 \delta_{t-1}}{\partial\theta_i \partial\theta_j} + x_{t-1}\frac{\partial^3 \tilde{\delta}_{t-1}(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \\
&= \sum_{j=1}^{\infty} \vartheta_{t-j} \prod_{i=1}^{j-1} x_{t-i}
\end{aligned}$$

where $\vartheta_t = \frac{\partial A_i}{\partial\theta_j}\frac{\partial z_t}{\partial\theta_k} + \frac{\partial A_i}{\partial\theta_k}\frac{\partial z_t}{\partial\theta_j} + A_i \frac{\partial^2 z_t}{\partial\theta_j \partial\theta_k} + \frac{\partial^2 \delta_t}{\partial\theta_i \partial\theta_k}\frac{\partial x_t}{\partial\theta_j} + \frac{\partial \delta_t}{\partial\theta_i}\frac{\partial^2 x_t}{\partial\theta_j \partial\theta_k} + \frac{\partial x_t}{\partial\theta_k}\frac{\partial^2 \delta_t}{\partial\theta_i \partial\theta_j}$. Each term of ϑ_t can be uniformly bounded in a neighborhood of θ_0 and a repeated application of the c_r inequality yields $\mathrm{E}\sup_{\theta \in N(\theta_0)} |\vartheta_t| < \infty$. Since $\mathrm{E}\prod_{i=1}^{j} |x_{t-i}| = O(\rho^j)$ as in the proof of Lemma 3, the infinite sum in (C.27) is convergent and we obtain

$$\mathrm{E} \sup_{\theta \in N(\theta_0)} \left| \frac{\partial^3 \tilde{\delta}_t(\theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < \infty$$

which implies (C.26). Finally, since by the ergodic theorem, given (C.26),

$$\limsup_{T\to\infty} \left\| \frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in N(\theta_0)} \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta \partial\theta'} \right\| = \mathrm{E} \left\| \sup_{\theta \in N(\theta_0)} \frac{\partial \mathrm{vec}}{\partial\theta'} \frac{\partial^2 \tilde{l}_t(\theta)}{\partial\theta \partial\theta'} \right\| < \infty$$

the desired result follows because $\hat{\theta}_T \to \theta_0$ a.s. by Theorem 1.

□

D Monte Carlo simulation

In this section we provide insights into the finite sample first and second order behaviour of the PML estimator in score-driven rotation models. We first outline the simulation design and subsequently discuss simulation outcomes. The section concludes with some recommendations that one might draw from the Monte-Carlo analysis for empirical work.

D.1 Simulation design

The purpose of the simulation study is to uncover the scope of the PML estimator provided in Section 3 for evaluating the dynamic model in (2.1) and (2.3) and, hence, its potential to extract the underlying independent components from orthogonalized data. The data generating processes are specified with two alternative choices of β, i.e. $\beta = 0.9, 0.95$, and three alternative choices of κ, i.e. $\kappa = 0.1, 0.05, 0.025$. For all data generating processes we set $\omega = 1.30$. The alternative choices of β and κ are combined to obtain overall six parameterizations of the underlying data generating process ($\theta = (\omega, \beta, \kappa)'$).

Throughout we generate data from marginal standardized Student-t random variables with $\nu_1 = 5$ and $\nu_2 = 6$ degrees of freedom. For PML estimation we distinguish two scenarios. First, we assume the correct degrees of freedom ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$). In the second scenario, the interest turns to the effects of performing PML estimation under a false assumption on the degrees of freedom ($\nu_1^{(0)} \neq \nu_1, \nu_2^{(0)} \neq \nu_2$). Specifically, we set $\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$ (instead of $\nu_1^{(0)} = 5$ and $\nu_2^{(0)} = 6$). Alternative sample sizes are $T = 1000, 2500, 5000$ and $T = 10000$. For each experiment we draw $T + 250$ sample observations, and disregard the first 250 observations to immunize the outcomes of the Monte-Carlo analysis against initial effects.

For the numerical log-likelihood maximization we use the initial parameters as

$\theta^{(0)} = (1.63, 0.98, 0.01)'$ in the case of $T = 10000$. For analysing shorter samples of length $T = 1000, 2500$ and $T = 5000$ we try four alternative initializations $\theta^{(0)}$ and select the particular choice which obtains the largest log-likelihood statistics (the alternative choices are (1.95,0.98,0.01)', (1.43,0.95,0.05)', (1.17,0.90,0.10)' and (0.98,0.70,0.15)'). We use the BHHH algorithm of Berndt et al. (1974) for log-likelihood maximization which has the convenient feature that its implementation only requires first order derivatives. In a stylized form the BHHH algorithm delivers updates of the estimated parameter vector as

$$\hat{\theta}^{(s)} = \hat{\theta}^{(s-1)} + \phi_s \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \Big|_{\theta=\hat{\theta}^{(s-1)}} \right)^{-1} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \Big|_{\theta=\hat{\theta}^{(s-1)}},$$

where s indexes the iteration sequence of the optimizer and $\phi_s > 0$ is used to modify the step length at each iteration. Specifically, we employ alternative choices $\phi_s \in \{1.0, 0.5, 0.25, 0.10, 0.05\}$ to detect most effective updates of the parameter vector at each iteration step.

Each Monte-Carlo experiment is performed with $R = 1000$ replications. To characterise the performance of PML estimation we document the following summary statistics for the typical elements of $\hat{\theta}$ denoted $\hat{\theta}_i$:

- Root mean squared error: $\mathrm{RMSE}(\hat{\theta}_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{r,i} - \theta_i)^2}$, where $\hat{\theta}_{r,i}$ is the estimator of the i-th component of θ obtained from the r-th Monte-Carlo replication.

- Bias: $\mathrm{BIAS}(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_{r,i} - \theta_i$

- Standard deviation: $\mathrm{SD}(\hat{\theta}_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_{r,i} - \bar{\theta}_i)^2}$, where $\bar{\theta}_i = \frac{1}{R}\sum_{r=1}^{R}\hat{\theta}_{r,i}$

- Approximation based on $\widehat{I}^{-1}$: $\widehat{\mathrm{SD}}_I(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\mathcal{S}_{ii}^{(r)}}$, with $\mathcal{S}^{(r)} = \widehat{I}_r^{-1}/T$

- Approximation based on $\widehat{V}$: $\widehat{\mathrm{SD}}_V(\hat{\theta}_i) = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\mathcal{S}_{ii}^{(r)}}$, with $\mathcal{S}^{(r)} = \widehat{V}_r/T$

D.2 Simulation results

Simulation results for DGPs parameterized with $\beta = 0.90$ and $\beta = 0.95$ are documented in Tables D.1 and D.2, respectively. The left (right) hand side panels of the tables show simulation outcomes for experiments with $\nu_i^{(0)} = \nu_i$, $i = 1, 2$ ($\nu_1^{(0)} \neq \nu_1$, $\nu_2^{(0)} \neq \nu_2$), i.e., PML estimation under knowledge (misspecification) of the actual underlying distributional models. The Monte-Carlo analysis allows for a couple of insights with respect to the first and second order properties of the nonlinear PML estimator:

1. Confirming our results of Theorem 1, all model parameters are estimated consistently. For instance, regarding the DGP with $\beta = 0.9$, $\kappa = 0.05$ the documented RMSE statistics shrink for all parameters with increasing sample size. Focussing on estimates for the response parameter κ, RMSE (absolute BIAS) statistics shrink from 10.68E-02 (1.66E-02) in case of $T = 1000$ to 1.71E-02 (0.17E-02) in case of $T = 10000$ (left hand side panels of Table D.1). With $T = 2500$ sample observations, these performance statistics are 4.99E-02 (RMSE) and 1.13E-02 (absolute BIAS). Since BIAS estimates vanish with increasing sample size, the RMSE statistics approach the documented standard errors of the estimators.

2. As it is implied by the documented BIAS statistics, it holds for all alternative DGPs that the PML estimator results in an underestimation of ω and β, and an overestimation of κ.

3. The consistency of the PML estimator holds irrespective of assuming correct or false degrees of freedom for the underlying distributions of independent innovations. However, assuming false degrees of freedom ($\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$) comes with a loss in estimator precision. For instance, under an assumption

of false degrees of freedom, the relative efficiency in terms of the RMSE of estimating κ is about 69% in the case of $T = 2500$ and about 64% for $T = 10000$ ($\beta = 0.9$, $\kappa = 0.05$). Hence, as an implication for empirical practice, it is worth to strive for a most suitable distribution, since it promises efficiency gains of parameter estimates.

4. Approximating estimation uncertainty by the asymptotic result in (3.8) suffers from strong biases in small samples. Under the assumption of a correctly specified distributional model ($I = J$) both covariance estimators $\widehat{I}^{-1}$ and $\widehat{V}$ can be expected to quantify estimation uncertainty reliably in sufficiently large samples. As it turns out, however, both estimators tend to underestimate the empirical variance of $\hat{\kappa}$ and overestimate the variances of $\hat{\omega}$ and $\hat{\beta}$, irrespective of the true response parameter and the actual sample size. The estimation biases for all these second order characteristics are sizeable in small samples ($T = 1000, 2500$) and vanish only slowly. For example, in the model with $\beta = 0.9$, $\kappa = 0.05$ and in case of $T = 2500$, the estimate of the standard deviation of $\hat{\kappa}$ using $\widehat{I}^{-1}$ falls short of the true empirical standard deviation by about 30%, while that of $\hat{\beta}$ exceeds the benchmark by about 3%.

5. Using the robust covariance estimator $\widehat{V}$ if the correct distributional model is employed for PML estimation ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$) results in a loss of estimator precision in comparison with using $\widehat{I}^{-1}$. For instance, in the scenario described before ($T = 2500, \beta = 0.9$, $\kappa = 0.05$), the assessment of estimation uncertainty by means of $\widehat{V}$ yields standard error approximations which fall short of the empirical standard error by 47% (SD of $\hat{\kappa}$) and 15% (SD of $\hat{\beta}$). In the large sample case of $T = 10000$, the simulation outcomes for both variance estimators approach each other. Moreover, these performance measures come

close to the empirical standard errors of the estimators.

6. For scenarios of misspecified distributional models ($\nu_1^{(0)} = 7$ and $\nu_2^{(0)} = 8$), the large sample case of $T = 10000$ is not sufficient to indicate an overall lead of the robust covariance estimator $\widehat{V}$ over the nonrobust estimator $\widehat{I}^{-1}$. For instance, for the DGP with $\beta = 0.9$ and $\kappa = 0.05$, the Monte-Carlo standard deviation of $\hat{\beta}$ (scaled by 100) is 9.16, while the approximations obtained from $\widehat{I}^{-1}$ and $\widehat{V}$ are, respectively, 7.90 and 5.77.

7. The documented results on finite sample biases of PML covariance estimation indicate that one should be careful when interpreting commonly provided inferential outcomes (e.g. t-ratios). For instance, sample-based assessments of I^{-1} or V might be employed to construct confidence intervals for estimators $\hat{\kappa}$. It might be of interest to obtain the frequencies at which such confidence intervals fail to include the true parameter κ, corresponding to the size of an equivalent Wald test of $H_0 : \kappa = \kappa_0$. One can also obtain the empirical non-coverage of $\kappa = 0$, which would correspond to the power under classical conditions. In our case, however, the case $\kappa = 0$ is excluded by Assumption (A8), i.e., κ is at the boundary of the admissible parameter space. Hence, there are unidentified parameters under the null hypothesis, and the asymptotic theory developed in Section 3 does not apply for this case. To formally test the hypothesis $H_0 : \kappa = 0$, i.e. constancy of the rotation angle δ_t, we recommend Lagrange Multiplier (LM) tests, e.g. Portmanteau-type tests for serial correlation of the scores, which we will use in the empirical illustration in Section 4, see e.g. Chapter 2.5 of Harvey (2013).

   Table D.3 shows for the case $\beta = 0.9$, $\kappa = 0.05$ empirical frequencies of non-coverage for such confidence intervals with nominal 99% and 95% coverage.

The results confirm that analytical approximations of estimation uncertainty for $\hat{\kappa}$ are too small, such that the constructed confidence intervals fail to achieve their nominal coverage throughout. In the large sample case of $T = 10000$, the empirical type-I error probabilities still exceed their nominal counterparts, where the empirical sizes are 9.7 and 5.2 for nominal significance levels of 5% and 1%, respectively. One might interpret the results documented for a nominal level of 99% to imitate a size adjusted test for the null hypothesis of insignificance of $\hat{\kappa}$ with 5% nominal significance. With this interpretation of 99% confidence intervals, PML inference shows considerable (size adjusted) power against a model with static rotation ($\kappa = 0$). Defined in this sense, (size adjusted) power increases from 18.4 ($T = 1000$) to 87.1 ($T = 10000$) percent.

D.3 Recommendations for empirical practice

The simulation results indicate that the parameters governing the dynamic rotation patterns can be retrieved consistently from the data. The scales of estimation errors shrink quickly in case of realistic sample sizes covering 4 or 10 years of daily data ($T = 1000, 2500$). To guard the empirical analysis against (inefficient) estimates derived from local extremes of the pseudo likelihood, it is worth to initialize non-linear optimizations by means of alternative parameter selections. The precision of parameter estimation benefits from conditioning PML estimation on the correct distributional assumptions ($\nu_i^{(0)} = \nu_i$, $i = 1, 2$). Hence, for empirical practice it is reasonable (i) to estimate the degrees of freedom (ν_i) and dynamic parameters in θ jointly or (ii) to condition the PML estimation of the dynamic parameters in θ on a variety of alternative assumptions with regard to the degrees of freedom parameters ($\nu_i^{(0)}$, $i = 1, 2$). Owing to computational complexity the advice in (ii) seems preferable to the suggestion in (i).

The approximation of estimation uncertainty by means of $\widehat{I}^{-1}$ and $\widehat{V}$ suffers from sizeable biases in small samples and converges only slowly to the true second order features of the estimators. In light of (i) the convenient first order properties of the PML estimator and (ii) its asymptotic normality shown in Theorem 2, bootstrap approaches may be suggested as an alternative for inferential analysis in finite samples. The underlying model assumption of independence largely facilitates the implementation of bootstrap schemes by means of independent drawings with replacement. For the empirical application in Section 4 we propose a specific bootstrap scheme to complement analytical covariance estimates.

κ	T	θ_i	RMSE	BIAS	SD	$\widehat{\mathrm{SD}}_I$	$\widehat{\mathrm{SD}}_V$	RMSE	BIAS	SD	$\widehat{\mathrm{SD}}_I$	$\widehat{\mathrm{SD}}_V$
.100	1000	ω	6.97	-0.48	6.95	46.73	16.33	8.45	-0.10	8.45	13.49	7.18
		β	32.35	-10.16	30.72	109.00	123.51	36.70	-11.80	34.75	28.67	12.68
		κ	9.04	0.69	9.01	5.20	2.38	12.29	4.10	11.59	9.64	10.12
	2500	ω	3.90	-0.17	3.90	10.12	2.61	4.03	-0.14	4.03	4.41	2.52
		β	14.72	-3.09	14.39	7.05	5.39	14.98	-3.23	14.63	10.30	6.08
		κ	3.76	0.53	3.72	2.87	1.99	6.28	3.66	5.10	4.98	2.42
	5000	ω	2.58	-0.10	2.58	2.57	3.35	2.61	-0.10	2.61	3.04	1.90
		β	5.13	-0.95	5.04	3.64	2.98	5.01	-0.96	4.92	4.69	2.85
		κ	2.30	0.16	2.30	1.88	1.43	4.29	2.96	3.11	3.17	1.78
	10000	ω	1.77	-0.11	1.77	1.65	1.48	1.84	-0.12	1.83	2.09	1.25
		β	2.29	-0.32	2.27	2.05	1.65	2.48	-0.40	2.44	2.65	1.61
		κ	1.39	0.02	1.39	1.25	1.05	3.36	2.72	1.97	2.08	1.18
.050	1000	ω	6.57	-0.62	6.54	59.25	15.51	6.62	-0.28	6.61	79.87	19.26
		β	55.73	-25.44	49.59	40.88	29.70	56.39	-25.82	50.13	551.26	434.31
		κ	10.68	1.66	10.55	5.91	3.22	14.15	3.54	13.70	11.16	4.54
	2500	ω	3.83	-0.22	3.83	16.76	16.46	3.80	-0.20	3.80	18.60	3.98
		β	34.02	-12.16	31.77	32.75	27.09	38.38	-13.64	35.87	41.80	23.62
		κ	4.99	1.13	4.86	3.38	2.62	7.26	2.84	6.68	5.93	3.34
	5000	ω	2.56	-0.12	2.56	2.98	3.45	2.58	-0.13	2.57	3.87	2.40
		β	19.90	-5.08	19.24	12.88	11.63	24.14	-6.54	23.24	19.04	12.74
		κ	2.86	0.47	2.82	2.23	1.94	4.23	2.03	3.71	3.82	2.47
	10000	ω	1.77	-0.07	1.76	1.73	1.68	1.77	-0.08	1.77	2.19	1.70
		β	7.98	-1.68	7.80	5.93	5.62	9.37	-1.96	9.16	7.90	5.77
		κ	1.71	0.17	1.70	1.47	1.36	2.68	1.58	2.16	2.44	1.72
.025	1000	ω	6.75	-0.18	6.75	54.37	22.47	6.63	-0.34	6.62	370.82	126.34
		β	62.52	-29.66	55.03	296.95	250.95	65.85	-31.46	57.84	4293.2	3278.3
		κ	11.13	1.64	11.01	5.94	2.61	15.08	2.08	14.93	11.19	3.64
	2500	ω	3.65	-0.32	3.63	48.65	15.90	3.68	-0.23	3.67	102.13	13.77
		β	55.13	-24.71	49.28	906.42	748.17	62.10	-28.47	55.20	105.56	46.73
		κ	6.00	1.37	5.85	3.54	2.80	8.21	2.27	7.89	6.22	3.40
	5000	ω	2.47	-0.06	2.47	14.74	3.03	5.54	-0.05	5.54	141.14	45.68
		β	42.99	-16.12	39.86	183.17	201.70	47.97	-18.32	44.34	41.65	27.24
		κ	3.43	0.72	3.36	2.40	2.07	4.72	1.52	4.47	4.07	3.01
	10000	ω	1.71	-0.07	1.70	18.77	1.90	1.72	-0.06	1.72	3.70	4.99
		β	25.56	-7.47	24.45	26.49	25.71	30.73	-9.48	29.24	29.21	21.10
		κ	1.95	0.36	1.91	1.59	1.47	2.83	1.18	2.57	2.66	1.88

Table D.1: Simulation results for DGPs with $\omega = 1.3$ and $\beta = 0.90$. The true parameter κ is given in the first column. PML estimation results are shown for presuming correct ($\nu_i^{(0)} = \nu_i$, left hand side panels) and false degrees of freedom ($\nu_i^{(0)} \neq \nu_1$, right hand side). Documented loss measures include root mean squared errors (RMSE), bias (BIAS) and standard deviation estimates (SD). Further performance statistics are PML approximations of SD based on estimates $\widehat{I}^{-1}$ and $\widehat{V}$ of the asymptotic covariance matrix. All summary statistics are multiplied with 100.

κ	T	θ_i	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$	RMSE	BIAS	SD	$\widehat{\text{SD}}_I$	$\widehat{\text{SD}}_V$
.100	1000	ω	11.44	-0.08	11.44	37.85	17.08	9.53	-0.04	9.53	22.83	44.18
		β	29.12	-7.93	28.02	910.49	713.45	31.22	-8.61	30.00	22.20	9.79
		κ	6.45	0.62	6.42	4.10	1.83	9.89	3.70	9.18	7.46	9.18
	2500	ω	4.53	-0.19	4.52	3.93	2.03	4.65	-0.16	4.64	5.04	0.58
		β	4.25	-0.95	4.14	2.61	1.99	4.79	-1.04	4.67	3.95	2.29
		κ	2.55	0.23	2.53	1.99	1.12	4.79	3.12	3.63	3.43	1.14
	5000	ω	2.95	-0.08	2.95	2.66	1.31	3.11	-0.08	3.11	3.37	0.76
		β	1.52	-0.32	1.48	1.23	0.79	1.56	-0.31	1.53	1.61	0.69
		κ	1.48	0.04	1.48	1.28	0.67	3.47	2.79	2.06	2.12	0.69
	10000	ω	1.96	-0.06	1.96	1.80	0.91	2.12	-0.11	2.11	2.25	0.56
		β	0.84	-0.13	0.83	0.73	0.41	0.89	-0.12	0.88	0.93	0.34
		κ	0.94	0.04	0.93	0.83	0.44	2.96	2.63	1.36	1.36	0.46
.050	1000	ω	11.02	- 0.09	11.02	477.52	115.01	10.76	-0.13	10.76	135.59	35.57
		β	47.50	-20.56	42.82	45.65	33.67	50.68	-22.29	45.51	4243.3	1847.8
		κ	9.65	2.09	9.42	5.32	2.71	13.85	4.26	13.17	9.86	3.30
	2500	ω	4.26	-0.22	4.26	9.58	4.16	4.28	-0.24	4.28	9.57	5.14
		β	23.34	-6.57	22.39	13.88	13.15	26.03	-7.22	25.01	17.89	11.61
		κ	4.07	0.91	3.97	2.64	2.13	5.72	2.48	5.16	4.51	2.27
	5000	ω	2.85	-0.09	2.85	2.78	2.54	2.85	-0.10	2.85	3.53	2.48
		β	11.16	-2.01	10.97	5.45	4.96	11.91	-2.14	11.72	7.98	5.55
		κ	2.02	0.24	2.00	1.63	1.44	3.08	1.70	2.57	2.73	1.80
	10000	ω	1.95	-0.06	1.94	1.92	1.78	1.96	-0.06	1.96	2.42	1.82
		β	2.74	-0.56	2.68	2.21	2.06	3.00	-0.62	2.93	2.91	2.08
		κ	1.15	0.07	1.15	1.06	0.94	2.10	1.43	1.54	1.75	1.24
.025	1000	ω	7.98	-0.19	7.98	70.32	27.12	6.70	-0.42	6.68	66.86	18.18
		β	67.20	-35.00	57.37	148.03	89.57	69.18	-35.98	59.09	66.44	52.41
		κ	10.73	1.99	10.55	5.75	2.88	14.88	1.95	14.75	10.84	3.50
	2500	ω	3.87	-0.28	3.86	33.10	15.90	3.96	-0.24	3.96	64.40	8.06
		β	48.57	-21.26	43.67	45.78	39.04	52.42	-22.88	47.17	49.82	31.83
		κ	5.41	1.42	5.22	3.23	2.58	7.69	2.53	7.26	5.59	3.01
	5000	ω	2.63	-0.09	2.62	64.50	102.22	2.64	-0.11	2.64	8.20	6.52
		β	33.78	-10.90	31.98	108.06	97.22	37.98	-12.80	35.76	33.68	25.11
		κ	2.83	0.66	2.75	1.97	1.73	3.84	1.42	3.56	3.34	2.21
	10000	ω	1.82	-0.06	1.82	1.84	1.82	1.83	-0.07	1.83	2.32	1.80
		β	16.25	-3.85	15.79	9.54	9.06	20.36	-4.88	19.77	12.81	9.33
		κ	1.52	0.28	1.49	1.23	1.16	2.28	1.06	2.02	2.06	1.48

Table D.2: Simulation results for DGPs with $\omega = 1.3$ and $\beta = 0.95$. For further comments see Table D.1.

	95% coverage				99% coverage			
T	1000	2500	5000	10000	1000	2500	5000	10000
$\kappa = 0.05$	26.8	18.2	12.9	9.7	20.6	12.7	8.4	5.2
$\kappa = 0$	30.5	46.7	73.4	97.2	18.7	24.3	48.2	87.1

Table D.3: Non-coverage frequencies for $\kappa = 0.05$ and $\kappa = 0$, of confidence intervals for $\hat{\kappa}$ that are determined by means of standard errors implied by $\widehat{I}^{-1}$. The true model parameters are $\kappa = 0.05$ and $\beta = 0.9$.

References

Berndt, E. K., B. H. Hall, R. E. Hall, and J. A. Hausman (1974). “Estimation and inference in nonlinear structural models”. *Annals of Economic and Social Measurement* 3/4, pp. 653–665.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Vol. 52. Cambridge University Press.

Lancaster, H. O. (1954). “Traces and cumulants of quadratic forms in normal variables”. *Journal of the Royal Statistical Society. Series B (Methodological)* 16.2, pp. 247–254.

Pfanzagl, J. (1969). “On the measurability and consistency of minimum contrast estimates”. *Metrika: International Journal for Theoretical and Applied Statistics* 14.1, pp. 249–272.

Scott, D. J. (1973). “Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach”. *Advances in Applied Probability* 5.1, pp. 119–137.

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