Reliable Residuals for Multivariate Nonlinear Time Series Models

Leena Kalliovirta, University of Helsinki[1],

Pentti Saikkonen, University of Helsinki

Abstract

We generalize and apply quantile residuals to multivariate nonlinear time series models for which conventional residuals are unreliable. We formulate a general framework of obtaining misspecification tests that also allows non-ergodic data and takes the effect of parameter estimation properly into account. Computationally simple tests developed to detect serial correlation, conditional heteroscedasticity, and non-normality in quantile residuals illustrate the usefulness of our approach. Our tests are generalizations of previous tests based on moments of conventional residuals and the Lagrange Multiplier principle. We apply the developed tests to exchange rate series. In simulations our tests show good size and power properties.

JEL classification: C32, C52

Keywords: Quantile residual, Lagrange Multiplier test, Generalized autoregressive conditional heteroscedasticity, Multivariate Generalized Orthogonal Factor GARCH model, Regime switching model.

[1] Department of Economic and Political Studies, Arkadiankatu 7 (P.O.Box 17), 00014 University of Helsinki, fax +358-9-191 28736, email: leena.kalliovirta@helsinki.fi

1 Introduction

This paper studies multivariate quantile residuals and, based on their asymptotic properties, develops misspecification tests in a general likelihood framework. These residuals exist for any parametric model with continuous (conditional) cumulative distribution function. Thus, they are applicable, for example, to various regime-switching models and other models based on mixture distributions or involving latent variables. With these models the use of traditional (or Pearson's) residuals, in contrast, leads to erroneous inference. Kalliovirta (2006) discusses this in a univariate setting and demonstrates the usefulness of quantile residual based tests. This previous work and the recent interest in multivariate models based on mixtures of distributions motivate the generalization of this paper.

The idea of quantile residuals originates from Rosenblatt (1952) and Cox and Snell (1968), and was developed, among others, by Smith (1985), Dunn and Smyth (1996), and Palm and Vlaar (1997). The term *quantile residual* is due to Dunn and Smyth (1996). Two transformations define quantile residuals: 1) the estimated cumulative distribution function implied by the model transforms the observations into approximately independent uniformly distributed random variables, and 2) the inverse of the cumulative distribution function of the standard normal distribution retransforms these variables into approximately independent standard normal. These results assume that the model is correctly specified and parameters are consistently estimated. If not, quantile residuals are expected to exhibit detectable departures from the characteristic properties described above.

We give regularity conditions under which a central limit theorem holds for smooth functions of quantile residuals. This result yields misspecification tests which, under correct specification, have limiting χ^2-distributions. These tests are applicable in standard stationary or ergodic models as well as in non-ergodic models such as co-integrated VAR-models with the number of unit roots correctly specified and other models involving trending variables. Furthermore, our approach is theoretically sound: it takes the uncertainty caused by parameter estimation properly into account. We illustrate the approach by deriving tests aimed to detect serial correlation, conditional heteroscedasticity, and non-normality in multivariate quantile residuals. Using our framework one can similarly design more tests to detect other departures from the characteristic properties of quantile residuals. Our three tests have the following advantageous properties. First, they are easy to use once the parameter estimates of the model are available. Second, they have Lagrange Multiplier (LM) or score test interpretations and are, therefore, asymptotically optimal against local alternatives. Third, in conventional models with normal errors, where quantile residuals are Pearson's residuals, our test statistics reduce to classical tests for autocorrelation, conditional heteroscedasticity and non-normality.

Several previous papers (see Kalliovirta (2006), and the references therein) consider quantile residuals. Most of them concentrate on out-of-sample forecast evaluation in a univariate setting and, contrary to us, lack proper theoretical justification for the employed procedures. Only Hong and Li (2005) and Bai and Chen (2008) study in-sample evaluation of multivariate models. Compared to the nonparametric approach of Hong and Li (2005), our approach, in addition to being theoretically optimal in parametric setting, is simpler in practice. The reliability of the generalized Kolmogorov-Smirnov test of Bai and Chen (2008) requires quantile residuals to be independent. This property, however, may not hold and should be tested. Our tests of serial correlation and conditional heteroscedasticity are designed for this purpose. Furthermore, we allow for non-ergodic data and, therefore, our approach applies to a wider class of models than the ones in Hong and Li (2005) and Bai and Chen (2008).

Multivariate quantile residuals are functions of marginal and conditional distributions at each time point. Even if the observations are multivariate, we suggest also to consider univariate quantities, henceforth joint quantile residuals, that are functions of products of marginal and conditional distributions. Previously, Diebold et al. (1999), Clements and Smith (2000), and Clements and Smith (2002) have applied this idea in the context of multivariate density forecast evaluation.

Our general testing principle applies to a wide range of multivariate models including mixture distribution based models or models with unobservable regime switching. These include Markov switching VAR models rigorously studied by Douc et al. (2004) and applied in special cases, for example, by Paap et al. (2009) and Lanne et al. (2010). A similar mixture distribution based VAR model was considered by Lanne and Lütkepohl (2010) whereas a different mixture VAR model, the ACR model, was studied and applied by Bec et al. (2008). Our approach also applies to nonlinear multivariate autoregressive models, multivariate GARCH models, and co-integrated VAR models with the number of unit roots correctly specified, because with these models quantile residuals reduce to conventional residuals. We apply our developed methods to the Multivariate Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen (2007). A mixture version of this model illustrates how our approach supports graphical analysis and is able to formally compare the goodness of fit between models based on different structural or distributional assumptions or both.

Our simulation study shows that the size properties of the proposed tests are satisfactory once a simulation method is used to compute a covariance matrix needed in the test statistics. We use a previous multinormality test as an example to demonstrate by simulation that the size of a test can be totally incorrect if traditional residuals are incorrectly used or if the uncertainty caused by parameter estimation is ignored.

The remainder of this paper is organized as follows. Section 2 defines both the multivariate and joint quantile residuals, and examines their theoretical properties, which are used in Section 3 to derive misspecification tests. Section 4 presents the empirical example, Section 5 shows simulation results, and Section 6 concludes.

2 Quantile residuals

This section recalls the definition of univariate quantile residuals, defines multivariate and joint quantile residuals in a general likelihood framework, and derives a general approach of obtaining misspecification tests.

2.1 Univariate case

Let $\mathbf{y} = (y_1, ..., y_T)$ be a vector of observations with density function $f(\boldsymbol{\theta}_0, \mathbf{y})$, where $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is the unknown true parameter value. Denote with $\mathcal{P} = \{f(\boldsymbol{\theta}, \mathbf{y}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k,\ \mathbf{y} \in \mathbb{R}^T\}$ the collection of potential models for $\mathbf{y}$. For each $f : \boldsymbol{\Theta} \times \mathbb{R}^T \to \mathbb{R}_+$ we can write

$$f(\boldsymbol{\theta}, \mathbf{y}) = \prod_{t=1}^{T} f_{t-1}(\boldsymbol{\theta}, y_t), \tag{1}$$

where $f_{t-1}(\boldsymbol{\theta},y_t) = f(\boldsymbol{\theta},y_t|\mathcal{G}_{t-1})$, $t \in \{1,...,T\}$, $y_t \in \mathbb{R}$, is the conditional density function given $\mathcal{G}_{t-1} = \sigma(\mathbf{Y}_0, Y_1, ..., Y_{t-1})$, the sigma-algebra generated by the random variables $\{\mathbf{Y}_0, Y_1, ..., Y_{t-1}\}$. The random vector $\mathbf{Y}_0$ represents the needed initial values. *The theoretical quantile residual* is defined by

$$R_{t,\boldsymbol{\theta}} = \Phi^{-1}(F_{t-1}(\boldsymbol{\theta}, Y_t)), \tag{2}$$

and *the observed quantile residual* is $r_{t,\widehat{\boldsymbol{\theta}}_T} = \Phi^{-1}(F_{t-1}(\widehat{\boldsymbol{\theta}}_T, y_t))$, where $\Phi^{-1}(\cdot)$ is the inversed cumulative distribution function of the standard normal distribution, $F_{t-1}(\boldsymbol{\theta}, y_t) = \int_{-\infty}^{y_t} f_{t-1}(\boldsymbol{\theta},u)du$ is the conditional cumulative distribution function of y_t, and $\widehat{\boldsymbol{\theta}}_T$ is an estimate of $\boldsymbol{\theta}_0$.

2.2 Multivariate case

Let $\mathbf{y}_1, ..., \mathbf{y}_T$ be vector valued observations with conditional density function $f_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t)$ defined for every $\mathbf{y}_t = (y_{1t}, \cdots, y_{nt})$. The collection of potential models is denoted by $\mathcal{P} = \left\{ f(\boldsymbol{\theta}, \mathbf{y}) : \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^k,\ \mathbf{y} \in \mathbb{R}^{nT} \right\}$.

If the components of $\mathbf{y}_t$ are independent, the quantile residuals extend straightforwardly to the vector case. Because the conditional cumulative distribution function of $\mathbf{y}_t$ has the product form $F_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t) = \prod_{j=1}^{n} F_{j,t-1}(\boldsymbol{\theta}, y_{jt})$, where $F_{j,t-1}(\boldsymbol{\theta}, y_{jt})$ is the marginal distribution function of the jth component, we can make the transformation (2) component-wise.

If the components of $\mathbf{y}_t$ are dependent, quantile residuals are defined as follows. Write the conditional density function of $\mathbf{y}_t$ in the product form

$$f_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t) = \prod_{j=1}^{n} f_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t}) \tag{3}$$

by conditioning with respect to any chosen order of the components. The index $j-1$ in the formula denotes conditioning with respect to the sigma-algebra $\mathcal{A}_{j-1} = \sigma\left\{Y_{i_1,t}, ..., Y_{i_{j-1},t}\right\}$ generated by the component variables. Interpret $f_{i_1,0,t-1}(\boldsymbol{\theta}, y_{i_1,t}) = f_{i_1,t-1}(\boldsymbol{\theta}, y_{i_1,t})$, and $F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t}) = \int_{-\infty}^{y_{i_j,t}} f_{i_j,j-1,t-1}(\boldsymbol{\theta},u)du$. Thus, the vector of *theoretical quantile residual* at time point t takes the form

$$\mathbf{R}_{t,\boldsymbol{\theta}} = \begin{bmatrix} R_{1t,\boldsymbol{\theta}} \\ R_{2t,\boldsymbol{\theta}} \\ \vdots \\ R_{nt,\boldsymbol{\theta}} \end{bmatrix} = \begin{bmatrix} \Phi^{-1}(F_{i_1,t-1}(\boldsymbol{\theta}, Y_{i_1,t})) \\ \Phi^{-1}(F_{i_2,1,t-1}(\boldsymbol{\theta}, Y_{i_2,t})) \\ \vdots \\ \Phi^{-1}(F_{i_n,n-1,t-1}(\boldsymbol{\theta}, Y_{i_n,t})) \end{bmatrix}. \tag{4}$$

This vector is non-unique, it can be formed in $n!$ different ways. The results presented in this paper are independent of the chosen order of conditioning, however. The vector of *observed quantile residual* at time point t is obtained by replacing $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}_T$, an estimate of $\boldsymbol{\theta}_0$, in (4).

One can also base the model evaluation on univariate quantities. Congruent with Clements and Smith (2000), we define *theoretical joint quantile residual* as

$$Q_{t,\boldsymbol{\theta}} = \Phi^{-1}(Z_{t,\boldsymbol{\theta}}), \tag{5}$$

where $Z_{t,\boldsymbol{\theta}} = X_{t,\boldsymbol{\theta}} \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} \left(\ln X_{t,\boldsymbol{\theta}}\right)^k$ with $X_{t,\boldsymbol{\theta}} = \prod_{j=1}^{n} F_{i_j,j-1,t-1}(\boldsymbol{\theta}, Y_{i_j,t})$. Clements and Smith (2000) and Clements and Smith (2002) studied this transformation in a bivariate case and applied it to evaluate forecast densities. Previously, no general form of joint quantile residuals has been suggested. The *observed joint quantile residual* at time point t is obtained by replacing $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}}_T$, an estimate of $\boldsymbol{\theta}_0$, in (5).

2.3 Theoretical properties

Under mild regularity conditions quantile residuals have properties that make them useful in model evaluation: 1) Lemma 2 shows that observed multivariate quantile residuals are asymptotically independently multinormally distributed, if the estimated model is correctly specified, 2) Lemma 3 yields the same result for the observed joint quantile residuals. The following Condition 1 is sufficient for Lemmas 2 and 3 to hold. Unless otherwise stated, all limit statements assume that $T \to \infty$. The symbol $\overset{W}{\to}$ signifies weak convergence, and $\overset{P}{\to}$ signifies convergence in probability.

Condition 1 *Let the following assumptions hold.*

(1) *The collection $\mathcal{P}$ is correctly specified, i.e., $f(\boldsymbol{\theta}_0, \mathbf{y}) \in \mathcal{P}$.*

(2) *$f_{t-1} : \boldsymbol{\Theta} \times \mathbb{R}^n \to \mathbb{R}$ is a continuous conditional density function for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $t = 1, ..., T$.*

(3) *$\widehat{\boldsymbol{\theta}}_T$ is an estimator of $\boldsymbol{\theta}_0$ such that $\widehat{\boldsymbol{\theta}}_T \overset{P}{\to} \boldsymbol{\theta}_0$.*

Lemma 2 *Under Condition 1,*

a) the distribution of the vector of quantile residuals $\begin{bmatrix} \mathbf{R}'_{1,\boldsymbol{\theta}_0} & \cdots & \mathbf{R}'_{T,\boldsymbol{\theta}_0} \end{bmatrix}'$ is multivariate standard normal, where $\mathbf{R}_{t,\boldsymbol{\theta}_0}$ is as in (4) with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

b) for any H fixed, the distribution of $\begin{bmatrix} \mathbf{R}'_{1,\widehat{\boldsymbol{\theta}}_T} & \cdots & \mathbf{R}'_{H,\widehat{\boldsymbol{\theta}}_T} \end{bmatrix}'$ is asymptotically multivariate standard normal, where $\mathbf{R}_{t,\widehat{\boldsymbol{\theta}}_T}$ is as in (4) with $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$, and

c) for any $s \geq 1$, $\mathbf{R}_{t+s,\boldsymbol{\theta}_0}$ is independent of $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$.

The proof is given in Appendix A. Parts a) and b) are used to obtain the tests and part c) is used in some subsequent derivations.

Lemma 3 *Under Condition 1,*

a) the distribution of the vector $\begin{bmatrix} Q_{1,\boldsymbol{\theta}_0} & \cdots & Q_{T,\boldsymbol{\theta}_0} \end{bmatrix}'$ *is multivariate standard normal, where* $Q_{t,\boldsymbol{\theta}_0}$ *is as in (5) with* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$,

b) for any H fixed the distribution of $\begin{bmatrix} Q_{1,\widehat{\boldsymbol{\theta}}_T} & \cdots & Q_{H,\widehat{\boldsymbol{\theta}}_T} \end{bmatrix}'$ *is asymptotically multivariate standard normal, where* $Q_{t,\widehat{\boldsymbol{\theta}}_T}$ *is as in (5) with* $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_T$, *and*

c) for any $s \geq 1$, $Q_{t+s,\boldsymbol{\theta}_0}$ *is independent of* $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$.

The proof is given in Appendix A. Again, we use parts a) and b) to obtain the tests and part c) in some subsequent derivations.

Using the preceding results, one can check the correct model specification by testing whether the observed multivariate or joint quantile residuals are normally and independently distributed. As already mentioned, previous literature mainly considers quantile residuals in the univariate setting without the normalizing transformation whose usefulness was pointed out by Dunn and Smyth (1996), Berkowitz (2001), and Kalliovirta (2006). Specifically, Kalliovirta (2006) shows (in a univariate setting) that the normalizing transformation makes possible to test independence and normality together in a simple way and even motivate the obtained tests by the LM principle (see also Section 2.5).

We, too, advocate the use of the normalizing transformation and show that similar optimality results exist in the multivariate setting. In contrast, such results are unavailable when independence and uniform distribution are tested (cf. Kalliovirta (2006)). Furthermore, the normalizing transformation makes it easy to test the independence using correlations of quantile residuals whereas an identical approach with uniform distribution appears unnatural. Presumably for this reason, previous literature, except Hong and Li (2005), ignores testing the independence hypothesis. The normalizing transformation is advantageous in practice, too. As pointed out by Dunn and Smyth (1996), practitioners often find graphs based on normally distributed residuals easier to interpret than graphs based on the uniform distribution. Moreover, as discussed in Kalliovirta (2006), the normalizing transformation implies that, in several standard models with Gaussian likelihood, quantile residuals simplify to conventional residuals. This property makes comparisons to standard models simple. In contrast, uniformly distributed quantile residuals lack such a convenience.

2.4 General test statistics

This section develops our general framework for obtaining misspecification tests based on multivariate and joint quantile residuals. With different choices of the function g to be introduced shortly, one can construct test statistics for different potential departures from the characteristic properties of quantile residuals. Because our framework is based on fairly standard likelihood theory we only describe the main features and assumptions needed in this section. Precise regularity conditions and supplementary discussion can be found in Appendix A.

Conditional on initial values, the log-likelihood function of the sample takes the form $l_T(\boldsymbol{\theta}, \mathbf{y}) = \sum_{t=1}^{T} l_t(\boldsymbol{\theta}, \mathbf{y}_t) = \sum_{t=1}^{T} \log f_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t)$. We define the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\theta}}_T$ to be any local maximizer of $l_T(\boldsymbol{\theta}; \mathbf{y})$, when such a maximum exists and $+\infty$ otherwise. We assume that $l_T(\boldsymbol{\theta}, \mathbf{y})$ is twice continuously differentiable with the score function $S_T(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} l_T(\boldsymbol{\theta}, \mathbf{Y})$ and the Hessian matrix $B_T(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} l_T(\boldsymbol{\theta}, \mathbf{Y})$. We scale the Hessian with known nonrandom $k \times k$ diagonal matrices ξ_T to obtain its scaled version $W_T(\boldsymbol{\theta}) = \xi_T^{-1} B_T(\boldsymbol{\theta}) \xi_T^{-1}$. We assume that the scaled Hessian $W_T(\boldsymbol{\theta}_0)$ converges weakly to a positive definite (possibly) random matrix $\mathcal{I}(\boldsymbol{\theta}_0)$. Thus, the matrices ξ_T define the rate of convergence of the Hessian matrix $B_T(\boldsymbol{\theta})$ at the true parameter value. They are similarly assumed to define the rate of (weak) convergence of the score function $S_T(\boldsymbol{\theta})$ and, furthermore, that of the MLE $\widehat{\boldsymbol{\theta}}_T$. The needed regularity conditions are presented in Condition 7 in Appendix A. These conditions are typically imposed to ensure the consistency and asymptotic (mixed) normality of a local maximizer of the conditional likelihood function. [2]

Condition 7 and Proposition 8 in Appendix A yield that $\xi_T(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$ is asymptotically mixed normal. Specifically, we have $\xi_T(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{W}{\to} \mathcal{I}(\boldsymbol{\theta}_0)^{-1/2} Z$ where Z $(k \times 1)$ is a standard normal random vector independent of $\mathcal{I}(\boldsymbol{\theta}_0)$. In standard cases, where conventional limit theorems apply, $\xi_T = \sqrt{T}\mathbf{I}_k$ and $\mathcal{I}(\boldsymbol{\theta}_0)$ is nonrandom so that $\sqrt{T}(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \overset{W}{\to} N(0, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$. In the non-ergodic case, that includes co-integrated VAR models with possibly deterministic trends and number of unit roots correctly specified, the matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ is random and the diagonal elements of the matrix ξ_T are powers of $\sqrt{T}$, a typical example being $\xi_T = diag\left[T^{1/2} I_{k_1} : T I_{k_2} : T^{3/2} I_{k_3}\right]$ (see Johansen (1995)). As will be seen below, we rule out the case where the diagonal elements of ξ_T increase at a rate slower than $\sqrt{T}$.

[2] In a similar context, Sweeting (1980) and Basawa and Scott (1983) consider the more general case where the matrices ξ_T need not be diagonal and can belong to a class of continuous functions of the parameter $\boldsymbol{\theta}$. All of our results can be extended to this case but, for simplicity, we do not pursue this extension.

Condition 7 in Appendix A that suffices for the aforementioned results can be verified in various contexts. In the standard stationary case the conditions given in Francq and Zakoian (2010) for multivariate GARCH models are sufficient and the same is true for the multivariate mixture distribution based ACR model of Bec et al. (2008). Regarding non-ergodic models, the needed conditions can be verified in co-integrated VAR models by using results in Johansen (1995).

Condition 4 below allows test statistics to be any continuously differentiable transformation of the multivariate or joint quantile residuals with zero expectation. A large number of different hypotheses are, therefore, testable within this framework.

Condition 4 *Let one of the following assumptions hold.*

(1a) $g : \mathbb{R}^{nm} \rightarrow \mathbb{R}^{w}$ *is a continuously differentiable function such that* $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})) = \mathbf{0}$, *where* $\mathbf{U}_{t,\boldsymbol{\theta}_0} = \begin{bmatrix} \mathbf{R}'_{t,\boldsymbol{\theta}_0} & \cdots & \mathbf{R}'_{t-m+1,\boldsymbol{\theta}_0} \end{bmatrix}' \in \mathbb{R}^{nm}$ *is a vector of quantile residuals defined in (4).*

(1b) $g : \mathbb{R}^{m} \rightarrow \mathbb{R}^{w}$ *is a continuously differentiable function such that* $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})) = \mathbf{0}$, *where* $\mathbf{U}_{t,\boldsymbol{\theta}_0} = \begin{bmatrix} Q_{t,\boldsymbol{\theta}_0} & \cdots & Q_{t-m+1,\boldsymbol{\theta}_0} \end{bmatrix}' \in \mathbb{R}^{m}$ *is a vector of joint quantile residuals defined in (5).*

Before we can state the theorem from which the limiting distributions of our test statistics are obtained, we need further notation and conditions. Again, we only describe the main features here and provide details and supplementary discussion in Appendix A where Condition 9 presents the needed regularity conditions. These conditions mainly concern moments of the function $g(\mathbf{U}_{t,\boldsymbol{\theta}_0})$ and its derivatives with $\mathbf{U}_{t,\boldsymbol{\theta}_0}$ as in Condition 4. Thus, we assume that the expectations

$\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ and $\mathbf{H} = \mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0}) g(\mathbf{U}_{t,\boldsymbol{\theta}_0})')$ exist and are finite with $\mathbf{H}$ positive definite. We also need a condition on (weak) convergence of sample cross moments between $g(\mathbf{U}_{t,\boldsymbol{\theta}_0})$ and $\frac{\partial}{\partial \boldsymbol{\theta}} l_t(\boldsymbol{\theta}, Y_t)$, and to this end we introduce the (possibly) random matrix $\boldsymbol{\Psi}$. In the standard stationary case it is only required that a (weak) law of large numbers applies to these sample cross moments so that in this case the matrix $\boldsymbol{\Psi}$ is nonrandom and equals $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0}) \frac{\partial}{\partial \boldsymbol{\theta}} l_t(\boldsymbol{\theta}, Y_t))$. Finally, we define a matrix $\mathbf{J}$ which is assumed to be the limit of $\sqrt{T} \boldsymbol{\xi}_T^{-1}$ so that it is known, nonrandom, and diagonal. In particular cases the elements of $\mathbf{J}$ consist of zeros and ones; the elements equal to 0 correspond to the components of $\widehat{\boldsymbol{\theta}}_T$ that converge at a rate faster than $\sqrt{T}$ whereas the elements equal to 1 correspond to the components of $\widehat{\boldsymbol{\theta}}_T$ that converge at the usual rate $\sqrt{T}$. Thus, in standard cases we have $\mathbf{J} = \mathbf{I}_k$.

Theorem 5 *Under Conditions 4 and Conditions 7 and 9 in Appendix A*

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}), \hat{\boldsymbol{\Omega}}_T \right) \xrightarrow{W} \left(\boldsymbol{\Omega}^{1/2} \mathcal{U}, \boldsymbol{\Omega} \right) \tag{6}$$

where $\mathcal{U} \sim N(0, \mathbf{I}_w)$,

$$\boldsymbol{\Omega} = \mathbf{G} \mathbf{J} \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{J}' \mathbf{G}' + \boldsymbol{\Psi} \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{J}' \mathbf{G}' + \mathbf{G} \mathbf{J} \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \boldsymbol{\Psi}' + \mathbf{H}, \tag{7}$$

and $\hat{\boldsymbol{\Omega}}_T$ *is computed by replacing the matrices* $\mathbf{G}$, $\mathcal{I}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Psi}$, *and* $\mathbf{H}$ *in the definition of* $\boldsymbol{\Omega}$ *by* $\hat{\mathbf{G}}_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T})$, $W_T(\widehat{\boldsymbol{\theta}}_T) = \boldsymbol{\xi}_T^{-1} B_T(\widehat{\boldsymbol{\theta}}_T) \boldsymbol{\xi}_T^{-1}$,
$\hat{\boldsymbol{\Psi}}_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_t(\widehat{\boldsymbol{\theta}}_T, Y_t) \right]' \boldsymbol{\xi}_T^{-1}$, *and* $\hat{\mathbf{H}}_T = \frac{1}{T} \sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T})'$.

The proof is in Appendix A.

The assumptions of Theorem 5 imply that the limiting (conditional) covariance matrix $\mathbf{\Omega}$ is positive definite (see Condition 9(2) in Appendix A). The first three terms of the matrix $\mathbf{\Omega}$ are due to the uncertainty caused by parameter estimation. If $\mathbf{G} = \mathbf{0}$, there is (asymptotically) no need to take the uncertainty caused by parameter estimation into account in the test statistics. Then the estimate of $\mathbf{\Omega}$ simplifies in an obvious way, because only the matrix $\mathbf{H}$ needs to be estimated. In particular cases, the matrix $\mathbf{H}$ may even be known, as seen in the next section.

The estimator $\hat{\mathbf{\Omega}}_T$ is convenient for most nonlinear models for which the components of $\mathbf{\Omega}$ are impossible or difficult to obtain analytically. One obtains the numerical value of $\hat{\mathbf{\Omega}}_T$ easily by employing the estimation algorithm, one only needs the knowledge of the estimate $\widehat{\boldsymbol{\theta}}_T$, the scaled Hessian matrix $W_T(\widehat{\boldsymbol{\theta}}_T)$, the log-likelihood function $l_t(\widehat{\boldsymbol{\theta}}_T, \mathbf{y}_t)$, and the derivatives $\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_T})$ and $\frac{\partial}{\partial \boldsymbol{\theta}} l_t(\widehat{\boldsymbol{\theta}}_T, \mathbf{y}_t)$. Lemma 10 in Appendix A provides explicit expressions for the derivatives $\frac{\partial}{\partial \boldsymbol{\theta}'} \mathbf{R}_{t,\boldsymbol{\theta}}$ and $\frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}}$ needed to form the estimate $\hat{\mathbf{G}}_T$. All needed derivatives are easy to compute numerically if their analytic values are unknown or difficult to obtain.

The size properties of our tests were occasionally unsatisfactory in models based on mixture distributions, when we used the estimator $\hat{\mathbf{\Omega}}_T$ in simulations. We, therefore, employed the following simulation method: 1) We estimated the parameters of the model; 2) we used these estimates to simulate a data set of 20000 observations from the model; 3) based on this large sample, we computed quantile residuals and numerical derivatives for both the log-likelihood function and quantile residuals; and 4) we used these quantities to compute an estimate of the covariance matrix $\mathbf{\Omega}$. Henceforth, we denote this estimate by $\tilde{\mathbf{\Omega}}_T$.[3] We then

[3] Specifically, the estimate $\tilde{\mathbf{\Omega}}_T$ is defined as $\tilde{\mathbf{\Omega}}_T = \tilde{\mathbf{G}}_T \mathbf{J} \tilde{W}_T(\widehat{\boldsymbol{\theta}}_T)^{-1} \mathbf{J}' \tilde{\mathbf{G}}_T' + \tilde{\mathbf{\Psi}}_T \tilde{W}_T(\widehat{\boldsymbol{\theta}}_T)^{-1} \mathbf{J}' \tilde{\mathbf{G}}_T' + \tilde{\mathbf{G}}_T \mathbf{J} \tilde{\mathbf{W}}_{\mathbf{T}}(\widehat{\boldsymbol{\theta}}_{\mathbf{T}})^{-\mathbf{1}} \tilde{\mathbf{\Psi}}_T' + \tilde{\mathbf{H}}_T$, where $\tilde{\mathbf{G}}_T$, $\tilde{W}_T(\widehat{\boldsymbol{\theta}}_T)^{-1}$, $\tilde{\mathbf{\Psi}}_T$, and $\tilde{\mathbf{H}}_T$ are as in Theorem 5 except that

used the quantile residuals of the original data and the estimate $\tilde{\boldsymbol{\Omega}}_T$ to compute values of the test statistics to be introduced next. This procedure is easy to use in practice. It adds little to the programming task, because one needs to write a code that simulates data from the estimated model. As far as computing time is concerned, the effect of using $\tilde{\boldsymbol{\Omega}}_T$ is insignificant.

Based on the results of Theorem 5, we can deduce that $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T})' \cdot \hat{\boldsymbol{\Omega}}_T^{-1} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) \xrightarrow{W} \mathcal{U}'\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{1/2}\mathcal{U} = \mathcal{U}'\mathcal{U}$. This yields a *general test statistic*

$$S = \frac{1}{T-m+1} \sum_{t=m}^{T} g(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_T})' \cdot \hat{\boldsymbol{\Omega}}_T^{-1} \cdot \sum_{t=m}^{T} g(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_T}) \xrightarrow{W} \chi^2(w), \tag{8}$$

where m and w are the dimensions defined in Condition 4.

A test based on Theorem 5 uses a strategy that requires no specification of an alternative hypothesis. Cox and Hinkley (1974) introduced tests of this type and called them pure significance tests. Such tests are robust, but generally lack optimality against particular alternatives. However, multivariate quantile residual based tests can possess LM test interpretations against particular alternatives, also. This is the case for the three tests to be derived in the next section. From this LM interpretation and results of Chesher and Smith (1997) it further follows that the autocorrelation test has also a LR test interpretation. For the tests based on joint quantile residuals we make no attempt to provide LM interpretations, because of certain difficulties in the formulation of the needed augmented model.

We determine the LM interpretations in the same way as Kalliovirta (2006) in the univariate case. We choose a suitable auxiliary model for quantile residuals

they are computed using the simulated data set and the estimate $\widehat{\boldsymbol{\theta}}_T$ based on the original data.

and incorporate it into the model of interest to obtain an extended likelihood function. We then obtain the test by using the LM principle to an appropriate null hypothesis in the auxiliary model (for details, see Appendix B). We note, however, that the above-mentioned auxiliary model is not suggested for use in practice. It is only a device to obtain a test and understand its properties.

As already mentioned, an advantage of the LM interpretation is that the obtained test is asymptotically optimal against local alternatives (see Basawa and Scott (1983)). In contrast, no similar results exist for uniformly distributed quantile residuals, because then the likelihood function of the resulting auxiliary model is not continuous, and hence, not regular enough.

3 Tests based on Quantile Residuals

Our general framework can be used to derive tests based on a continuously differentiable function of multivariate or joint quantile residuals. This includes, but is unlimited to, higher moments of multivariate or joint quantile residuals. We exemplify this by deriving separate misspecification tests for serial correlation, conditional heteroscedasticity, and non-normality of multivariate and joint quantile residuals.[4] We suggest that one should use these tests jointly. As our tests check for both normality and independence of quantile residuals non-rejections in all of them can be considered as convincing evidence that the model is adequate. Instead of these separate tests, we could have chosen to generalize the approach in Jarque and Bera (1980), and use our framework to derive a joint test for these

[4] If the uncertainty caused by parameter estimation has effect only in small samples, these three test statistics are asymptotically independent. However, dynamic models or even conventional models with normal errors rarely meet this property. On the contrary, in most cases the effect maintains asymptotically.

three features. However, because the sensitiveness of the individual tests against different misspecifications varies, outcomes of separate tests can give useful hints of the reasons of a potential misspecification. In addition, separate tests complement the information provided by graphical methods such as histograms, Q-Q plots, autocorrelation and cross correlation functions of quantile residuals and squared quantile residuals. We illustrate this by providing confidence bounds for graphs of autocorrelation functions and, thereby, justify their use.

In this section we assume a correct model specification, thus $\mathbf{R}_{t,\boldsymbol{\theta}_0} \sim n.i.d.(0, I_n)$ and $Q_{t,\boldsymbol{\theta}_0} \sim n.i.d.(0,1)$.

3.1 Test for Autocorrelation

For our autocorrelation test based on multivariate quantile residuals we introduce the general null hypothesis $H_0 : \mathbb{E}(\mathbf{R}_{t,\boldsymbol{\theta}_0}\mathbf{R}'_{t-s,\boldsymbol{\theta}_0}) = 0$ for all t and $s > 0$. The test is based on the statistics $\hat{\mathbf{C}}_s = \frac{1}{T-s}\sum_{t=1+s}^{T} \mathbf{r}_{t,\widehat{\boldsymbol{\theta}}_T}\mathbf{r}'_{t-s,\widehat{\boldsymbol{\theta}}_T}$, $s = 1, ..., K_1$. Thus, we assume that the first K_1 autocovariance matrices reflect the potential inadequacy of the model. The use of uncentered sample autocovariance matrices is reasonable here, because theoretically $\mathbb{E}(\mathbf{R}_{t,\boldsymbol{\theta}_0}) = 0$, even though in general $\bar{\mathbf{r}}_{\widehat{\boldsymbol{\theta}}_T} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{r}_{t,\widehat{\boldsymbol{\theta}}_T} \neq \mathbf{0}$. Also, Chitturi (1974) used a similar test statistic based on autocorrelations of traditional Pearson's residuals in a VAR model.

To apply our general approach we define the (continuously differentiable) function $g : \mathbb{R}^{n(K_1+1)} \rightarrow \mathbb{R}^{n^2K_1}$ as

$$g(\mathbf{u}_{t,\boldsymbol{\theta}}) = vec\begin{bmatrix}\mathbf{r}_{t,\boldsymbol{\theta}}\mathbf{r}'_{t-1,\boldsymbol{\theta}} & \cdots & \mathbf{r}_{t,\boldsymbol{\theta}}\mathbf{r}'_{t-K_1,\boldsymbol{\theta}}\end{bmatrix}, \tag{9}$$

where vec denotes the columnwise vectorization of a matrix. Then clearly $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$

$= \mathbf{0}$. Of the matrices $\mathbf{H} = \mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})g(\mathbf{U}_{t,\boldsymbol{\theta}_0})')$ and $\mathbf{G} =\mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'}g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ in the covariance matrix $\boldsymbol{\Omega}$ (see Theorem 5), the former is equal to $\mathbf{I}_{n^2 K_1}$ by properties of the standard multinormal distribution whereas the latter is given in Derivatives section in Appendix A. These analytic computations are unnecessary, when one employs the numerically obtained value $\hat{\boldsymbol{\Omega}}_T$ (or $\tilde{\boldsymbol{\Omega}}_T$), however.

Theorem 5 yields the test statistic A_{K_1} computed as the general test statistic (8) with $m = K_1 + 1$, $w = n^2 K_1$, and $g(\mathbf{u}_{t,\hat{\boldsymbol{\theta}}_T})$ defined by (9). We also use the function defined in (9) to obtain the estimate $\hat{\boldsymbol{\Omega}}_T$ (or alternatively $\tilde{\boldsymbol{\Omega}}_T$). One can interpret this test statistic as a LM test when a K_1th order autoregressive model is specified for quantile residuals. Appendix B provides details on this.

In addition to the overall test statistic A_{K_1}, it is always useful to consider individual autocovariance and cross covariance estimates $\hat{c}_{ijs}$. A large value of $\hat{c}_{ijs}$ compared to its approximate standard error, obtained from the relevant diagonal element of the matrix $T^{-1}\hat{\boldsymbol{\Omega}}_T$ (or $T^{-1}\tilde{\boldsymbol{\Omega}}_T$), suggests model inadequacy. Therefore, a useful model criticism procedure is to plot $\hat{c}_{ij1}, ..., \hat{c}_{ijr}$, divided by their standard errors for each j and some r, and compare them with their approximate 95% critical bounds, as already suggested in McLeod (1978). For each j this procedure corresponds to performing r individual tests, and, therefore, the resulting joint significance level lies between the maximum p-value of the individual tests and their sum.

Now consider obtaining an autocorrelation test based on joint quantile residuals. The relevant null hypothesis is $H_0 : \mathbb{E}(Q_{t,\boldsymbol{\theta}_0} Q_{t-s,\boldsymbol{\theta}_0}) = 0$ for all t and $s > 0$. The test is obtained by applying the preceding autocorrelation test in univariate form. Thus, $\mathbf{r}_{t,\boldsymbol{\theta}}, ..., \mathbf{r}_{t-K_1,\boldsymbol{\theta}}$ are replaced with $q_{t,\boldsymbol{\theta}}, ..., q_{t-K_1,\boldsymbol{\theta}}$ in the function $g(\cdot)$, and appropriate changes are made in the matrices $\mathbf{G}$, $\boldsymbol{\Psi}$, and $\mathbf{H}$ in the covariance matrix $\mathbf{\Omega}$ (see Theorem 5). We denote the resulting test statistic by $A^J_{K_1}$.

3.2 Test for Conditional Heteroscedasticity

Again, we first use multivariate quantile residuals to obtain a test of potential conditional heteroscedasticity. As usual, we relate conditional heteroscedasticity to correlation of squares and consider the general null hypothesis H_0 : $\mathbb{E}(R^2_{it,\boldsymbol{\theta}_0}, R^2_{j,t-s,\boldsymbol{\theta}_0}) = 0$ for $i,j \in \{1,...,n\}$, all t, and $s > 0$. This is a natural generalization of the hypothesis used in the corresponding univariate test in Kalliovirta (2006). As in that paper, we modify the ideas suggested in McLeod and Li (1983) and Ling and Li (1997) and base the test on the autocovariance type statistics $\hat{d}_{ijs} = \frac{1}{T-s}\sum_{t=1+s}^{T}\left(r^2_{it,\widehat{\boldsymbol{\theta}}_T} - 1\right)\left(r^2_{j,t-s,\widehat{\boldsymbol{\theta}}_T} - 1\right)$ $i,j \in \{1,...,n\}$, $s = 1,...,K_2$. We assume that a relatively small number of these statistics reflect sufficiently the potential inadequacy of the model. Appendix B shows that one can motivate the resulting test as a LM test when multivariate quantile residuals follow a K_2th order multivariate ARCH model.

Let $\mathbf{U}_{t,\boldsymbol{\theta}} = \begin{bmatrix}\mathbf{R}'_{t,\boldsymbol{\theta}} & \cdots & \mathbf{R}'_{t-K_2,\boldsymbol{\theta}}\end{bmatrix}'$, and, according to the preceding discussion, define the function $g : \mathbb{R}^{n(K_2+1)} \to \mathbb{R}^{n^2K_2}$ as

$$g(\mathbf{u}_{t,\boldsymbol{\theta}}) = vec\begin{bmatrix}\mathbf{v}_{t,\boldsymbol{\theta}}\mathbf{v}'_{t-1,\boldsymbol{\theta}} & \cdots & \mathbf{v}_{t,\boldsymbol{\theta}}\mathbf{v}'_{t-K_2,\boldsymbol{\theta}}\end{bmatrix} \tag{10}$$

with $\mathbf{v}_{t-s,\boldsymbol{\theta}} = \begin{bmatrix}r^2_{1,t-s,\boldsymbol{\theta}} - 1 & \cdots & r^2_{n,t-s,\boldsymbol{\theta}} - 1\end{bmatrix}'$, $s = 0,1,...,K_2$. Then $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})) = \mathbf{0}$, and Derivatives section in Appendix A gives the matrix $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ of Condition 9. Properties of the standard multinormal distribution give that the matrix $\mathbf{H} = \mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})g(\mathbf{U}_{t,\boldsymbol{\theta}_0})')$ in the covariance matrix $\mathbf{\Omega}$ (see Theorem 5) is equal to $\mathbf{I}_n \otimes 4\mathbf{I}_{K_2}$, where $\otimes$ denotes the Kronecker product. Theorem 5 yields the test statistic H_{K_2} computed as the general test statistic (8) with $m = K_2 + 1$, $w = n^2 K_2$, and $g(\mathbf{u}_{t,\hat{\boldsymbol{\theta}}_T})$ and, further, $\hat{\boldsymbol{\Omega}}_T$ (or $\tilde{\boldsymbol{\Omega}}_T$) defined by (10).

Similarly to the autocorrelation test, it is useful to supplement the overall test statistic H_{K_2} by plotting individual estimates $\widehat{d}_{ijs}$ divided by their approximate standard errors obtained from the diagonal elements of the matrix $T^{-1}\hat{\boldsymbol{\Omega}}_T$ or $(T^{-1}\tilde{\boldsymbol{\Omega}}_T)$.

A heteroscedasticity test can be based on joint quantile residuals, also. Then one tests the null hypothesis $H_0 : \mathbb{E}(Q^2_{t,\boldsymbol{\theta}_0}, Q^2_{t-s,\boldsymbol{\theta}_0}) = 0$ for all t and $s > 0$ and modifies the preceding test statistic to obtain a test statistic $H^J_{K_2}$ in the following way: Choose $\mathbf{U}_{t,\boldsymbol{\theta}} = \begin{bmatrix} Q_{t,\boldsymbol{\theta}} & \cdots & Q_{t-K_2,\boldsymbol{\theta}} \end{bmatrix}'$ so that $\mathbf{r}_{t,\boldsymbol{\theta}}, ..., \mathbf{r}_{t-K_2,\boldsymbol{\theta}}$ are replaced with $q_{t,\boldsymbol{\theta}}, ..., q_{t-K_2,\boldsymbol{\theta}}$ in the function $g(\cdot)$, and make appropriate changes in the matrices $\mathbf{G}$, $\boldsymbol{\Psi}$, and $\mathbf{H}$ in the covariance matrix Ω (see Theorem 5).

3.3 Normality tests

Our multinormality tests use ideas in Lomnicki (1961), Bera and John (1983), Jarque and Bera (1987), and Doornik and Hansen (2008). First we choose $\mathbf{U}_{t,\boldsymbol{\theta}} = \mathbf{R}_{t,\boldsymbol{\theta}}$ in Condition 4, and develop a multinormality test for multivariate quantile residuals. The null hypothesis we employ uses three moments of multivariate quantile residuals, that is, $H_0 : \mathbb{E}\begin{bmatrix} R^2_{jt,\boldsymbol{\theta}_0} - 1 & R^3_{jt,\boldsymbol{\theta}_0} & R^4_{jt,\boldsymbol{\theta}_0} - 3 \end{bmatrix} = 0$ for all $j \in \{1, ..., n\}$ and t. This hypothesis is true if $R_{jt,\boldsymbol{\theta}_0} \sim n.i.d.(0,1)$. The independence structure of theoretical quantile residuals within and between observations allows us to test multinormality in a similar manner as in Doornik and Hansen (2008).

Now define the function $g : \mathbb{R}^n \to \mathbb{R}^{3n}$ as

$$g(\mathbf{u}_{t,\boldsymbol{\theta}}) = \begin{bmatrix} g_1(r_{1t,\boldsymbol{\theta}})' & \cdots & g_n(r_{nt,\boldsymbol{\theta}})' \end{bmatrix}', \tag{11}$$

where $g_j(r_{jt,\boldsymbol{\theta}}) = \begin{bmatrix} r_{jt,\boldsymbol{\theta}}^2 - 1 & r_{jt,\boldsymbol{\theta}}^3 & r_{jt,\boldsymbol{\theta}}^4 - 3 \end{bmatrix}'$.[5] Properties of the standard multinormal distribution give $\mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})) = \mathbf{0}$. The matrix $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ in the covariance matrix $\boldsymbol{\Omega}$ (see Theorem 5) is given in Derivatives section in Appendix A, whereas

$$\mathbf{H} = \mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0}) g(\mathbf{U}_{t,\boldsymbol{\theta}_0})') = \mathbf{I}_n \otimes \begin{bmatrix} 2 & 0 & 12 \\ 0 & 15 & 0 \\ 12 & 0 & 96 \end{bmatrix}. \tag{12}$$

Theorem 5 yields the test statistic N computed as the general test statistic (8) with $m = 1$, $w = 3n$, and $g(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_T})$ and, further, $\hat{\boldsymbol{\Omega}}_T$ (or $\tilde{\boldsymbol{\Omega}}_T$) defined by (11).

Arguments similar to those in Bera and John (1983) show that test statistic N can be motivated by the LM principle (find details in Appendix B).

We test the normality of the joint quantile residuals by choosing the null hypothesis $H_0 : \mathbb{E}\begin{bmatrix} Q_{t,\boldsymbol{\theta}_0}^2 - 1 & Q_{t,\boldsymbol{\theta}_0}^3 & Q_{t,\boldsymbol{\theta}_0}^4 - 3 \end{bmatrix} = 0$ for all t. This hypothesis is true if $Q_{t,\boldsymbol{\theta}_0} \sim n.i.d.(0,1)$. Therefore, we construct a univariate form of the normality test obtained above and denote the resulting test statistic by N^J.. Thus, we set $\mathbf{U}_{t,\boldsymbol{\theta}} = Q_{t,\boldsymbol{\theta}}$ in $g(\mathbf{u}_{t,\widehat{\boldsymbol{\theta}}_T})$, and we define accordingly the matrices $\mathbf{G}$, $\boldsymbol{\Psi}$, and $\mathbf{H}$ of Condition 9.

[5] Compared to earlier normality tests based on Pearson's residuals, we include the term $r_{t,\boldsymbol{\theta}}^2 - 1$. We have found that the addition of this term improves small sample properties of the test for nonlinear models. It has to be removed, however, if the variance of quantile residuals of the estimated model is automatically one. Otherwise, the matrix defined in (12) is not positive definite and the asymptotic result does not hold. This happens, for example, when models can be estimated using ordinary least squares (OLS).

4 Empirical example

This section analyzes properties of exchange rate series by applying multivariate GARCH models, multivariate and joint quantile residuals, and tests based on them. These considered models employ mixture distributions and, therefore, the interpretation of traditional residuals is unreliable.

4.1 The Model

The Multivariate Generalized Orthogonal Factor GARCH model uses generalized orthogonal factors to solve some typical problems encountered in multivariate GARCH models. The aim is to determine a relatively small number of factors that describe the multivariate conditional variance structure of the data adequately. Compared to the model proposed by Lanne and Saikkonen (2007), we consider a slightly generalized version.

Let $\mathbf{y}_t$ be a n dimensional process with a conditional density function of the form

$$
\begin{aligned}
f_{t-1}(\mathbf{y}_t) = {} & p(2\pi)^{-\frac{n}{2}} \det(\mathbf{W}\mathbf{H}_{1t}\mathbf{\Phi}_1^{-1}\mathbf{W}')^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mathbf{y}_t' \left(\mathbf{W}\mathbf{H}_{1t}\mathbf{\Phi}_1^{-1}\mathbf{W}'\right)^{-1} \mathbf{y}_t\} \\
& +(1-p)(2\pi)^{-\frac{n}{2}} \det(\mathbf{W}\mathbf{H}_{2t}\mathbf{\Phi}_2^{-1}\mathbf{W}')^{-\frac{1}{2}} \exp\{-\frac{1}{2}\mathbf{y}_t' \left(\mathbf{W}\mathbf{H}_{2t}\mathbf{\Phi}_2^{-1}\mathbf{W}'\right)^{-1} \mathbf{y}_t\},
\end{aligned}
\quad (13)
$$

where $p \in (0,1)$, $\mathbf{W}$ $(n \times n)$, $\mathbf{\Phi}_1 = p\mathbf{I}_n + (1-p)\mathbf{\Phi}$, $\mathbf{\Phi}_2 = \mathbf{\Phi}_1\mathbf{\Phi}^{-1}$, and $\mathbf{\Phi} = diag\,[\phi_1 \cdots \phi_n]$ are parameter matrices with $\mathbf{W}$ nonsingular, and $\mathbf{H}_{1t}$ and $\mathbf{H}_{2t}$ $(n \times n)$ are stochastic diagonal matrices defined below. We assume that the matrix $\mathbf{\Phi}$ has positive diagonal elements; $\phi_i > 0$ for all $i \in \{1, ..., n\}$. The stochastic diagonal matrices $\mathbf{H}_{1t}$ and $\mathbf{H}_{2t}$ describe conditional heteroscedasticity in the process $\mathbf{y}_t$.

They are of the form $\mathbf{H}_{jt} = diag[\mathbf{V}_{jt} : \mathbf{I}_{n-r}]$ with $\mathbf{V}_{jt} = diag\left[v_{1t}^{(j)} \cdots v_{rt}^{(j)}\right]$, where

$$v_{it}^{(j)} = (1 - \alpha_{ji} - \beta_{ji}) + \beta_{ji} v_{i,t-1}^{(j)} + \alpha_{ji}(\mathbf{b}_i' \mathbf{y}_{t-1})^2, \quad i = 1, ..., r, \text{ and } j = 1, 2, \quad (14)$$

and $\mathbf{b}_i'$ is the ith row of the parameter matrix $\mathbf{B}' = \mathbf{W}^{-1}$. Thus, each of the processes $v_{it}^{(j)}$ is a conventional (univariate) GARCH(1,1) process except that we normalize the intercept terms so that the components of $\mathbf{B}'\mathbf{y}_t$ have unit unconditional variance. As in GARCH(1,1) models, the parameters $\alpha_{1i}, \alpha_{2i}, \beta_{1i}, \beta_{2i}$ in (14) are assumed to satisfy $\alpha_{ji} > 0$, $\beta_{ji} \geq 0$, and $\alpha_{ji} + \beta_{ji} < 1$ for all i and j. Thus, the conditional distribution of $\mathbf{y}_t$ is a mixture of two normal distributions with $\mathbb{E}_{t-1}(\mathbf{y}_t) = 0$ and $cov_{t-1}(\mathbf{y}_t) = p\mathbf{W}\mathbf{H}_{1t}\mathbf{\Phi}_1^{-1}\mathbf{W}' + (1-p)\mathbf{W}\mathbf{H}_{2t}\mathbf{\Phi}_2^{-1}\mathbf{W}'$. The model is identified up to multiplying the columns of $\mathbf{B}$ by minus one.

The model has an alternative representation as a function of parameters and two unobservable random variables

$$\mathbf{y}_t = \mathbf{W}\left(I(s_t = 0) \cdot \mathbf{H}_{1t}^{1/2}\mathbf{\Phi}_1^{-1/2} + I(s_t = 1) \cdot \mathbf{H}_{2t}^{1/2}\mathbf{\Phi}_2^{-1/2}\right)\boldsymbol{\varepsilon}_t, \quad (15)$$

where $I(\cdot)$ is the indicator function, $\boldsymbol{\varepsilon}_t \sim n.i.d.(0, \mathbf{I}_n)$, and s_t is an $i.i.d.$ random variable with $\Pr(s_t = 0) = p$ and $\Pr(s_t = 1) = 1 - p$. Moreover, the processes $\{\boldsymbol{\varepsilon}_t\}$ and $\{s_t\}$ are independent. From (15) one clearly obtains the conditional density (13) for $\mathbf{y}_t$. The representation (15) is easily compared with the model in Lanne and Saikkonen (2007), where

$$\mathbf{y}_t = \mathbf{W}\mathbf{H}_t^{1/2}\left(I(s_t = 0) + I(s_t = 1) \cdot \mathbf{\Phi}^{1/2}\right)\mathbf{\Phi}_1^{-1/2}\boldsymbol{\varepsilon}_t. \quad (16)$$

This model can be obtained from (15) by imposing the parameter restrictions

$\alpha_{1i} = \alpha_{2i}$ and $\beta_{1i} = \beta_{2i}$ for all $i = 1, ..., r$. Thus, in our model the mixture structure is not limited to the distribution of the error term ε_t, it also affects parameters in the conditional covariance matrix.

The restrictions imposed on the parameters in (14), with some further assumptions, imply that the process $\mathbf{y}_t$, defined by (16), is strictly stationary and ergodic and also second order stationary (see Lanne and Saikkonen (2007) and the references therein). To the best of our knowledge, conditions that guarantee similar results in the more general model (15) are unknown. Therefore, we assume that the model is stationary and ergodic under the assumptions made on the parameters. Then standard limit theorems apply, and verification of the high level conditions imposed in Section 2 becomes possible with $\boldsymbol{\xi}_T=\sqrt{T}\mathbf{I}_k$ and $\mathbf{J} = \mathbf{I}_k$. Due to space constraints, no attempt is made to provide details, however.

We analyze 4 weekly exchange rate series of the French Franc (FRF), Dutch Guilder (NLG), German Mark (DEM) and Swiss Franc (CHF) against the U.S. Dollar (USD) for the years 1984–1997. That makes 782 observations. This data was also employed in Lanne and Saikkonen (2007). We used Gauss 10.0 and the algorithm library cml MT-package to compute maximum likelihood estimates of the parameters. We calculated the initial values of $\widehat{\mathbf{H}}_{11}$ and $\widehat{\mathbf{H}}_{21}$ using the sample variances of $\widehat{\mathbf{b}}_j'\mathbf{y}_t$, thus, the initial values differed at each iteration. Before estimation, we centered the series for the mean to be zero.

4.2 Comparison of the estimated models

We use quantile residuals to compare four different Multivariate Generalized Orthogonal Factor GARCH models. Two of them, already estimated in Lanne and

Saikkonen (2007), are a two factor model under normality (Model 1) and a one factor mixture-normal model (Model 2), both based on equation (16). The estimation results of these two models of can be found in their paper. Tables 1 and 2 give the estimation results of a one factor mixture-normal model (Model 3) and a two factor mixture-normal model (Model 4) based on equation (15). We estimated these two new models to discover whether we could remove the inadequacies in the previous models (1 and 2) detected by our analysis. We used two factors in Model 4 because, according to the tests derived in Lanne and Saikkonen (2007), the null hypothesis of two conditionally heteroscedastic factors was not rejected at the 5% significance level.

We computed the quantile residuals according to equations (4) and (5), and chose the conditioning order $(i_1, i_2, i_3, i_4) = (1, 2, 3, 4)$. Hence, our observed multivariate quantile residuals are

$$\mathbf{r}_{t,\widehat{\boldsymbol{\theta}}_T} = \begin{bmatrix} r_{1t,\widehat{\boldsymbol{\theta}}_T} \\ r_{2t,\widehat{\boldsymbol{\theta}}_T} \\ r_{3t,\widehat{\boldsymbol{\theta}}_T} \\ r_{4t,\widehat{\boldsymbol{\theta}}_T} \end{bmatrix} = \begin{bmatrix} \Phi^{-1}(F_{1,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{1t})) \\ \Phi^{-1}(F_{2,1,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{2t})) \\ \Phi^{-1}(F_{3,2,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{3t})) \\ \Phi^{-1}(F_{4,3,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{4t})) \end{bmatrix}.$$

The factorization of the joint density into a product of one marginal and three conditional densities was eased by the fact that for the family of mixtures of multinormal distributions the marginal and conditional distributions belong to the same family of distributions. Thus, one can solve the residuals for each observation iteratively by solving the parameters of one marginal and one conditional distribution at a time. For completeness we provide details on this in Appendix C.

Table 3 shows the values of the test statistics, developed in Section 3, for each model along with the values of two information criteria, AIC and BIC. They are computed as $AIC = 2 \cdot k - 2 \cdot l_T$ and $BIC = k \cdot \log(T-u) - 2 \cdot l_T$, where l_T is the value of the maximized log-likelihood of the sample, k is the dimension of the parameter vector, T is the sample size, and u is the number of needed initial values. Similarly to Kalliovirta (2006), we compute the values of the tests statistics in Table 3 with a simulated covariance matrix $\tilde{\mathbf{\Omega}}_T$ (for definition, see Section 2) because, according to simulations in Section 5, they provide more reliable versions of the tests.[6]

Table 3 shows that the autocorrelation test A_3^J based on three lags of joint quantile residuals is not critical on any of the models. One observes the same by looking at autocorrelation functions based on the joint quantile residuals (reported only for Model 4 in Figure 1). The normality test based on joint quantile residuals N^J rejects all models at 1% significance level. This is also the case for the conditional heteroscedasticity test H_3^J based on three lags of joint quantile residuals with the exception of Model 4, and the $p-$value is as small as 1.1% even for this model. Figure 1, that depicts both the autocorrelation graphs of the joint quantile residuals and squared joint quantile residuals of Model 4 along with 99% critical bounds[7], illustrates this further. Tests based on multivariate quantile residuals reject the models at conventional critical levels. Overall the tests are least critical towards Model 4 that is also favoured by the information criteria.

The third multivariate quantile residual $r_{3t,\widehat{\boldsymbol{\theta}}_T}$ is negatively autocorrelated at

[6] The Gauss code to implement our tests is available from the author upon request.

[7] Section 3 explains how these 99% critical bounds are derived. Because we actually test several tests at the same time, we should make the Bonferroni correction. If we use 99% confindence bounds for 5 tests at the same time, we are, according to the correction, actually basing our inference on 95% confidence bounds.

lag one with the absolute value around 0.20 in all four models. Figure 1 depicts this for Model 4 along with the 99% critical bounds. This figure explains why the autocorrelation test A_3 rejects all the models, and indicates that the mean might be time-varying. We ignore this problem, but acknowledge that it can cause bias in our analysis. The squared multivariate quantile residuals are autocorrelated especially in Model 1. The autocorrelation is smaller in mixture distribution based models. But even for Model 4, autocorrelation exists in the series of $r^2_{4t,\widehat{\boldsymbol{\theta}}_T}$ and, therefore, H_3 rejects (see Figure 1).

The multivariate quantile residual series of Models 1 and 4, depicted in Figures 2 and 3, show that Model 4 captures the fluctuations of the data much better than Model 1. The same is true when Model 4 is compared with Models 2 and 3 (the graphs not shown). An inspection of the distributional fit by other methods, like histograms and normal probability plots based on multivariate quantile residuals (not reported), favour Models 3 and 4.

To conclude, the tests and figures are more informative than previously available AIC and BIC. The graphs based on multivariate quantile residuals indicate that the mixture models provide better descriptions for the exchange rate series than the normal distribution based Model 1. A further advantage of the graphs is that they suggest possible reasons of misspecification. We wish to emphasize, however, that our aim has been to illustrate how different models with non-nested structures can be analyzed with the methods proposed in this paper. Therefore, it is beyond the scope of this work to consider new specifications even though the diagnostics accepted none of the examined models.

5 Simulations

This section studies the size and power properties of the proposed tests based on joint and multivariate quantile residuals ($A^J_{K_1}$, $H^J_{K_2}$, N^J, A_{K_1}, H_{K_2}, and N). Our simulations consider the sample sizes 250, 500, 750, and 1000, depending on the model to be estimated. All results are based on 2000 replications. We report empirical rejection frequencies when one considers tests at 5% and 1% significance levels. To avoid the initial value problem, 200 extra observations were simulated and removed from the beginning of every sample. We obtained the MLEs of the parameters of the considered models using the cml MT library in GAUSS Windows Version 10.0. We used the inverse of the cross-product of the first derivatives to compute the approximate covariance matrix of estimators. This procedure guarantees positive semidefinite estimates.

As already mentioned, the size properties of the tests were occasionally unsatisfactory when we used the covariance matrix estimator $\hat{\mathbf{\Omega}}_T$ in models based on mixture distributions. We, therefore, employed the estimate $\tilde{\mathbf{\Omega}}_T$ to compute values of the test statistics.

5.1 Models

We study the size properties and the ability of the considered tests to reveal misspecification with simple bivariate models. In power comparisons we do not adjust the tests for size distortions, because the sizes are quite accurate and the adjustment is impossible to do in empirical applications. Thus, our simulation study conforms to actual testing situation.

We use Models S.1, S.3, and S.5 to examine the size properties and Models S.1, S.2, and S.4 to study power. Results are in Tables 4, 5, 6, and 7.

Model S.1

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \text{ where } \boldsymbol{\varepsilon}_t \sim n.i.d.(\mathbf{0}, \boldsymbol{\Sigma}), \text{ and } \boldsymbol{\Sigma} > \mathbf{0}.$$

Model S.2:

$$\mathbf{y}_t = (\boldsymbol{\mu}_1 + \boldsymbol{\varepsilon}_t) \cdot I(\eta_t \leq c) + (\boldsymbol{\mu}_2 + \boldsymbol{\epsilon}_t) \cdot I(\eta_t > c),$$

where $\eta_t \sim N(0,1)$, $\boldsymbol{\varepsilon}_t \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_1)$, and $\boldsymbol{\epsilon}_t \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_2)$ are mutually independent unobservable i.i.d. random variables with $\boldsymbol{\Sigma}_1 > 0$ and $\boldsymbol{\Sigma}_2 > 0$.

Model S.3:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \text{ where } \boldsymbol{\varepsilon}_t \sim n.i.d.(\mathbf{0}, \boldsymbol{\Sigma}) \text{ and } \boldsymbol{\Sigma} > \mathbf{0}.$$

Model S.4:

$$\mathbf{y}_t = \boldsymbol{\mu} + I(\eta_t \leq c) \cdot \mathbf{A}_1\mathbf{y}_{t-1} + I(\eta_t > c) \cdot \mathbf{A}_2\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where $\eta_t \sim n.i.d.(0,1)$ and $\boldsymbol{\varepsilon}_t \sim n.i.d.(\mathbf{0}, \boldsymbol{\Sigma})$ are mutually independent with $\boldsymbol{\Sigma} > \mathbf{0}$.

Model S.5:

$$\mathbf{y}_t = \boldsymbol{\Sigma}_t^{\frac{1}{2}} \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim n.i.d.(\mathbf{0}, \boldsymbol{\Sigma}_t)$ with $\boldsymbol{\Sigma}_t = \mathbf{W}\mathbf{H}_t\mathbf{W}' > 0$ for all t, $\mathbf{H}_t = diag \begin{bmatrix} v_t & 1 \end{bmatrix}$, and $v_t = (1-\alpha) + \alpha(\mathbf{b}_1'\mathbf{y}_{t-1})^2$ with $\mathbf{b}_1$being the first row of matrix $(\mathbf{W}')^{-1}$.

5.2 Size and power properties

The size properties of the tests are satisfactory, though the conditional heteroscedasticity test based on multivariate quantile residuals (H_3) is somewhat oversized (Tables 4, 5, and 6). For sample sizes larger than 1000, however, the size properties are quite accurate even for this test (not reported).

We study the power properties of the tests by simulating Models S.2 and S.5, and then estimating Model S.1 (Table 4). Model S.2 is a mixture of two normal distributions with small differences in the means and covariance structures based on i.i.d. innovations. Therefore, it is unsurprising that the tests show only little power. One also expects the normality tests to react, and indeed the multivariate quantile residual based version N has power in larger samples. When the difference in means increases, the tests become powerful in small samples as well (not reported). When Model S.5 generates the data, one expects that especially the conditional heteroscedasticity tests H_3 and H_3^J have power. This is the case for the multivariate quantile residual based test (H_3) even though the value of the parameter α is relatively small (Table 4). However, the corresponding test based on joint quantile residuals (H_3^J) shows only little power and is even less powerful than the normality test based on multivariate quantile residuals (N). When larger values of α are used, both of the two conditional heteroscedasticity tests are very powerful, as expected. The autocorrelation tests A_3 and A_3^J are at their nominal levels. This is unsurprising, because Models S.2 and S.5 contain no time varying conditional mean (Table 4).

The conditional heteroscedasticity test based on multivariate quantile residuals (H_3) shows power, when Model S.4 is simulated and Model S.3 is estimated (Table

5). The corresponding test based on joint quantile residuals (H_3^J) also has fairly good power compared with the other joint quantile residual based tests. The other tests have low power except the autocorrelation test based on multivariate quantile residuals (A_3). However, even the power of this test is rather modest and decreases when the sample size increases. The cause for this may be that, compared to Models S.3, the additional regime is conditionally heteroscedastic, but not autocorrelated in Model S.4.

To conclude, according to our simulations the tests based on multivariate quantile residuals are more powerful than their counterparts based on joint quantile residuals. However, despite their relatively low power, tests based on joint quantile residuals may nevertheless be useful especially when both the dimension and the number of the observations are large.

5.3 Comparisons to other tests

Kalliovirta (2006) found in the univariate case that, when a mixture model is simulated and estimated, tests based on Pearson's residuals[8] are unreliable. Table 7 illustrates this in the multivariate case using mixture distribution based Model S.2 and the normality test BJ^p of Bera and John (1983). The size properties of Pearson's residual based test are unacceptable: the true null hypothesis is rejected with 100% certainty. If we employ the normality test of Doornik and Hansen (2008), the size properties are exactly the same (not reported).

Kalliovirta (2006) also found that previous tests based on moments can suffer from severe size distortions when employed without taking the uncertainty caused

[8] We define Pearson's residual as $\widehat{Cov_{t-1}(\mathbf{y}_t)}^{-\frac{1}{2}} \cdot (\mathbf{y}_t - \widehat{E_{t-1}\mathbf{y}_t})$, where $\widehat{E_{t-1}\mathbf{y}_t}$ is estimated conditional expectation vector and $\widehat{Cov_{t-1}(\mathbf{y}_t)}$ is estimated conditional covariance matrix.

by parameter estimation into account. For this reason, we consider in Model S.2 quantile residual based test N^q, computed using $\mathbf{\Omega} = \mathbf{H}$. This test is heavily undersized (Table 7). When the effect of parameter estimation is taken into account, the size properties are satisfactory. This is displayed in Table 7 with both our multivariate quantile residual based test N and joint quantile residual based test N^J. To compare our tests with existing uniformly distributed quantile residual based test, we also computed the pooled test statistic of Bai and Chen (2008). This test turned out to be oversized. We tried several different parameter values in Model S.2 (not reported), but our findings were the same as in Table 7.

To conclude, the above results demonstrate that with mixture distribution based models it is important to use quantile residual based tests that properly take the effect of parameter estimation into account. Similar results may occur with other models as well. Especially with such models where no analytic results exist to guarantee that the uncertainty caused by parameter estimation vanishes asymptotically.

6 Conclusion

We studied multivariate and joint quantile residuals that are generalizations of traditional residuals. Under regularity conditions, we stated the theoretical properties of quantile residuals, developed a general framework, and used it to obtain misspecification tests based on quantile residuals. Our tests are theoretically sound in that they take the uncertainty caused by parameter estimation into account. We illustrated how to apply our framework by deriving tests for serial correlation, conditional heteroscedasticity, and non-normality in quantile residuals. These tests are simple to compute once the parameters of the model are estimated, and their application only requires the conventional χ^2 criterion.

We enlarged the set of models for which traditional graphical diagnostics and related statistical tests are applicable. Examples of models now included in this set are models involving mixture distributions or latent variables that have recently found applications in econometrics. Our simulations showed that for these models diagnostics based on traditional Pearson's residuals can be unreliable. Our misspecification tests are reliable and applicable for all models for which quantile residuals are suited. This includes models for which also traditional residuals work. Because our testing approach properly takes the uncertainty caused by parameter estimation into account, it can even improve size properties of previous tests which ignore the estimation uncertainty. We demonstrated this by using a normality test and simulation. We illustrated the practical usefulness of our approach by an empirical example that applied mixtures of Multivariate Generalized Orthogonal Factor GARCH models.

A Appendix: Proofs

We assume the usual framework of a parametric model, where $(\Omega, \mathfrak{A}, \mathbb{P})$ is a fixed probability space with a complete measure $\mathbb{P}$, $\mathbf{Y}_{\boldsymbol{\theta}} : \Omega \rightarrow \mathbb{R}^{nT}$ a family of random variables indexed by the parameter $\boldsymbol{\theta}$ belonging to the set $\boldsymbol{\Theta} \subset \mathbb{R}^{k}$, and $(\mathbb{R}^{nT}, \mathfrak{B}^{nT}, \mathbb{P}_{\boldsymbol{\theta}})$ the probability space induced by $\mathbf{Y}_{\boldsymbol{\theta}}$. Then $\mathcal{P} = \{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is a collection of probability measures defined on $\mathfrak{B}^{nT}$, the Borel sigma-algebra of $\mathbb{R}^{nT}$. The collection $\mathcal{P}$ can equally well be defined by the density functions $f(\boldsymbol{\theta}, \mathbf{y})$, $\mathcal{P} = \left\{ f(\boldsymbol{\theta}, \mathbf{y}) : \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \mathbf{y} \in \mathbb{R}^{nT} \right\}$, the definition we use in the main text.

Proof of Lemma 2. Following the proof of Rosenblatt (1952) and the notation in the main text, we write $Z_{jt} = F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, Y_{i_j,t})$ for each $j = 1, ..., n$ and $t = 1, ..., T$. We fix the point $(\mathbf{z}_1, ..., \mathbf{z}_T) \in (0,1)^{nT}$, where $\mathbf{z}_t = (z_{1t}, ..., z_{nt})$. Then for each z_{jt} exists unique $y_{i_j,t}$ such that $z_{jt} = F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, y_{i_j,t})$ for all j and t. This follows from the fact that the distributions $F_{i_j,j-1,t-1}$ are absolutely continuous w.r.t. Lebesgue measure. We denote

$$A = \left\{ \prod_{t=1}^{T} \prod_{j=1}^{n} \left(-\infty, F^{-1}_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, z_{jt}) \right] ;\ j = 1, ..., n \text{ and } t = 1, ..., T \right\} \subset \mathbb{R}^{nT}$$

and $B = \left\{ \prod_{t=1}^{T} \prod_{j=1}^{n} (0, z_{jt}] ;\ j = 1, ..., n \text{ and } t = 1, ..., T \right\} \subset (0,1)^{nT}$. Now,

$$\begin{aligned}
F_{(\mathbf{Z}_1,...,\mathbf{Z}_T|\mathbf{Y}_0)}(\mathbf{z}_1, ..., \mathbf{z}_T|\mathcal{G}_0) &= \mathbb{P}(\mathbf{Z}_1 \leq \mathbf{z}_1, ..., \mathbf{Z}_T \leq \mathbf{z}_T|\mathcal{G}_0) \\
&= \mathbb{P}(Y_{i_j,t} \leq F^{-1}_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, z_{jt}) \text{ for all } j \text{ and } t|\mathcal{G}_0) \\
&= \int_A \prod_{t=1}^{T} \prod_{j=1}^{n} f_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, u_{i_j,t}) du_{i_j,t} \\
&= \int_B \prod_{t=1}^{T} \prod_{j=1}^{n} dv_{jt} = \prod_{t=1}^{T} \prod_{j=1}^{n} z_{jt}.
\end{aligned}$$

The second equality follows from absolute continuity of F_{t-1}. The third equality uses equations (1) and (3) to rewrite the joint density. The fourth equality follows by change of variable $v_{jt} = F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, u_{i_j,t})$, and the fifth by integration. Therefore, $Z_{11}, ..., Z_{nT}$ are independent (conditional on $\mathbf{Y}_0$)[9], and each $Z_{jt} \sim Uniform(0,1)$. Because Φ^{-1} is continuous, it is measurable. Then $\mathbf{R}_{1,\boldsymbol{\theta}_0}, ..., \mathbf{R}_{T,\boldsymbol{\theta}_0}$, where $\mathbf{R}_{t,\boldsymbol{\theta}_0} = \begin{bmatrix} \Phi^{-1}(Z_{1t}) & \cdots & \Phi^{-1}(Z_{nt}) \end{bmatrix}'$, are independent as measurable mappings of independent random variables. Clearly, $R_{jt,\boldsymbol{\theta}_0} \sim N(0,1)$ for each j and t, and

$$\begin{bmatrix} \mathbf{R}'_{1,\boldsymbol{\theta}_0} & \cdots & \mathbf{R}'_{T,\boldsymbol{\theta}_0} \end{bmatrix}' = \begin{bmatrix} R_{11,\boldsymbol{\theta}_0} & \cdots & R_{nT,\boldsymbol{\theta}_0} \end{bmatrix}' \sim N(0, I_{nT}).$$

Because the mapping $F_{i_j,j-1,t-1} : \boldsymbol{\Theta} \times \mathbb{R} \to (0,1)$ is continuous with respect to $\boldsymbol{\theta}$, the Continuous Mapping Theorem (see for example van der Vaart (1998), page 7) and Condition 1(3) together imply that $F_{i_j,j-1,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{i_j,t}) \xrightarrow{P} F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, y_{i_j,t})$ whereas the continuity of $\Phi^{-1} : (0,1) \to \mathbb{R}$ yields

$$R_{jt,\widehat{\boldsymbol{\theta}}_T} = \Phi^{-1}\left(F_{i_j,j-1,t-1}(\widehat{\boldsymbol{\theta}}_T, y_{i_j,t})\right) \xrightarrow{P} \Phi^{-1}\left(F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, y_{i_j,t})\right) = R_{jt,\boldsymbol{\theta}_0}$$

for each j and t. Then $\begin{bmatrix} \mathbf{R}'_{1,\widehat{\boldsymbol{\theta}}_T} & \cdots & \mathbf{R}'_{H,\widehat{\boldsymbol{\theta}}_T} \end{bmatrix}' \xrightarrow{W} N(0, I_{nH})$ for H fixed.

The independence of $\mathbf{R}_{t+s,\boldsymbol{\theta}_0}$ and $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$ for $s \geq 1$ follows easily using the results above: $\mathbf{R}_{t+s,\boldsymbol{\theta}_0}$ is independent of $\{\mathbf{R}_{1,\boldsymbol{\theta}_0}, ..., \mathbf{R}_{t,\boldsymbol{\theta}_0}\}$, and $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$ is a measurable mapping of $\{\mathbf{R}_{1,\boldsymbol{\theta}_0}, ..., \mathbf{R}_{t,\boldsymbol{\theta}_0}\}$, because $Y_{i_j,t} = F^{-1}_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, \Phi(R_{jt,\boldsymbol{\theta}_0}))$.

■

[9] This remark holds for every independence proven in this paper and is hereafter omitted.

A.1 Proof of Lemma 3

We apply Lemma 6 in the proof of Lemma 3.

Lemma 6 *Let* $X_1, ..., X_n$ *be i.i.d. uniform random variables on* $(0, 1)$ *and* $X = \prod_{i=1}^{n} X_i$*, then* $f_n(X) \sim Uniform(0, 1)$*, where* $f_n(x) = x \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} (\ln x)^i$.

Proof. Let $n = 2$, and denote $Z_1 = X_1 X_2$ and $Z_2 = X_2$. The Jacobian determinant of the inverse transformation is $\frac{1}{z_2}$, and hence we get the joint density function $f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{z_2}$, when $0 < z_1 < z_2 < 1$, and $f_{Z_1,Z_2}(z_1, z_2) = 0$ otherwise. Integrating with respect to z_2 over the range $(z_1, 1)$ yields the marginal density function $f_{Z_1}(z_1) = -\ln z_1$, and the cumulative distribution function $F_{Z_1}(z_1) = z_1 - z_1 \ln z_1$. From the proof of Lemma 2, $F_{Z_1}(Z_1) \sim Uniform(0, 1)$, as required.

We make an induction assumption that the result holds for $n = k - 1$, and show that it holds for $n = k$. Denote $Z = \prod_{i=1}^{k-1} X_i$. The induction assumption gives $F_Z(z) = z \sum_{i=0}^{k-2} \frac{(-1)^i}{i!} (\ln z)^i$. Therefore, derivation with respect to z yields the density function of the variable Z

$$f_Z(z) = \frac{(-1)^{k-2}}{(k-2)!} (\ln z)^{k-2}. \tag{17}$$

Denote $V_1 = Z X_k$ and $V_2 = X_k$. Because Z and X_k are independent, the joint density function of V_1 and V_2 is obtained by calculating the Jacobian determinant for the inverse transformation and applying the standard formula to obtain

$$f_{V_1,V_2}(v_1, v_2) = f_Z(\frac{v_1}{v_2}) f_{X_k}(v_2) \frac{1}{v_2} = \frac{(-1)^{k-2}}{(k-2)!} (\ln \frac{v_1}{v_2})^{k-2} \frac{1}{v_2},$$

when $0 < v_1 < v_2 < 1$, and $f_{V_1,V_2}(v_1, v_2) = 0$ otherwise. Because

$$\frac{d}{dv_2}(\ln \frac{v_1}{v_2})^{k-1} = (-1)(k-1)(\ln \frac{v_1}{v_2})^{k-2}\frac{1}{v_2}$$

and $\ln 1 = 0$, the density function of V_1 is

$$f_{V_1}(v_1) = \frac{(-1)^{k-2}}{(k-2)!}\int_{v_1}^{1}(\ln \frac{v_1}{v_2})^{k-2}\frac{1}{v_2}dv_2 = \frac{(-1)^{k-1}}{(k-1)!}(\ln v_1)^{k-1}.$$

Integrating by parts we get the distribution function of V_1

$$\begin{aligned} F_{V_1}(v_1) &= \int_0^{v_1} \frac{(-1)^{k-1}}{(k-1)!}(\ln x)^{k-1}dx \\ &= v_1\frac{(-1)^{k-1}}{(k-1)!}(\ln v_1)^{k-1} - \lim_{x\to 0} x\frac{(-1)^{k-1}}{(k-1)!}(\ln x)^{k-1} + \int_0^{v_1}\frac{(-1)^{k-2}}{(k-2)!}(\ln x)^{k-2}dx \end{aligned}$$

for $0 < v_1 < 1$.

Using (17), we see that $\int_0^{v_1} \frac{(-1)^{k-2}}{(k-2)!}(\ln x)^{k-2}dx = F_Z(v_1) = v_1 \sum_{i=0}^{k-2} \frac{(-1)^i}{i!}(\ln v_1)^i$. An application of L'Hospital's Rule $(k-1)$ times yields $\lim_{x\to 0} \frac{(-1)^{k-1}}{(k-1)!}\frac{(\ln x)^{k-1}}{x^{-1}} = \lim_{x\to 0} x = 0$. Therefore,

$$F_{V_1}(v_1) = v_1 \sum_{i=0}^{k-1} \frac{(-1)^i}{i!}(\ln v_1)^i,$$

and, from the proof of Lemma 2, $F_{V_1}(v_1) \sim Uniform(0,1)$. Since $V_1 = \prod_{i=1}^k X_i$, the induction principle completes the proof. ■

Proof of Lemma 3. Write $X_{t,\boldsymbol{\theta}} = \prod_{j=1}^{n} F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})$ using (3). Lemma 2 shows that $F_{i_j,j-1,t-1}(\boldsymbol{\theta}_0, Y_{i_j,t})$ are i.i.d. uniform variables, so that by Lemma 6,

$$Z_{t,\boldsymbol{\theta}_0} \sim Uniform(0,1),$$

where $Z_{t,\boldsymbol{\theta}_0} = X_{t,\boldsymbol{\theta}} \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} (\ln X_{t,\boldsymbol{\theta}})^i$. Clearly, $Q_{t,\boldsymbol{\theta}_0} = \Phi^{-1}(Z_{t,\boldsymbol{\theta}_0}) \sim N(0,1)$. Because $\{X_{1,\boldsymbol{\theta}_0}, ..., X_{T,\boldsymbol{\theta}_0}\}$ are independent (Lemma 2), $Z_{t,\boldsymbol{\theta}_0}$s and $Q_{t,\boldsymbol{\theta}_0}$s are independent as measurable transformations of independent variables. Therefore, part a) of the Lemma follows.

The random variables $Q_{t,\boldsymbol{\theta}}$ are continuous in $\boldsymbol{\theta}$ for all t. Then, Condition 1(3), the Continuous Mapping Theorem, and part a) together yield part b).

Lemma 2 c) shows that $X_{t+s,\boldsymbol{\theta}_0} = \prod_{j=1}^{n} F_{i_j,j-1,t+s-1}(\boldsymbol{\theta}_0, Y_{i_j,t})$ and $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$ are independent for $s \geq 1$. Hence, $Q_{t,\boldsymbol{\theta}_0}$ and $\{\mathbf{Y}_1, ..., \mathbf{Y}_t\}$ are independent for $s \geq 1$. ■

A.2 Conditions for Theorem 5

Condition 7 below is a slightly modified version of conditions presented in Sweeting (1980) and Basawa and Scott (1983), and guarantees that the asymptotic distribution of the MLE is mixed normal. We use $\|\cdot\|$ to signify the Euclidean norm.

Condition 7 *Let the following assumptions hold.*

(1) $\boldsymbol{\Theta} \subset \mathbb{R}^k$ *is an open set.*

(2) *The model is correctly specified, i.e.,* $f(\boldsymbol{\theta}_0, \mathbf{y}) \in \mathcal{P}$.

(3) *For every* $(\boldsymbol{\theta}, \mathbf{x}) \in \boldsymbol{\Theta} \times D$, *where* $D \subset \mathbb{R}^n$, *and every* $t = 1, ..., T$, $f_{t-1}(\boldsymbol{\theta}, \mathbf{x}) > 0$ *and the second partial derivatives* $\frac{\partial^2}{\partial\theta_i \partial\theta_j} f_{t-1}(\boldsymbol{\theta}, \mathbf{x})$, $i, j = 1, ..., k$, *exist and are continuous.*

(4) *There exist nonrandom nonsingular* $k \times k$ *diagonal matrices* $\boldsymbol{\xi}_T$ *such that* $\boldsymbol{\xi}_T^{-1} \to 0$ *and, for all* $c > 0$,

$$\sup_{\boldsymbol{\theta} \in M_{T,c}} \left\| \boldsymbol{\xi}_T^{-1} \left[B_T(\boldsymbol{\theta}) - B_T(\boldsymbol{\theta}_0) \right] \boldsymbol{\xi}_T^{-1} \right\| \xrightarrow{P} 0,$$

where $M_{T,c} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \|\boldsymbol{\xi}_T(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\| \leq c\}$ *and*
$B_T(\boldsymbol{\theta}) = -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} l_T(\boldsymbol{\theta}, \mathbf{Y}) = -\left[\sum_{t=1}^{T} \frac{\partial^2 l_t(\boldsymbol{\theta}, \mathbf{Y}_t)}{\partial\theta_i\partial\theta_j}\right]_{i,j=1}^{k}$.

(5) *Let* $S_T(\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} l_T(\boldsymbol{\theta}, \mathbf{Y}) = \sum_{t=1}^{T} \frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}, \mathbf{Y}_t)$ *be the score function and* $W_T(\boldsymbol{\theta}_0) \equiv \boldsymbol{\xi}_T^{-1} B_T(\boldsymbol{\theta}_0)\boldsymbol{\xi}_T^{-1}$ *a scaled Hessian matrix. There exists a (possibly) random matrix* $\mathcal{I}(\boldsymbol{\theta}_0)$ *such that*

$$\left[W_T(\boldsymbol{\theta}_0) \quad \left[\boldsymbol{\xi}_T^{-1} S_T(\boldsymbol{\theta}_0)\right]'\right]' \xrightarrow{W} \left[\mathcal{I}(\boldsymbol{\theta}_0) \quad \left[\mathcal{I}(\boldsymbol{\theta}_0)^{1/2}\mathbf{Z}\right]'\right]',$$

where $\mathbb{P}_{\boldsymbol{\theta}_0}(\mathcal{I}(\boldsymbol{\theta}_0) > 0) = 1$, *and* $\mathbf{Z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$ *is independent of* $\mathcal{I}(\boldsymbol{\theta}_0)$.

Condition 7(3) imposes fairly standard regularity conditions on the conditional density functions. Combined with Condition 7(1), it implies the applicability of the Mean-Value Theorem for the score function in any convex set $C \subset \boldsymbol{\Theta}$. The correct model specification is necessary for Proposition 8 below and for testing purposes. Condition 7(4) is technical and gives a uniform convergence in probability of the Hessian of the log-likelihood on special compact sets that contain the true parameter value $\boldsymbol{\theta}_0$. Using the general weight matrices $\boldsymbol{\xi}_T$ in the condition and allowing the matrix $\mathcal{I}(\boldsymbol{\theta}_0)$, the limit of the scaled Hessian $W_T(\boldsymbol{\theta}_0)$ in Condition 7(5), to be random makes the framework applicable also in the non-ergodic case. Condition 7(5) is a high level assumption needed to obtain asymptotic mixed normality of the MLE. For this, Condition 7(1) is also pertinent because it guarantees that the MLE is an inner point. In standard cases one typically verifies Condition 7(4) using an appropriate uniform law of large numbers and Condition 7(5) using a martingale central limit theorem. In these cases the MLE is asymptotically normally distributed with a constant covariance matrix $\mathcal{I}(\boldsymbol{\theta}_0)$. In co-integrated VAR models, the most typical multivariate non-ergodic examples in econometrics, one uses a functional central limit theorem instead of its conventional counterpart.

Proposition 8 *Under Condition 7, there exists a sequence of local maximizers* $\widehat{\boldsymbol{\theta}}_T$ *such that* $\left\{\boldsymbol{\xi}_T(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)\right\}_{T\in\mathbb{N}}$ *is bounded in probability and*

$$\boldsymbol{\xi}_T^{-1} S_T(\boldsymbol{\theta}_0) - W_T(\boldsymbol{\theta}_0)\boldsymbol{\xi}_T(\widehat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{P} 0.$$

The proof is given in Sweeting (1980). Proposition 8 yields the asymptotic mixed normality of the MLE mentioned in Section 2.4.

Condition 9 below is necessary for the general framework of obtaining tests. Therefore, it contains functions of quantile residuals along with quantities derived from the log-likelihood function.

Condition 9 *Let the vector* $\mathbf{U}_{t,\boldsymbol{\theta}_0}$ *and the function* g *be as in Condition 4 and let the following assumptions hold.*

(1) *For all* $c > 0$

$$\sup_{\boldsymbol{\theta}\in M_{T,c}} \left\| \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}}) - \mathbf{G} \right\| \xrightarrow{P} 0, \quad \sup_{\boldsymbol{\theta}\in M_{T,c}} \left\| \frac{1}{T}\sum_{t=1}^{T} g(\mathbf{U}_{t,\boldsymbol{\theta}})g(\mathbf{U}_{t,\boldsymbol{\theta}})' - \mathbf{H} \right\| \xrightarrow{P} 0,$$

and

$$\sup_{\boldsymbol{\theta}\in M_{T,c}} \left\| \frac{1}{\sqrt{T}} \sum\nolimits_{t=1}^{T} g(\mathbf{U}_{t,\boldsymbol{\theta}}) \left[\frac{\partial}{\partial\boldsymbol{\theta}} l_t(\boldsymbol{\theta}, Y_t)\right]' \boldsymbol{\xi}_T^{-1} - \boldsymbol{\Psi} \right\| \xrightarrow{P} 0,$$

where $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial\boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ *and* $\mathbf{H} = \mathbb{E}(g(\mathbf{U}_{t,\boldsymbol{\theta}_0})g(\mathbf{U}_{t,\boldsymbol{\theta}_0})')$ *exist and are finite and* $\boldsymbol{\Psi}$ *is a (possibly) random matrix. Moreover, the matrix* $\mathbf{H}$ *is positive definite.*

(2) *There exists a nonrandom $k \times k$ matrix $\mathbf{J}$ such that $\left\| \sqrt{T}\boldsymbol{\xi}_T^{-1} - \mathbf{J} \right\| \to 0$, and*

$$\left[W_T(\boldsymbol{\theta}_0) \quad : \quad \left[\boldsymbol{\xi}_T^{-1} S_T\left(\boldsymbol{\theta}_0\right)\right]' \quad \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(\mathbf{U}_{t,\boldsymbol{\theta}_0})' \right]'$$
$$\xrightarrow{W} \left[\mathcal{I}(\boldsymbol{\theta}_0) \quad : \quad \left[\boldsymbol{\Sigma}^{1/2}\mathcal{Z}\right]' \right]',$$

where $\mathcal{Z} \sim N_{k+w}(\mathbf{0}, \mathbf{I}_{k+w})$ is independent of $\boldsymbol{\Sigma} = \begin{bmatrix} \mathcal{I}(\boldsymbol{\theta}_0) & \boldsymbol{\Psi}' \\ \boldsymbol{\Psi} & \mathbf{H} \end{bmatrix}$, a positive definite matrix with (possibly random) elements defined above in (1) and Condition 7(5).

(3) *The cumulative distribution function $F_{t-1} : \boldsymbol{\Theta} \times \mathbb{R}^n \to (0,1)$ is continuously differentiable in $(\boldsymbol{\theta}, \mathbf{x}) \in \boldsymbol{\Theta} \times \mathbb{R}^n$ for all $t = 1, ..., T$.*

Condition 9(1) imposes uniform convergence in probability on special compact sets similar to that in Condition 7(4). Together these two conditions define the matrix $\boldsymbol{\Sigma}$ in Condition 9(2). One can verify the joint weak convergence in Condition 9(2) by using an appropriate (functional) central limit theorem. Condition 9(2) contains Condition 7(5) as a special case. Condition 9(3) complements Condition 7(3) and guarantees the existence of derivatives of quantile residuals.

Proof of Theorem 5. We can assumed that $\widehat{\boldsymbol{\theta}}_T \neq \infty$, because $\lim_{T\to\infty} \mathbb{P}(\widehat{\boldsymbol{\theta}}_T \neq \infty) = 1$ by Proposition 8. Here $\mathbb{P} = \mathbb{P}_{\boldsymbol{\theta}_0}$ is the probability measure induced by the true parameter value $\boldsymbol{\theta}_0$. Again by Proposition 8, for every $\varepsilon > 0$ exists c_0 and T_0 such that $\mathbb{P}(\widehat{\boldsymbol{\theta}}_T \in M_{T,c_0}) > 1 - \varepsilon$ for all $T > T_0$. By the first uniform convergence in Condition 9(1), $\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}_T}) \xrightarrow{P} \mathbf{G}$ for all $\tilde{\boldsymbol{\theta}}_T \in M_{T,c}$ and $c > 0$, thus, $\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) \xrightarrow{P} \mathbf{G}$.

The Mean-Value Theorem and Conditions 4 and 9(3) together imply that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}})(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)+\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}), \tag{18}$$

where $\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}}) = \begin{bmatrix}\frac{\partial}{\partial\boldsymbol{\theta}}g_1(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}^{(1)}}) & \cdots & \frac{\partial}{\partial\boldsymbol{\theta}}g_n(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}^{(n)}})\end{bmatrix}'$ is a $(n\times k)$ Jacobian-matrix with $\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}^{(j)}} = \begin{bmatrix}\mathbf{R}'_{t,\tilde{\boldsymbol{\theta}}^{(j)}} & \cdots & \mathbf{R}'_{t-m+1,\tilde{\boldsymbol{\theta}}^{(j)}}\end{bmatrix}'$ (or $\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}^{(j)}} = \begin{bmatrix}Q_{t,\tilde{\boldsymbol{\theta}}^{(j)}} & \cdots & Q_{t-m+1,\tilde{\boldsymbol{\theta}}^{(j)}}\end{bmatrix}'$ depending on the choice in Condition 4), $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\theta}}^{(1)},\ldots,\tilde{\boldsymbol{\theta}}^{(n)})$, and $\left\|\tilde{\boldsymbol{\theta}}^{(j)}-\boldsymbol{\theta}_0\right\| < \left\|\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0\right\|$ for each $j=1,...,n$. Proposition 8 and Condition 7(5) give

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) = \sqrt{T}\boldsymbol{\xi}_T^{-1}W_T(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\xi}_T^{-1}S_T(\boldsymbol{\theta}_0)+o_P(\mathbf{1}), \tag{19}$$

because, by Condition 9(2), $\sqrt{T}\boldsymbol{\xi}_T{}^{-1}W_T(\boldsymbol{\theta}_0)^{-1}\cdot o_P(\mathbf{1}) = o_P(\mathbf{1})$.
Because (see Condition 9(1)) $\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}})\cdot o_P(\mathbf{1}) = o_P(\mathbf{1})$, equations (18) and (19) yield

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T})\\ &= \begin{bmatrix}\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}})\sqrt{T}\boldsymbol{\xi}_T^{-1}W_T(\boldsymbol{\theta}_0)^{-1} & : & \mathbf{I}_w\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}_T^{-1}S_T(\boldsymbol{\theta}_0)\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(\mathbf{U}_{t,\boldsymbol{\theta}_0})\end{bmatrix}+o_P(\mathbf{1}).\end{aligned}$$

Conditions 9(1), 9(2), and the Continuous Mapping Theorem ensure that

$$\begin{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\tilde{\boldsymbol{\theta}}})\boldsymbol{\xi}_T^{-1}W_T(\boldsymbol{\theta}_0)^{-1} & : & \mathbf{I}_w\end{bmatrix} \xrightarrow{W} \begin{bmatrix}\mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} & : & \mathbf{I}_w\end{bmatrix}. \tag{20}$$

Finally, using (20), Condition 9(2), and the Continuous Mapping Theorem

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g(\mathbf{U}_{t,\widehat{\boldsymbol{\theta}}_T}) \stackrel{W}{\to} \left[\mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \quad : \quad \mathbf{I}_w\right] \cdot \boldsymbol{\Sigma}^{1/2}\mathcal{Z},$$

where $\mathcal{Z} \sim N_{k+w}(\mathbf{0}, \mathbf{I}_{k+w})$. Furthermore, setting

$$\begin{aligned}\boldsymbol{\Omega} &= \left[\mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \quad : \quad \mathbf{I}_w\right] \boldsymbol{\Sigma} \begin{bmatrix} \mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{J}'\mathbf{G}' \\ \mathbf{I}_w \end{bmatrix} \\ &= \mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{J}'\mathbf{G}' + \boldsymbol{\Psi}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\mathbf{J}'\mathbf{G}' + \mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{\Psi}' + \mathbf{H}\end{aligned}$$

we can write

$$\left[\mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \quad : \quad \mathbf{I}_w\right] \boldsymbol{\Sigma}^{1/2}\mathcal{Z} = \boldsymbol{\Omega}^{1/2}\mathcal{U},$$

where $\mathcal{U} \sim N_w(\mathbf{0}, \mathbf{I}_w)$. Independence of $\mathcal{U}$ and $\boldsymbol{\Omega}$ follows from that of $\mathcal{Z}$ and $\left[\mathbf{GJ}\mathcal{I}(\boldsymbol{\theta}_0)^{-1} \quad : \quad \mathbf{I}_w\right] \boldsymbol{\Sigma}^{1/2}$.

By Condition 7(4) and Proposition 8, $\left\|W_T(\widehat{\boldsymbol{\theta}}_T) - W_T(\boldsymbol{\theta}_0)\right\| \stackrel{P}{\to} 0$. Because $W_T(\boldsymbol{\theta}_0) \stackrel{W}{\to} \mathcal{I}(\boldsymbol{\theta}_0)$, by Condition 7(5), then $W_T(\widehat{\boldsymbol{\theta}}_T) \stackrel{W}{\to} \mathcal{I}(\boldsymbol{\theta}_0)$. The Continuous Mapping Theorem yields $W_T(\widehat{\boldsymbol{\theta}}_T)^{-1} \stackrel{W}{\to} \mathcal{I}(\boldsymbol{\theta}_0)^{-1}$. Finally, Condition 9(1) together with the Continuous Mapping Theorem yield $\hat{\boldsymbol{\Omega}}_T \stackrel{W}{\to} \boldsymbol{\Omega}$.

The joint convergence follows using $\left\|W_T(\widehat{\boldsymbol{\theta}}_T) - W_T(\boldsymbol{\theta}_0)\right\| \stackrel{P}{\to} 0$ and Conditions 9(1) and 9(2). ■

A.3 Derivatives

Lemma 10

$$\frac{\partial}{\partial \boldsymbol{\theta}} R_{jt,\boldsymbol{\theta}} = \left[\phi\left(R_{jt,\boldsymbol{\theta}}\right)\right]^{-1} \frac{\partial}{\partial \boldsymbol{\theta}}(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, Y_{i_j,t})),$$

and

$$\frac{\partial}{\partial \boldsymbol{\theta}} Q_{t,\boldsymbol{\theta}} = \left[\phi\left(Q_{t,\boldsymbol{\theta}}\right)\right]^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} Z_{t,\boldsymbol{\theta}},$$

where $\frac{\partial}{\partial \boldsymbol{\theta}} Z_{t,\boldsymbol{\theta}} = \frac{(-1)^{n-1}}{(n-1)!}\left(\log X_{t,\boldsymbol{\theta}}\right)^{n-1} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} X_{t,\boldsymbol{\theta}}$ *and* $\frac{\partial}{\partial \boldsymbol{\theta}} X_{t,\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}\left(\prod_{j=1}^{n} F_{i_j,j-1,t-1}(\boldsymbol{\theta}, Y_{i_j,t})\right)$. *Here* ϕ *is the density of the standard normal distribution.*

Proof. Let $r_{jt,\boldsymbol{\theta}} = \Phi^{-1}(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t}))$. The fact that $\phi(x) > 0$ for all $x \in \mathbb{R}$ ensures that $\frac{d}{dy}\Phi^{-1}(y) = \frac{1}{(\frac{d}{dx}\Phi)(x)} = \frac{1}{\phi(x)}$, where $x = \Phi^{-1}(y)$, exists for each $y \in (0,1)$. This and Condition 9(3) give

$$\begin{aligned}
\frac{\partial}{\partial \theta_s} r_{jt,\boldsymbol{\theta}} &= \frac{\partial}{\partial \theta_s} \Phi^{-1}(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})) \\
&= \left[\left(\Phi^{-1}\right)'\left(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})\right)\right] \frac{\partial}{\partial \theta_s}\left(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})\right) \\
&= \left[\Phi'\left[\Phi^{-1}(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t}))\right]\right]^{-1} \frac{\partial}{\partial \theta_s}\left(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})\right) \\
&= \left[\phi\left(r_{jt,\boldsymbol{\theta}}\right)\right]^{-1} \cdot \frac{\partial}{\partial \theta_s}\left(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})\right)
\end{aligned}$$

for all $s = 1, ..., k$. Since $\frac{\partial}{\partial \theta_s} r_{jt,\boldsymbol{\theta}}$ is continuous, $\frac{\partial}{\partial \theta_s} R_{jt,\boldsymbol{\theta}}$ is a well defined random variable.

Similarly, Condition 9(3) implies $\frac{\partial}{\partial \boldsymbol{\theta}} q_{t,\boldsymbol{\theta}} = \left[\phi\left(q_{t,\boldsymbol{\theta}}\right)\right]^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} z_{t,\boldsymbol{\theta}}$. Because

$$\frac{d}{dx} f_n(x) = \frac{d}{dx}\left(x \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} (\log x)^i\right) = \frac{(-1)^{n-1}}{(n-1)!} (\log x)^{n-1},$$

then $\frac{\partial}{\partial \boldsymbol{\theta}} z_{t,\boldsymbol{\theta}} = \frac{(-1)^{n-1}}{(n-1)!}\left(\log x_{t,\boldsymbol{\theta}}\right)^{n-1} \cdot \frac{\partial}{\partial \boldsymbol{\theta}} x_{t,\boldsymbol{\theta}}$, and $\frac{\partial}{\partial \boldsymbol{\theta}} x_{t,\boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}}\left(\prod_{j=1}^{n} F_{i_j,j-1,t-1}(\boldsymbol{\theta}, y_{i_j,t})\right)$. Because $\frac{\partial}{\partial \boldsymbol{\theta}} q_{t,\boldsymbol{\theta}}$ is continuous, $\frac{\partial}{\partial \boldsymbol{\theta}} Q_{t,\boldsymbol{\theta}}$ is a well defined random variable. ■

Remark 11 *The random variables*

(1) $\frac{\partial}{\partial \boldsymbol{\theta}'} R_{i,t-s,\boldsymbol{\theta}_0}$ *and* $R_{jt,\boldsymbol{\theta}_0}$ *are independent for all* $i, j \in \{1, ..., n\}$ *and* $s \geq 1$,

and

(2) $\frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t-s,\boldsymbol{\theta}_0}$ *and* $Q_{t,\boldsymbol{\theta}_0}$ *are independent for all* $s \geq 1$.

Proof. According to Lemma 10

$$\frac{\partial}{\partial \boldsymbol{\theta}} R_{i,t-s,\boldsymbol{\theta}_0} = [\phi(R_{i,t-s,\boldsymbol{\theta}_0})]^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} (F_{m_i,i-1,t-s-1}(\boldsymbol{\theta}_0, Y_{m_i,t-s}))$$

is a measurable function of the random variables $\{\mathbf{Y}_0, \mathbf{Y}_1, ..., \mathbf{Y}_{t-s}\}$. Lemma 2 c) shows the independence of $R_{jt,\boldsymbol{\theta}_0}$ and $\{\mathbf{Y}_0, \mathbf{Y}_1, ..., \mathbf{Y}_{t-s}\}$ for all $s \geq 1$, which implies result (1). Likewise, Lemma 3 c) yields the independence of $Q_{t,\boldsymbol{\theta}_0}$ and $\frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t-s,\boldsymbol{\theta}_0}$ for all $s \geq 1$. ■

Using Remark 11(1), we see that a typical row of the matrix $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ in the *autocorrelation test* with $\mathbf{U}_{t,\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{R}'_{t,\boldsymbol{\theta}} & \cdots & \mathbf{R}'_{t-K_1,\boldsymbol{\theta}} \end{bmatrix}'$ is

$$\begin{aligned} \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'}(R_{i,t-s,\boldsymbol{\theta}_0} R_{jt,\boldsymbol{\theta}_0})) &= \mathbb{E}(R_{i,t-s,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} R_{jt,\boldsymbol{\theta}_0}) + \mathbb{E}(R_{jt,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} R_{i,t-s,\boldsymbol{\theta}_0}) \\ &= \mathbb{E}(R_{i,t-s,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} R_{jt,\boldsymbol{\theta}_0}), \end{aligned}$$

where $\frac{\partial}{\partial \boldsymbol{\theta}'} R_{jt,\boldsymbol{\theta}_0}$ is the vector of derivatives given in Lemma 10, $i, j \in \{1, ..., n\}$, and $s = 1, ..., K_1$.

If $\mathbf{U}_{t,\boldsymbol{\theta}} = \begin{bmatrix} Q_{t,\boldsymbol{\theta}} & \cdots & Q_{t-K_1,\boldsymbol{\theta}} \end{bmatrix}'$ in the autocorrelation test, then using Remark 11(2) we see that the sth row of the matrix $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ is

$$\begin{aligned} \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'}(Q_{t-s,\boldsymbol{\theta}_0} Q_{t,\boldsymbol{\theta}_0})) &= \mathbb{E}(Q_{t-s,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}_0}) + \mathbb{E}(Q_{t,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t-s,\boldsymbol{\theta}_0}) \\ &= \mathbb{E}(Q_{t-s,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}_0}), \end{aligned}$$

where $\frac{\partial}{\partial \boldsymbol{\theta}} Q_{t,\boldsymbol{\theta}}$ is given in Lemma 10.

Remark 12 *The random variables*

(1) $R^2_{it,\boldsymbol{\theta}_0}$ *and* $R_{j,t-s,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}R_{j,t-s,\boldsymbol{\theta}_0}$ *are independent for all* $s\geq 1$,

and

(2) $Q^2_{t,\boldsymbol{\theta}_0}$ *and* $Q_{t-s,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}Q_{t-s,\boldsymbol{\theta}_0}$ *are independent for all* $s\geq 1$.

Proof. $R^2_{it,\boldsymbol{\theta}_0}$ is a measurable function of $R_{it,\boldsymbol{\theta}_0}$, and $\frac{\partial}{\partial\boldsymbol{\theta}}(F_{m_j,j-1,t-s-1}(\boldsymbol{\theta}_0, Y_{m_j,t-s}))$, $\left[\phi\left(R_{j,t-s,\boldsymbol{\theta}_0}\right)\right]^{-1}$, and $R_{j,t-s,\boldsymbol{\theta}_0}$ are measurable functions of $\{\mathbf{Y}_0, Y_1, ..., Y_{t-s}\}$. The independence follows using Lemma 2 c). Similarly, Lemma 3 c) yields the independence of $Q^2_{t,\boldsymbol{\theta}_0}$ and $Q_{t-s,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}Q_{t-s,\boldsymbol{\theta}_0}$ for all $s\geq 1$. ■

Using Remark 12(1), a typical row of the matrix $\mathbf{G}=\mathbb{E}(\frac{\partial}{\partial\boldsymbol{\theta}'}g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ in the *conditional heteroscedasticity test* based on $\mathbf{U}_{t,\boldsymbol{\theta}}=\begin{bmatrix}\mathbf{R}'_{t,\boldsymbol{\theta}} & \cdots & \mathbf{R}'_{t-K_2,\boldsymbol{\theta}}\end{bmatrix}'$ is

$$\begin{aligned}
& \mathbb{E}\left[\frac{\partial}{\partial\boldsymbol{\theta}'}\left(R^2_{i,t-s,\boldsymbol{\theta}_0}-1\right)\left(R^2_{jt,\boldsymbol{\theta}_0}-1\right)\right] \\
= \;& 2\mathbb{E}\left[\left(R^2_{i,t-s,\boldsymbol{\theta}_0}-1\right)R_{jt,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}R_{jt,\boldsymbol{\theta}_0}+\left(R^2_{jt,\boldsymbol{\theta}_0}-1\right)R_{i,t-s,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}R_{i,t-s,\boldsymbol{\theta}_0}\right] \\
= \;& 2\mathbb{E}\left[\left(R^2_{i,t-s,\boldsymbol{\theta}_0}-1\right)R_{jt,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}R_{jt,\boldsymbol{\theta}_0}\right],
\end{aligned}$$

where $\frac{\partial}{\partial\boldsymbol{\theta}'}R_{jt,\boldsymbol{\theta}_0}$ is the vector of derivatives given in Lemma 10, $i,j\in\{1,...,n\}$, and $s=1,...,K_2$.

If $\mathbf{U}_{t,\boldsymbol{\theta}}=\begin{bmatrix}Q_{t,\boldsymbol{\theta}} & \cdots & Q_{t-K_2,\boldsymbol{\theta}}\end{bmatrix}'$ in the conditional heteroscedasticity test, then Remark 12(2) yields

$$\begin{aligned}
& \mathbb{E}(\frac{\partial}{\partial\boldsymbol{\theta}'}\left(Q^2_{t-s,\boldsymbol{\theta}_0}-1\right)\left(Q^2_{t,\boldsymbol{\theta}_0}-1\right)) \\
= \;& 2\mathbb{E}(\left(Q^2_{t-s,\boldsymbol{\theta}_0}-1\right)Q_{t,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}Q_{t,\boldsymbol{\theta}_0})+\mathbb{E}(\left(Q^2_{t,\boldsymbol{\theta}_0}-1\right)Q_{t-s,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}Q_{t-s,\boldsymbol{\theta}_0}) \\
= \;& 2\mathbb{E}(\left(Q^2_{t-s,\boldsymbol{\theta}_0}-1\right)Q_{t,\boldsymbol{\theta}_0}\frac{\partial}{\partial\boldsymbol{\theta}'}Q_{t,\boldsymbol{\theta}_0}),
\end{aligned}$$

as the sth row of matrix $\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0}))$ and $\frac{\partial}{\partial \boldsymbol{\theta}} Q_{t,\boldsymbol{\theta}}$ is given in Lemma 10.

In the *multinormality test* with $\mathbf{U}_{t,\boldsymbol{\theta}} = \mathbf{R}_{t,\boldsymbol{\theta}}$, we have

$$\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(\mathbf{U}_{t,\boldsymbol{\theta}_0})) = \begin{bmatrix} \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}} g_1(R_{1t,\boldsymbol{\theta}_0})) & \cdots & \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}} g_n(R_{nt,\boldsymbol{\theta}_0})) \end{bmatrix}',$$

where

$$\mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}} g_j(R_{jt,\boldsymbol{\theta}_0})) = \mathbb{E} \begin{bmatrix} 2R_{jt,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}} R_{jt,\boldsymbol{\theta}_0} & 3R^2_{jt,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}} R_{jt,\boldsymbol{\theta}_0} & 4R^3_{jt,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}} R_{jt,\boldsymbol{\theta}_0} \end{bmatrix}$$

and $\frac{\partial}{\partial \boldsymbol{\theta}} R_{jt,\boldsymbol{\theta}_0}$ is given in Lemma 10. If $\mathbf{U}_{t,\boldsymbol{\theta}} = Q_{t,\boldsymbol{\theta}}$, then

$$\mathbf{G} = \mathbb{E}(\frac{\partial}{\partial \boldsymbol{\theta}'} g(Q_{t,\boldsymbol{\theta}_0})) = \mathbb{E} \begin{bmatrix} 2Q_{t,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}_0} & 3Q^2_{t,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}_0} & 4Q^3_{t,\boldsymbol{\theta}_0} \frac{\partial}{\partial \boldsymbol{\theta}'} Q_{t,\boldsymbol{\theta}_0} \end{bmatrix}$$

with $\frac{\partial}{\partial \boldsymbol{\theta}} Q_{t,\boldsymbol{\theta}}$ given in Lemma 10.

B Appendix: LM interpretations

B.1 Autocorrelation test

We obtain the LM interpretation of the multivariate autocorrelation test by applying an auxiliary vector autoregressive model. Hosking (1981) used a similar approach and Pearson's residuals.

Consider an auxiliary VAR(p) model for quantile residuals $\mathbf{R}_{t,\boldsymbol{\theta}} = \sum_{s=1}^{p} \boldsymbol{\Gamma}_s \mathbf{R}_{t-s,\boldsymbol{\theta}} + \boldsymbol{\varepsilon}_t$, where $\boldsymbol{\varepsilon}_t \sim n.i.d.(0, \mathbf{I}_n)$, $\boldsymbol{\Gamma}_s = \begin{bmatrix} \gamma_{uv,s} \end{bmatrix}$, $u, v = 1, ...n$, and $\mathbf{R}_{t-s,\boldsymbol{\theta}} = \begin{bmatrix} R_{1,t-s,\boldsymbol{\theta}} & \cdots & R_{n,t-s,\boldsymbol{\theta}} \end{bmatrix}'$, for $s = 1, ..., p$, and $t = 1, ..., T$, with $\mathbf{R}_{t,\boldsymbol{\theta}} = \mathbf{0}$ for $t \leq 0$. The Jacobian is triangular, because $\frac{\partial r_{mt,\theta}}{\partial y_{i_j}} = 0$ for all $m < j$. Therefore, the Jacobian determinant is $\left|\prod_{m=1}^{n}\frac{\partial r_{mt,\theta}}{\partial y_{im}}\right| = \prod_{m=1}^{n}\left[\phi\left(r_{mt,\boldsymbol{\theta}}\right)\right]^{-1} f_{i_m,m-1,t-1}(\boldsymbol{\theta},y_{i_m,t}) = \left[\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta},\mathbf{y}_t)$. Thus, the joint density function of the observations is

$$f(\boldsymbol{\theta},\boldsymbol{\Gamma}_1,...,\boldsymbol{\Gamma}_p,\mathbf{y}) = \prod_{t=1}^{T}\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}} - \sum_{s=1}^{p}\boldsymbol{\Gamma}_s\mathbf{R}_{t-s,\boldsymbol{\theta}}\right)\left[\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta},\mathbf{y}_t),$$

and the log-likelihood function

$$\begin{aligned}\tilde{l}(\boldsymbol{\theta},\boldsymbol{\Gamma}_1,...,\boldsymbol{\Gamma}_p,\mathbf{y}) &= -\frac{1}{2}\sum_{t=1}^{T}(\mathbf{R}_{t,\boldsymbol{\theta}} - \sum_{s=1}^{p}\boldsymbol{\Gamma}_s\mathbf{R}_{t-s,\boldsymbol{\theta}})'(\mathbf{R}_{t,\boldsymbol{\theta}} - \sum_{s=1}^{p}\boldsymbol{\Gamma}_s\mathbf{R}_{t-s,\boldsymbol{\theta}}) \\ &\quad +\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}R_{it,\boldsymbol{\theta}}^2 + \sum_{t=1}^{T}\log f_{t-1}(\boldsymbol{\theta},y_t).\end{aligned}$$

Thus, for each $u,v = 1,...,n,\ s = 1,...,p$,

$$\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\Gamma}_1,...,\boldsymbol{\Gamma}_p,\mathbf{y})}{\partial\gamma_{uv,s}} = \sum_{t=1}^{T}R_{ut,\boldsymbol{\theta}}R_{v,t-s,\boldsymbol{\theta}} - R_{v,t-s,\boldsymbol{\theta}}\sum_{s=1}^{p}\sum_{l=1}^{n}\gamma_{ul,s}R_{l,t-s,\boldsymbol{\theta}}.$$

The quantile residuals are independent, when $\boldsymbol{\Gamma}_s = \mathbf{0}$ for all $s = 1,...,p$, and the summands in $\frac{\partial\tilde{l}(\boldsymbol{\theta},\mathbf{0},...,\mathbf{0},\mathbf{y})}{\partial\gamma_{uv,s}} = \sum_{t=1}^{T}R_{ut,\boldsymbol{\theta}}R_{v,t-s,\boldsymbol{\theta}}$ are equal to the function $g(\mathbf{r}_{t,\boldsymbol{\theta}})$ of our autocorrelation test. Thus, $\frac{\partial\tilde{l}(\boldsymbol{\theta},\mathbf{0},...,\mathbf{0},\mathbf{y})}{\partial\boldsymbol{\theta}}$ is equal to the score in the main text, and the score $\tilde{\mathbf{s}}(\boldsymbol{\theta},\mathbf{0}) = \begin{bmatrix}\frac{\partial\tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial\boldsymbol{\theta}'} & \frac{\partial\tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial\boldsymbol{\rho}'}\end{bmatrix}'$ contains also the function $g(\mathbf{r}_{t,\boldsymbol{\theta}})$. The LM test based on $\frac{\partial\tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial\boldsymbol{\rho}}$ is, therefore, identical to our autocorrelation test.

Denote $\boldsymbol{\gamma} = vec\begin{bmatrix}\boldsymbol{\Gamma}_1 & \cdots & \boldsymbol{\Gamma}_p\end{bmatrix}$. The well-known regularity of the score function yields

$$\mathbb{E}\begin{bmatrix}\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}}\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}'} & \frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}}\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}'} \\ \frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}}\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}'} & \frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}}\frac{\partial\tilde{l}(\boldsymbol{\theta},\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}'}\end{bmatrix} = -\mathbb{E}\begin{bmatrix}\frac{\partial^2\tilde{l}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} & \frac{\partial^2\tilde{l}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\gamma}'} \\ \frac{\partial^2\tilde{l}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\theta}'} & \frac{\partial^2\tilde{l}(\boldsymbol{\theta}_0,\boldsymbol{\gamma}_0,\mathbf{y})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\end{bmatrix}.$$

Because $\mathbb{E}\left[\frac{\partial \tilde{l}(\boldsymbol{\theta}_0,\gamma_0,\mathbf{y})}{\partial \boldsymbol{\theta}}\frac{\partial \tilde{l}(\boldsymbol{\theta}_0,\gamma_0,\mathbf{y})}{\partial \gamma'}\right] = \boldsymbol{\Psi}'$ and $\mathbb{E}\left[\frac{\partial^2 \tilde{l}(\boldsymbol{\theta}_0,\gamma_0,\mathbf{y})}{\partial \boldsymbol{\theta}\partial \gamma'}\right] = \mathbf{G}'$, we have $\boldsymbol{\Psi} = -\mathbf{G}$. In finite samples the estimates of the corresponding expectations are naturally different, however. Thus, we estimated both statistics separately from the data.

B.2 Conditional heteroscedasticity test

We base the LM interpretation of the conditional heteroscedasticity test on an multivariate ARCH model.

Consider an auxiliary multivariate ARCH(q) model for quantile residuals

$$\mathbf{R}_{t,\boldsymbol{\theta}} = \mathbf{H}_t^{1/2}\boldsymbol{\varepsilon}_t,$$

where $\varepsilon_t \sim n.i.d.(0,1)$, $\mathbf{H}_t = diag\left[1+\sum_{s=1}^{q}\rho_{1t,s} \quad \cdots \quad 1+\sum_{s=1}^{q}\rho_{nt,s}\right]$, $\rho_{it,s} = \sum_{j=1}^{n}\alpha_{ij,s}\left(R^2_{j,t-s,\boldsymbol{\theta}}-1\right)$ for $i = 1,...,n$, $s = 1,...,q$, and $t = 1,...,T$, with $\mathbf{R}_{t,\boldsymbol{\theta}} = \mathbf{0}$ for $t \leq 0$. Denote with $\mathbf{a}$ the n^2q -vector of parameters $\alpha_{ij,s}$, $i,j = 1,...,n$ and $s = 1,...,q$.

The Jacobian of the transformation is triangular, because $\frac{\partial r_{lt,\theta}}{\partial y_{i_j}} = 0$ for all $l < j$. Therefore, the Jacobian determinant is $\left|\prod_{l=1}^{n}\frac{\partial r_{lt,\theta}}{\partial y_{i_l}}\right| = \prod_{l=1}^{n}\left[\phi\left(r_{lt,\boldsymbol{\theta}}\right)\right]^{-1} f_{i_l,l-1,t-1}(\boldsymbol{\theta},y_{i_l,t}) = \left[\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta},\mathbf{y}_t)$. Thus, the joint density function of the observations

$$f(\boldsymbol{\theta},\mathbf{a},\mathbf{y}) = \prod_{t=1}^{T}\mathbf{H}_t^{-1/2}\cdot\phi\left(\mathbf{H}_t^{-1/2}\mathbf{r}_{t,\boldsymbol{\theta}}\right)\left[\phi\left(\mathbf{r}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta},\mathbf{y}_t),$$

and the log-likelihood function

$$\tilde{l}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{y}) = -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\log(1+\sum_{s=1}^{q}\rho_{it,s}) - \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{r_{it,\boldsymbol{\theta}}^2}{1+\sum_{s=1}^{q}\rho_{it,s}} + \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n} r_{it,\boldsymbol{\theta}}^2 + \sum_{t=1}^{T}\log f_{t-1}(\boldsymbol{\theta}, y_t).$$

Thus, for each $i, j = 1, ..., n$ and $s = 1, ..., q$,

$$\frac{\partial \tilde{l}(\boldsymbol{\theta}, \mathbf{a}, \mathbf{y})}{\partial \alpha_{ij,s}} = \frac{1}{2}\sum_{t=1}^{T}\left[\frac{r_{j,t-s,\boldsymbol{\theta}}^2 - 1}{1+\sum_{s=1}^{q}\rho_{it,s}}\left(\frac{r_{it,\boldsymbol{\theta}}^2}{1+\sum_{s=1}^{q}\rho_{it,s}} - 1\right)\right].$$

Quantile residuals are homoscedastic when $\mathbf{a} = \mathbf{0}$, and the summands in $\frac{\partial \tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial \alpha_{ij,s}} = \frac{1}{2}\sum_{t=1}^{T}\left(r_{j,t-s,\boldsymbol{\theta}}^2 - 1\right)\left(r_{it,\boldsymbol{\theta}}^2 - 1\right)$ are equal to the components of the function $g(\mathbf{r}_{t,\boldsymbol{\theta}})$ of our conditional heteroscedasticity test. Under the null hypothesis, $\frac{\partial \tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial \boldsymbol{\theta}}$ is equal to the score in the main text, and the score function $\tilde{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{0}) = \left[\frac{\partial \tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial \boldsymbol{\theta}'} \quad \frac{\partial \tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial \mathbf{a}'}\right]'$ contains also the function $g(\mathbf{r}_{t,\boldsymbol{\theta}})$. Thus, the LM test based on $\frac{\partial \tilde{l}(\boldsymbol{\theta},\mathbf{0},\mathbf{y})}{\partial \mathbf{a}}$ is identical to our conditional heteroscedasticity test.

B.3 Normality test

We obtain the LM interpretation of the normality test as in Jarque and Bera (1987) and Kalliovirta (2006). Thus, consider the Pearson family of univariate distributions characterized by the differential equation $\frac{d\log(f(u))}{du} = -\frac{u}{b_0+b_1u+b_2u^2}$, $-\infty < u < \infty$, where $f(u)$ is the density of the random variable U, and $\boldsymbol{\beta} = (b_0, b_1, b_2)$ is a parameter vector. When $\boldsymbol{\beta} = (1, 0, 0) \equiv \boldsymbol{\beta}_0$, $f(u)$ is the density of a standard normal distribution. Denote $q(t) = -\int \frac{t}{b_0+b_1t+b_2t^2}dt$, so that $\log f(t) = q(t)+C$, where C is such that $\int_{-\infty}^{\infty} f(u)du = 1$. Then the above differential equation has a solution $f(u) = \exp\{q(u)\}/\int\limits_{-\infty}^{\infty}\exp\{q(t)\}dt$.

For simplicity, we now assume that the components of multivariate quantile residuals $R_{jt,\boldsymbol{\theta}} = \Phi^{-1}(F_{i_j,j-1,t-1}(\boldsymbol{\theta}, Y_{i_j,t}))$ are *independent* and have densities $f_j(u)$ with parameters $\boldsymbol{\beta}_j = [b_{j0}, b_{j1}, b_{j2}]'$ for each $j = 1, ..., n$. The same results follow, if we relax this assumption and use the more general definition for the Pearson family given in Bera and John (1983).

The Jacobian is triangular, because $\frac{\partial r_{lt,\theta}}{\partial y_{i_j}} = 0$ for all $l < j$. Thus, the Jacobian determinant is $\left|\prod\limits_{l=1}^{n} \frac{\partial r_{lt,\theta}}{\partial y_{i_l}}\right| = \prod\limits_{l=1}^{n} \left[\phi\left(r_{lt,\boldsymbol{\theta}}\right)\right]^{-1} f_{i_l,l-1,t-1}(\boldsymbol{\theta}, y_{i_l,t}) = \left[\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t)$. Therefore, the joint density function of the observations is $f(\boldsymbol{\theta}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n, \mathbf{y}) = \prod\limits_{t=1}^{T}\prod\limits_{j=1}^{n} f_j\left(r_{jt,\boldsymbol{\theta}}\right)\left[\phi\left(\mathbf{R}_{t,\boldsymbol{\theta}}\right)\right]^{-1} f_{t-1}(\boldsymbol{\theta}, \mathbf{y}_t)$, and the log-likelihood function

$$\tilde{l}(\boldsymbol{\theta}, \boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_n, \mathbf{y}) = -\sum_{t=1}^{T}\sum_{j=1}^{n}\int \frac{r_{jt,\boldsymbol{\theta}}}{b_{j0} + b_{j1}r_{jt,\boldsymbol{\theta}} + b_{j2}r_{jt,\boldsymbol{\theta}}^2} dr_{jt,\boldsymbol{\theta}}$$

$$-T\sum_{j=1}^{n}\log\int\limits_{-\infty}^{\infty}\exp\{-\int \frac{u}{b_{j0} + b_{j1}u + b_{j2}u^2}du\}du + \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n} R_{it,\boldsymbol{\theta}}^2 + \sum_{t=1}^{T}\log f_{t-1}(\boldsymbol{\theta}, Y_t).$$

Thus, under the null hypothesis (i.e. $\boldsymbol{\beta}_j = \boldsymbol{\beta}_0$ for all $j = 1, ..., n$)

$$\frac{\partial\tilde{l}(\boldsymbol{\theta}, \boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_0, \mathbf{y})}{\partial\boldsymbol{\beta}_j} = -\sum_{t=1}^{T}\begin{bmatrix}\frac{1}{2}\left(r_{jt,\boldsymbol{\theta}}^2 - 1\right) & \frac{1}{3}r_{jt,\boldsymbol{\theta}}^3 & \frac{1}{4}\left(r_{jt,\boldsymbol{\theta}}^4 - 3\right)\end{bmatrix}'$$

for each $j = 1, ..., n$. The summands are, apart from constants, equal to the function $g(\mathbf{r}_{t.\boldsymbol{\theta}})$ of our normality test. Under the null hypothesis, $\frac{\partial\tilde{l}(\boldsymbol{\theta}, \boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_0, \mathbf{y})}{\partial\boldsymbol{\theta}}$ is equal to the score in the main text, and the score

$$\tilde{\mathbf{s}}(\boldsymbol{\theta}, \boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_0) = \begin{bmatrix}\frac{\partial\tilde{l}(\boldsymbol{\theta}, \boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_0, \mathbf{y})}{\partial\boldsymbol{\theta}'} & \frac{\partial\tilde{l}(\boldsymbol{\theta}, \boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_0, \mathbf{y})}{\partial(\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_n')'}\end{bmatrix}'$$

contains also function $g(\mathbf{r}_{t.\boldsymbol{\theta}})$. Thus, the LM test based on the component $\frac{\partial \tilde{l}(\boldsymbol{\theta},\boldsymbol{\beta}_0,\ldots,\boldsymbol{\beta}_0;\mathbf{y})}{\partial(\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_n')'}$ is equal to our normality test.

C Appendix: Factorization of the joint density

We show that for the family of mixtures of multinormal distributions the marginal and conditional distributions belong to the same family of distributions.

Denote with $\mathbf{X}$ $(n \times 1)$ a random vector that follows a mixture of two[10] multinormal distributions. The density of $\mathbf{X}$ is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= p\,(2\pi)^{-n/2}\det(\boldsymbol{\Sigma})^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\,\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \\ &\quad + (1-p)\,(2\pi)^{-n/2}\det(\Omega)^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\nu})'\,\boldsymbol{\Omega}^{-1}(\mathbf{x}-\boldsymbol{\nu})\right\} \\ &= p\cdot MN_n(\boldsymbol{\mu},\boldsymbol{\Sigma}) + (1-p)\cdot MN_n(\boldsymbol{\nu},\boldsymbol{\Omega}), \end{aligned}$$

where $MN_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $MN_n(\boldsymbol{\nu},\boldsymbol{\Omega})$ denote the densities of multinormal distribution with expectations $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, and covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$, respectively. Make a partition on $\mathbf{X} = \begin{bmatrix}\mathbf{X}^{(1)\prime} & \mathbf{X}^{(2)\prime}\end{bmatrix}'$ and conformable partitions on the expectations $\boldsymbol{\mu} = \begin{bmatrix}\boldsymbol{\mu}_1' & \boldsymbol{\mu}_2'\end{bmatrix}'$, $\boldsymbol{\nu} = \begin{bmatrix}\boldsymbol{\nu}_1' & \boldsymbol{\nu}_2'\end{bmatrix}'$ and covariance matrices

$$\boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \begin{bmatrix}\boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22}\end{bmatrix}.$$

If the dimensions of the random vectors $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are k and $n-k$, respectively,

[10] The following applies also to mixtures of multinormal distributions with three or even more components.

then the marginal distribution of $\mathbf{X}^{(2)}$ is a mixture of two normal distributions with density

$$\begin{aligned}
f_{\mathbf{X}^{(2)}}\left(\mathbf{x}^{(2)}\right) &= p\,(2\pi)^{-(n-k)/2}\det(\boldsymbol{\Sigma}_{22})^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{x}^{(2)}-\boldsymbol{\mu}_2\right)'\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)}-\boldsymbol{\mu}_2\right)\right\} \\
&\quad + (1-p)\,(2\pi)^{-(n-k)/2}\det(\boldsymbol{\Omega}_{22})^{-1/2}\exp\left\{-\frac{1}{2}\left(\mathbf{x}^{(2)}-\boldsymbol{\nu}_2\right)'\boldsymbol{\Omega}_{22}^{-1}\left(\mathbf{x}^{(2)}-\boldsymbol{\nu}_2\right)\right\} \\
&= p\cdot MN_{n-k}\left(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22}\right) + (1-p)\cdot MN_{n-k}\left(\boldsymbol{\nu}_2,\boldsymbol{\Omega}_{22}\right).
\end{aligned}$$

This can be seen by integrating the joint density with respect to $\mathbf{x}^{(1)}$ and using well-known properties of the normal distribution.

In order to obtain the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ we define $\boldsymbol{\Sigma}_{11\cdot 2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, the Schur complement of $\boldsymbol{\Sigma}_{22}$. From the identity

$$\begin{bmatrix} \mathbf{I}_k & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_{n-k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

it follows that

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_{n-k} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11\cdot 2}^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_{n-k} \end{bmatrix}.$$

Thus, $\det\left(\boldsymbol{\Sigma}\right) = \det\left(\boldsymbol{\Sigma}_{11\cdot 2}\right)\det\left(\boldsymbol{\Sigma}_{22}\right)$. This and the notation
$\mathbf{x}^{(1)} - \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}_2\right) = \mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right)$ together give

$$\begin{aligned}
\left(\mathbf{x}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right) &= \left(\mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right)\right)'\boldsymbol{\Sigma}_{11\cdot 2}^{-1}\left(\mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right)\right) \\
&\quad + \left(\mathbf{x}^{(2)} - \boldsymbol{\mu}_2\right)'\boldsymbol{\Sigma}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\mu}_2\right).
\end{aligned}$$

The same holds, when we replace $\mathbf{\Sigma}$ and $\left(\mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right)\right)$ with $\mathbf{\Omega}$ and $\mathbf{x}^{(1)} - b\left(\mathbf{x}^{(2)}\right) = \mathbf{x}^{(1)} - \boldsymbol{\nu}_1 - \mathbf{\Omega}_{12}\mathbf{\Omega}_{22}^{-1}\left(\mathbf{x}^{(2)} - \boldsymbol{\nu}_2\right)$. Therefore, we can write the joint density function of $\mathbf{X}$ as

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= p \cdot MN_k\left(\mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right), \mathbf{\Sigma}_{11\cdot 2}\right) \cdot MN_{n-k}\left(\boldsymbol{\mu}_2, \mathbf{\Sigma}_{22}\right) \\ &\quad + (1-p) \cdot MN_k\left(\mathbf{x}^{(1)} - b\left(\mathbf{x}^{(2)}\right), \mathbf{\Omega}_{11\cdot 2}\right) \cdot MN_{n-k}\left(\boldsymbol{\nu}_2, \mathbf{\Omega}_{22}\right). \end{aligned}$$

The conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ is

$$\begin{aligned} f_{\mathbf{X}^{(1)}|\mathbf{X}^{(2)}}(\mathbf{x}^{(1)}|\mathbf{X}^{(2)} = \mathbf{x}^{(2)}) = \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}^{(2)}}\left(\mathbf{x}^{(2)}\right)} &= p\left(\mathbf{x}^{(2)}\right) \cdot MN_k\left(\mathbf{x}^{(1)} - a\left(\mathbf{x}^{(2)}\right), \mathbf{\Sigma}_{11\cdot 2}\right) \\ &\quad + \left(1 - p\left(\mathbf{x}^{(2)}\right)\right) \cdot MN_k\left(\mathbf{x}^{(1)} - b\left(\mathbf{x}^{(2)}\right), \mathbf{\Omega}_{11\cdot 2}\right), \end{aligned}$$

where

$$p\left(\mathbf{x}^{(2)}\right) = \frac{p \cdot MN_{n-k}\left(\boldsymbol{\mu}_2, \mathbf{\Sigma}_{22}\right)}{p \cdot MN_{n-k}\left(\boldsymbol{\mu}_2, \mathbf{\Sigma}_{22}\right) + (1-p) \cdot MN_{n-k}\left(\boldsymbol{\nu}_2, \mathbf{\Omega}_{22}\right)}$$

is a function of $\mathbf{x}^{(2)}$ and the parameters $p, \boldsymbol{\mu}_2, \boldsymbol{\nu}_2, \mathbf{\Sigma}_{22}$, and $\mathbf{\Omega}_{22}$.

Thus, one can solve the quantile residuals for each observation iteratively by solving the parameters of one marginal and one conditional distribution at a time. Each iteration involves the computation of the new expectation vectors $a\left(\mathbf{x}^{(2)}\right)$ and $b\left(\mathbf{x}^{(2)}\right)$, covariance matrices $\mathbf{\Sigma}_{11\cdot 2}$ and $\mathbf{\Omega}_{11\cdot 2}$, and the mixing proportion $p\left(\mathbf{x}^{(2)}\right)$ .These values form the set of parameters for the new conditional distribution. At the same time we solve one marginal distribution, which we then integrate to solve a desired component of the multivariate quantile residual vector at a fixed time point. One can use this procedure for the models in our empirical example whatever the chosen order of conditioning in the multivariate quantile residuals.

In general, one can always compute multivariate quantile residuals with numerical integration. This task becomes very burdensome as the dimension of the time series grows. Therefore, any theory that yields analytical results on the solution of the marginal and conditional distributions is useful. Results, similar to those presented here, can be obtained within the families of elliptical and spherical distributions. See Fang et al. (1990) and the references therein for general treatments on these families.

Acknowledgements:

We thank Markku Lanne, Jukka Nyblom and Bent Nielsen for comments that improved the presentation of this work. Part of this research was done while the first author was visiting the Department of Mathematical Sciences, University of Copenhagen. We are grateful to the Academy of Finland (LK and PS), the Yrjö Jahnsson foundation (LK), the Okobank Group Research Foundation (LK and PS), and the Finnish Foundation for Advancement of Securities Markets for financial support (LK).

References

Bai, J. and Z. Chen, 2008, Testing multivariate distributions in GARCH models, Journal of Econometrics 143, 19–36.

Basawa, I.V. and D.J. Scott, 1983, Asymptotic Optimal Inference for Non-ergodic Models, Lecture Notes in Statistics, Springer-Verlag, New York.

Bec, F., Rahbek, A. and N. Shephard, 2008, The ACR model: A multivariate dynamic mixture autoregression, Oxford Bulletin of Economics and Statistics 70(5), 583–618.

Bera, A. and S. John, 1983, Tests for multivariate normality with Pearson alternatives, Communications in Statistics - Theory and Methods 12(1), 103–117.

Berkowitz, J., 2001, Testing density forecasts, with applications to risk management, Journal of Business and Economic Statistics 19(4), 465–474.

Chesher, A. and R.J. Smith, 1997, Likelihood ratio specification tests, Econometrica 65(3), 627–646.

Chitturi, R.V., 1974, Distribution of residual autocorrelations in multiple autoregressive schemes, Journal of the American Statistical Association 69(348), 928–934.

Clements, M.P. and J. Smith, 2000, Evaluating the forecast densities of linear and non-linear models: Application to output growth and unemployment, Journal of Forecasting 19, 255–276.

Clements, M.P. and J. Smith, 2002, Evaluating multivariate forecast densities: a comparison of two approaches, International Journal of Forecasting 18, 397–407.

Cox, D.R. and D.V. Hinkley, 1974, Theoretical statistics, Chapman and Hall, London.

Cox, D.R. and E.J. Snell, 1968, A general definition of residuals, Journal of the

Royal Statistical Society, Series B 30(2), 248–275.

Diebold, F.X., J. Hahn and A.S. Tay, 1999, Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange, The Review of Economics and Statistics 81(4), 661–673.

Doornik, J.A. and H. Hansen, 2008, An omnibus test for univariate and multivariate normality, Oxford Bulletin of Economics and Statistics 70(s1), 927–939.

Douc, R., E. Moulines, and T. Rydén, 2004, Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime, The Annals of Statistics 32(5), 2254–2304.

Dunn, P.K. and G.K. Smyth, 1996, Randomized quantile residuals, Journal of Computational and Graphical Statistics 5(3), 236–244.

Fang, K., S. Kotz, and K. Ng, 1990, Symmetric Multivariate and Related Distributions, Chapman and Hall.

Francq, C. and J.-M. Zakoian, 2010, QML estimation of a class of multivariate GARCH models without moment conditions on the observed process. MPRA Paper No. 20779.

Hong, Y. and H. Li, 2005, Nonparametric specification testing for continuous-time models with applications to term stucture of interest rates, The Review of Financial Studies 18(1), 37–84.

Hosking, J., 1981, Lagrange-multiplier tests of multivariate time-series models, Journal of the Royal Statistical Society. Series B (Methodological) 43(2), 219–230.

Jarque, C.M. and A.K. Bera, 1980, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, Economics Letters 6, 255–259.

Jarque, C.M. and A.K. Bera, 1987, A test for normality of observations and regression residuals, International Statistical Review 55(2), 163–172.

Johansen, S., 1995, Likelihood-based Inference in Cointegrated Vector Autoregressive Models, Oxford University Press.

Kalliovirta, L., 2006, Misspecification tests based on quantile residuals. HECER Discussion Paper, No 124.

Lanne, M. and H. Lütkepohl, 2010, Structural Vector Autoregressions with nonnormal residuals, Journal of Business and Economic Statistics 28(1), 159–168.

Lanne, M., H. Lütkepohl and K. Maciejowska, 2010, Structural Vector Autoregressions with Markov switching, Journal of Economic Dynamics and Control 34(2), 121–131.

Lanne, M. and P. Saikkonen, 2007, A multivariate generalized orthogonal factor GARCH model, Journal of Business and Economic Statistics 25(1), 61–75.

Ling, S. and W.K. Li, 1997, Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors, Journal of Time Series Analysis 18(5), 447–464.

Lomnicki, Z.A., 1961, Tests for departure from normality in the case of linear stochastic processes, Metrika 4, 37–62.

McLeod, A.I., 1978, On the distribution of residual autocorrelations in Box-Jenkins models, Journal of the Royal Statistical Society. Series B (Methodological) 40(3), 296–302.

McLeod, A.I. and W.K. Li, 1983, Diagnostic checking ARMA time series models using squared-residual autocorrelations, Journal of Time Series Analysis 4(4), 269–273.

Paap, R., R. Segers and D. van Dijk, 2009, Do leading indicators lead peaks more than troughs?, Journal of Business and Economic Statistics 27(4), 528–543.

Palm, F.C. and P.J.G. Vlaar, 1997, Simple diagnostic procedures for modelling financial time series, Allgemeines Statistisches Archiv 81, 85–101.

Rosenblatt, M., 1952, Remarks on a multivariate transformation, The Annals of Mathematical Statistics 23(3), 470–472.

Smith, J.Q., 1985, Diagnostic checks of non-standard time series models, Journal of Forecasting 4, 283–291.

Sweeting, T.J., 1980, Uniform asymptotic normality of the maximum likelihood estimator, The Annals of Statistics 8(6), 1375–1381.

van der Vaart, A.W., 1998, Asymptotic Statistics, Cambridge University Press.

Table 1: Estimation results of Model 3 based on equation (15).

Parameter				
α_1	0.011		α_2	0.311
	(0.007)			(0.083)
β_1	0.986		β_2	0.616
	(0.007)			(0.082)
$\mathbf{B}$	0.025	3.271	0.436	-0.090
	(0.205)	(0.210)	(0.353)	(0.286)
	-9.322	-1.329	1.047	-1.602
	(3.809)	(0.405)	(2.268)	(0.655)
	9.294	-1.867	-0.102	0.437
	(3.797)	(0.423)	(2.149)	(0.619)
	-0.072	0.010	-0.776	1.509
	(0.105)	(0.056)	(0.574)	(0.317)
$\mathbf{\Phi}$	0.266	0.050	0.307	0.471
	(0.059)	(0.008)	(0.057)	(0.102)
p	0.130			
	(0.021)			

NOTE: The estimated standard errors are in the parentheses. They are computed using the cross-product of the first derivatives of the log-likelihood function.

Table 2: Estimation results of Model 4 based on equation (15).

Parameter				
α_{11}	0.020		α_{21}	0.273
	(0.012)			(0.061)
β_{11}	0.979		β_{21}	0.644
	(0.014)			(0.055)
			α_{22}	0.168
				(0.022)
			β_{22}	0.824
				(0.022)
$\mathbf{B}$	0.190	2.738	0.719	0.053
	(0.123)	(0.308)	(0.185)	(0.251)
	-8.914	-1.120	1.196	-1.827
	(2.915)	(0.252)	(0.850)	(0.800)
	8.748	-1.604	-0.504	0.553
	(2.856)	(0.293)	(0.692)	(0.718)
	-0.092	0.007	-0.806	1.490
	(0.061)	(0.035)	(0.373)	(0.220)
$\mathbf{\Phi}$	0.111	0.139	0.201	0.388
	(0.037)	(0.037)	(0.047)	(0.095)
p	0.059			
	(0.012)			

NOTE: The estimated standard errors are in the parentheses. They are computed using the cross-product of the first derivatives of the log-likelihood function.

Table 3: P-values of test statistics in percentages and values of the information criteria computed for Models 1-4.

Model	A_3^J	H_3^J	N^J	A_3	H_3	N	AIC	BIC
1	10	0	0	$3.4{\cdot}10^{-4}$	0	0	2789	2882
2	58	0.2	$6.2{\cdot}10^{-3}$	$7.9{\cdot}10^{-3}$	0	0	2495	2602
3	56	0.3	$8.9{\cdot}10^{-4}$	$5.6{\cdot}10^{-3}$	0	0	2483	2599
4	52	1.1	$1{\cdot}10^{-2}$	$5.3{\cdot}10^{-4}$	0	0	2328	2453

NOTE: We computed the autocorrelation test based on three lags (A_3), the conditional heteroscedasticity test based on three lags (H_3), and the normality test (N) with the simulated covariance matrix estimate $\tilde{\boldsymbol{\Omega}}_T$. The superscript J indicates tests based on *joint* quantile residuals. P-value 0 means a value $< 1{\cdot}10^{-5}$.

Table 4: Rejection frequencies of tests with Model S.1.

	A_3^J		H_3^J		N^J		A_3		H_3		N	
	Model S.1	simulated										
T	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**
250	4.7	1.0	5.6	1.7	4.6	1.5	4.8	1.0	7.6	3.4	4.7	1.5
500	5.5	1.0	4.5	1.4	5.5	1.8	6.4	1.0	6.8	2.3	5.2	2.1
750	5.5	0.8	5.9	1.3	5.5	0.8	5.6	1.6	6.4	1.9	5.8	1.8
1000	5.6	1.0	5.4	1.6	5.6	1.0	5.2	1.1	5.8	1.6	5.9	1.7
	Model S.2	simulated										
250	4.9	1.0	4.3	2.1	6.4	2.4	4.6	0.7	9.1	3.3	10.2	4.9
500	4.8	1.1	5.4	1.7	7.2	2.6	4.9	0.7	8.8	3.6	13.8	6.9
750	3.9	0.5	5.0	1.5	8.7	2.3	5.8	1.2	9.3	2.8	16.4	7.4
1000	4.2	0.8	5.8	1.8	9.5	2.9	5.7	0.9	9.4	3.6	19.8	9.5
	Model S.5	simulated										
250	5.9	1.7	7.7	3.6	6.4	3.3	7.4	2.1	47.5	35.6	16.2	9.7
500	6.2	2.0	10.8	5.1	9.3	4.2	7.2	1.9	73.7	61.0	24.3	16.5
750	7.1	1.9	11.1	5.4	10.5	4.6	7.7	1.8	89.9	81.1	29.7	19.1
1000	7.2	1.9	12.6	6.0	11.0	5.0	8.5	2.4	95.4	90.7	36.2	25.7

NOTE: For each sample size we provide the percentage of rejections at 5% and 1% levels. Results are based on 2000 replications, and we computed the test statistics using the simulated covariance matrix estimate $\tilde{\mathbf{\Omega}}_T$. We estimated Model S.1 with OLS, thus, the normality test statistics N^J and N lack the term $r^2_{t,\boldsymbol{\theta}} - 1$. The parameter values are: 1) Model S.1 $\boldsymbol{\mu} = (0,0)$ and $\mathbf{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 3 \end{bmatrix}$; 2) Model S.2 $\boldsymbol{\mu}_1 = (0,0)$, $\boldsymbol{\mu}_2 = (1,1)$, $\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 3 \end{bmatrix}$, $\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 4 \end{bmatrix}$, and $c = \Phi^{-1}(0.15)$; and 3) Model S.5 $\alpha = 0.25$ and $(\mathbf{W}')^{-1} = \begin{bmatrix} 0.49 & -1.32 \\ 2.27 & 0.5 \end{bmatrix}$.

Table 5: Rejection frequencies of tests with Model S.3.

	A_3^J		H_3^J		N^J		A_3		H_3		N	
	Model S.3		simulated									
T	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**
250	3.9	0.8	4.6	1.4	3.8	1.3	6.5	1.8	7.3	2.6	5.5	2.1
500	4.1	1.2	5.1	1.2	4.1	1.4	4.9	1.2	6.2	2.2	5.6	2.3
750	4.9	1.3	5.8	2.0	4.8	1.5	5.3	1.2	7.7	2.2	5.4	1.9
1000	4.4	1.0	5.3	1.5	5.0	1.0	5.8	1.4	7.0	1.8	5.7	1.6
	Model S.4		simulated									
250	7.3	1.7	25.3	15.1	5.1	1.9	49.6	39.4	46.0	32.9	6.9	3.1
500	7.8	2.3	46.3	29.3	7.2	2.3	31.5	22.8	77.9	64.5	10.4	4.3
750	9.1	2.4	60.7	44.0	8.0	2.9	25.1	16.9	92.3	84.7	10.3	4.9
1000	9.4	2.9	74.0	58.2	7.9	2.6	19.5	11.8	98.2	94.8	11.4	5.5

NOTE: For each sample size we provide the percentage of rejections at 5% and 1% levels. Results are based on 2000 replications, and we computed the test statistics using the simulated covariance matrix estimate $\tilde{\boldsymbol{\Omega}}_T$. We estimated Model S.3 with OLS, thus, the normality test statistics N^J and N lack the term $r^2_{t,\boldsymbol{\theta}} - 1$. The parameter values are: 1) Model S.3 $\boldsymbol{\mu} = (0,0)$, $\mathbf{A} = \begin{bmatrix} 0.9 & 0.2 \\ 0 & 0.6 \end{bmatrix}$, and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$; and 2) Model S.4 $\boldsymbol{\mu} = (0,0)$, $\mathbf{A}_1 = \mathbf{0}$, $\mathbf{A}_2 = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.6 \end{bmatrix}$, $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$, and $c = \Phi^{-1}(0.35)$.

Table 6: Rejection frequencies of tests with Model S.5.

	A_3^J		H_3^J		N^J		A_3		H_3		N	
	Model S.5		simulated									
T	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**
250	5.3	1.3	5.3	2.0	4.8	1.7	6.3	1.9	9.9	4.9	6.7	2.3
500	5.7	1.5	6.6	2.2	5.6	1.6	5.6	1.4	11.0	3.7	6.3	2.2
750	4.6	1.3	6.1	1.7	4.9	1.6	5.6	1.1	8.8	3.0	5.6	1.2
1000	5.3	0.9	5.4	1.4	5.1	1.5	5.0	1.4	9.8	3.2	5.8	1.2

NOTE: For each sample size we provide the percentage of rejections at 5% and 1% levels. Results are based on 2000 replications, and we computed the test statistics using the simulated covariance matrix estimate $\tilde{\mathbf{\Omega}}_T$. In Model S.5 the parameter values are $\alpha = 0.25$ and $(\mathbf{W}')^{-1} = \begin{bmatrix} 0.49 & -1.32 \\ 2.27 & 0.5 \end{bmatrix}$.
We imposed some restrictions on parameters to guarantee successful estimation. We used the actual parameter values as starting values for the estimation algorithm, because the optimization of the likelihood function of Model S.5 was difficult in smaller samples.

Table 7: Size properties of different tests with Model S.2.

	BJ^p		N^q		N		N^J		BC	
T	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**	**5**	**1**
500	100	100	0	0	6.7	2.6	5.7	2.3	19.7	9.5
1000	100	100	0	0	6.7	2.1	6.3	2.0	10.1	5.2

NOTE: For each sample size we provide the percentage of rejections at 5% and 1% levels based on 2000 replications. The test statistics BJ^p is the normality test of Bera and John (1983), a multivariate version of the test considered in Jarque and Bera (1987). This test employs Pearson's residuals and ignores the effect of parameter estimation. The effect of parameter estimation is also ignored in the normality test N^q that employs quantile residuals and is computed with $\mathbf{\Omega}$ equal to $\mathbf{H}$ defined in (12). We computed our normality tests, N (based on multivariate quantile residuals) and N^J (based on joint quantile residuals), using the simulated covariance matrix estimate $\tilde{\mathbf{\Omega}}_T$. The test statistic BC is the pooled test statistic of Bai and Chen (2008). In Model S.2, the parameter values are $\boldsymbol{\mu}_1 = (0, 0)$, $\boldsymbol{\mu}_2 = (12, 12)$, $\mathbf{\Sigma}_1 = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 3 \end{bmatrix}$, $\mathbf{\Sigma}_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 4 \end{bmatrix}$, and $c = \Phi^{-1}(0.15)$.
We imposed some restrictions on parameters to guarantee successful estimation. We used the actual parameter values as starting values for the estimation algorithm, because the optimization of the likelihood function of Model S.2 was difficult in smaller samples.

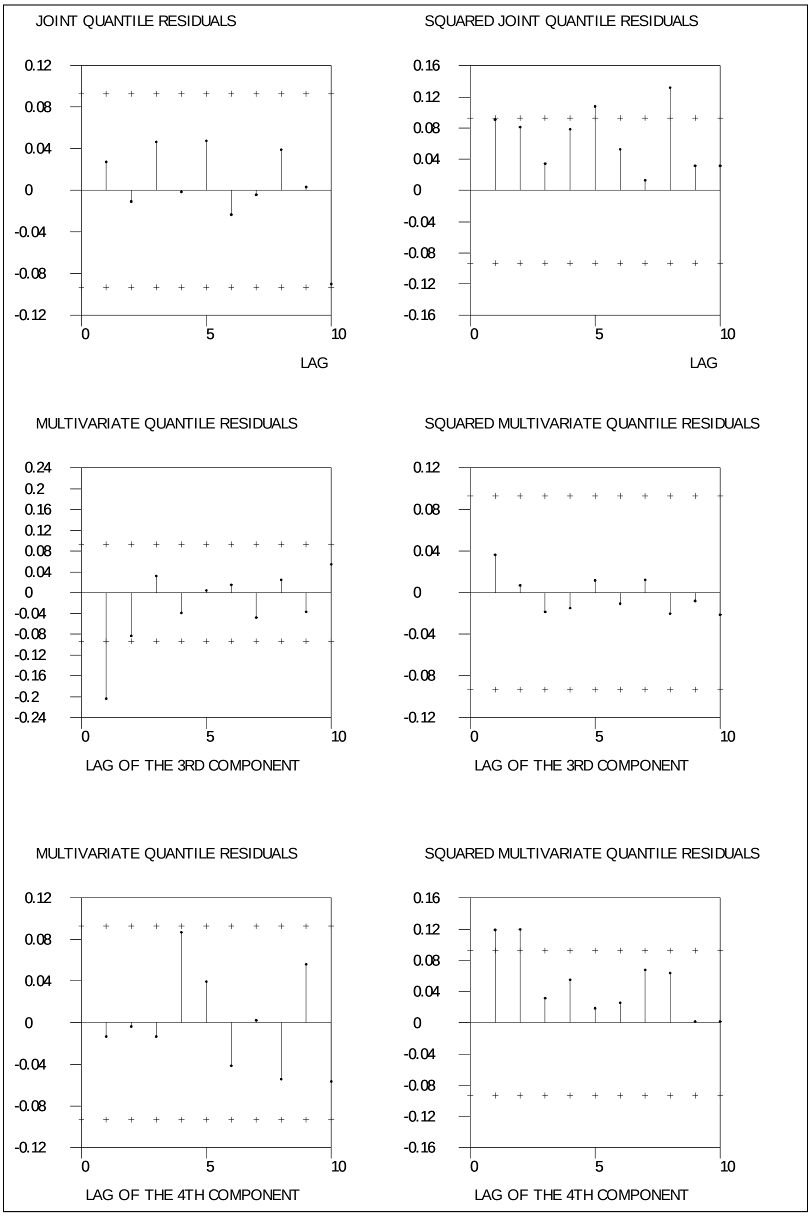

Figure 1: Autocovariance functions of joint and multivariate quantile residuals and squared joint and multivariate quantile residuals of Model 4 divided by their approximate standard errors. The standard errors base on the simulated covariance matrix estimate $T^{-1}\tilde{\mathbf{\Omega}}_T$. Approximate 99% critical bounds are denoted with plus signs for each lag.

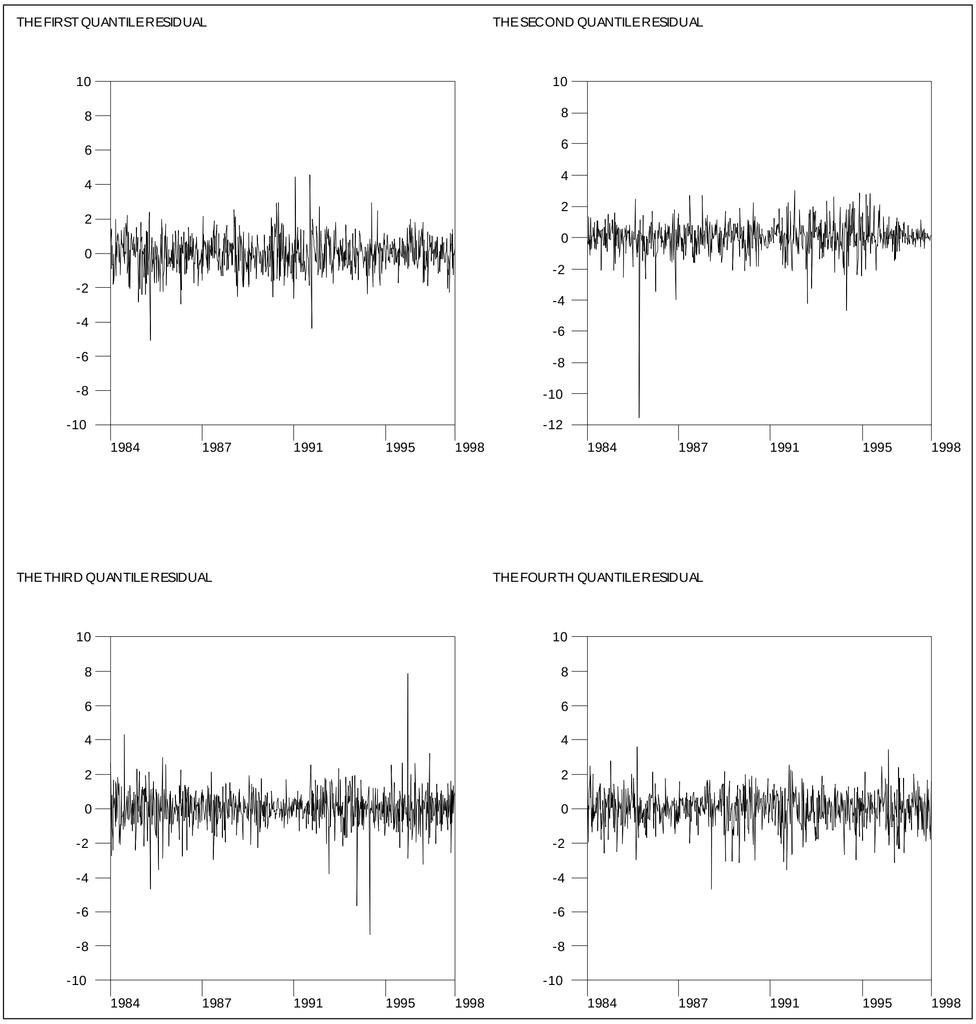

Figure 2: Residual series for two factor model under normality (Model 1).

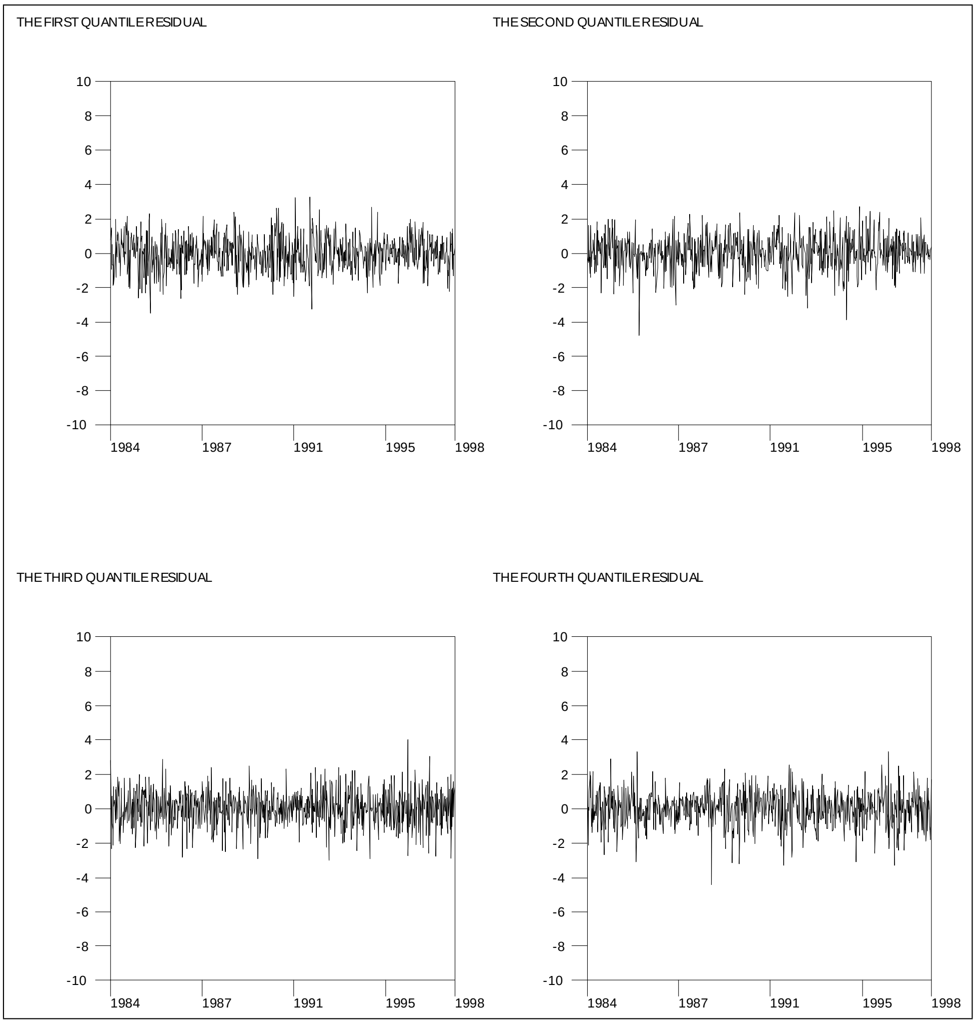

Figure 3: Quantile residual series for two factor mixture normal model (Model 4)