Quantile Residuals for Multivariate Models*

Leena Kalliovirta
http://www.valt.helsinki.fi/blogs/lkvaisan/
Department of Economics,
Research Unit of Economic Structures and Growth, HECER,
and
Department of Mathematics and Statistics
University of Helsinki
December 5, 2007

Abstract

So-called quantile residuals are generalized for multivariate time series models. These residuals are applicable for example to nonlinear time series models based on mixture distributions for which conventional residuals are not well suited. We show that under mild regularity conditions multivariate quantile residuals are approximately independent with standard normal distribution. A general framework of obtaining tests based on smooth functions of quantile residuals and the likelihood function is formulated. In addition to conventional ergodic models our framework also allows for certain non-ergodic models. The tests based on the framework can be thought of as pure significance type tests, and they take uncertainty caused by parameter estimation properly into account. Under regularity conditions the test statistics are asymptotically chi-square distributed. We use the framework to develop misspecification tests aimed at detecting non-normality, serial correlation, and conditional heteroscedasticity in quantile residuals. An empirical example on exchange rate series illustrates the application of Multivariate Generalized Orthogonal Factor GARCH models and the tests of the paper.

JEL classification: C32, C52

Keywords: Generalized autoregressive conditional heteroscedasticity, Multivariate Generalized Orthogonal Factor GARCH model, Nonlinear time series model, Quantile residuals, Misspecification test, Pure significance test.

*The author would like to thank Pentti Saikkonen for pointing out and supervising this work. The author is responsible for any remaining errors. Part of this research was done while visiting the Department of Mathematical Sciences, University of Copenhagen. The author is grateful to the Research Unit of Economic Structures and Growth (RUESG) in the University of Helsinki and the Yrjö Jahnsson foundation for financial support.
1 Introduction

Checking the specification of a statistical model usually involves both statistical tests and graphical methods based on residuals. However, in some recent models based on mixtures of distributions conventional residuals, often called Pearson residuals, are not convenient or ideal. The approach taken in this paper makes use of residuals sometimes referred to as quantile residuals. These residuals can be defined for any parametric model by using the cumulative distribution function of the observations. 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), whereas Palm and Vlaar (1997) speak of normalized residuals. Smith (1985) calls them normal forecast transformed residuals. Quantile residuals are defined by two transformations. First, the estimated cumulative distribution function implied by the model is used to transform the observations into approximately independent uniformly distributed random variables. This is the so-called probability integral transformation. Second, the inverse of the cumulative distribution function of the standard normal distribution is used to get variables which are approximately independent with standard normal distribution. 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.

In this paper, we study multivariate quantile residuals and their asymptotic properties in a general likelihood framework. We give regularity conditions under which a central limit theorem holds for smooth functions of quantile residuals. This result can be used to obtain misspecification tests which, under correct specification, have limiting $\chi^2$—distributions. Our approach is theoretically sound in that it takes the uncertainty caused by parameter estimation into account and it is applicable also with non-ergodic models where estimators have different rates of convergence. The approach is illustrated by deriving tests aimed at detecting non-normality, serial correlation, and conditional heteroscedasticity in quantile residuals. Tests for other departures from the characteristic properties of quantile residuals can be obtained similarly from the general framework. Because the tests of the paper are derived without any particular alternative hypothesis in mind, they can be thought of as pure significance tests introduced by Cox and Hinkley (1974).

Quantile residuals have been considered in many papers (see Kalliovirta (2006), and references therein). Most of them concentrate on out-of-sample forecast evaluation of a univariate model and, unlike we, do not give proper theoretical justification for the employed procedures. This paper generalizes the work of these authors by showing how misspecification tests based on multivariate quantile residuals can be obtained in a general likelihood framework. We use the idea suggested in Diebold, Hahn, and Tay (1999), Clements and Smith (2000), and Clements and Smith (2002) in the context of multivariate density forecast evaluation, and base
our analysis on two different types of multivariate quantile residuals. One type is derived using marginal and conditional distribution functions at each time point, and the other, henceforth referred to as joint quantile residuals, is based on the product of marginal and conditional distribution functions.

The diagnostic methods developed are applied to a Multivariate Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen (2007). A modified version of their model is estimated and analyzed. Conventional residuals are not well suited for these models and, therefore, the value of the log-likelihood function and model selection criteria such as AIC (Akaike 1973) or BIC (Schwarz 1978) can be employed to discriminate between candidate models. Our approach provides a useful addition to model selection criteria. In particular, it can be used to support graphical analysis and to formally compare the goodness of fit between models based on different structural or distributional assumptions.

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 gives an empirical example, and Section 5 contains concluding remarks.

2 Quantile residuals

2.1 Definition - univariate data

Let \( y = \{y_1, ..., y_T\} \) be a vector of observations with density function \( f(\theta_0, y) \), where \( \theta_0 \in \Theta \) is the unknown true parameter value. Denote with \( \mathcal{P} = \{f(\theta, y) : \theta \in \Theta \subset \mathbb{R}^k, y \in \mathbb{R}^T\} \) the collection of potential models for \( y \). For each \( f : \Theta \times \mathbb{R}^T \rightarrow \mathbb{R}_+ \), we can write

\[
f(\theta, y) = \prod_{t=1}^{T} f_{t-1}(\theta, y_t),
\]

where \( f_{t-1}(\theta, y_t) = f(\theta, y_t|G_{t-1}) \), \( t \in \{1, ..., T\} \), \( y_t \in \mathbb{R} \) is the conditional density function given \( G_{t-1} = \sigma(Y_0, Y_1, ..., Y_{t-1}) \), the sigma-algebra generated by the random variables \( \{Y_0, Y_1, ..., Y_{t-1}\} \), i.e., the history at a time \( t \). The random vector \( Y_0 \) represents the needed initial values.

According to Dunn and Smyth (1996), the theoretical quantile residual is defined by

\[
R_{t, \theta} = \Phi^{-1}(F_{t-1}(\theta, Y_t)),
\]

and the observed quantile residual is \( r_{t, \hat{\theta}_T} = \Phi^{-1}(F_{t-1}(\hat{\theta}_T, y_t)) \), where \( \Phi^{-1}(\cdot) \) is the inverted cumulative distribution function of the standard normal distribution, \( F_{t-1}(\theta, y_t) = \int_{-\infty}^{y_t} f_{t-1}(\theta, u)du \) is the conditional cumulative distribution function of \( y_t \), and \( \hat{\theta}_T \) is an estimate of \( \theta_0 \). The
second transformation leading to normal distribution is recommended inter alia by Dunn and Smyth (1996) and Berkowitz (2001). Normal variation is that which most people have practice interpreting graphically and existing tools for testing i.i.d. standard normality are available.

If the data are independently and identically distributed, then formula (2) is a special case of the “crude” residual of Cox and Snell (1968). Note also that the quantile residuals of a standard linear model with normal errors are identical to conventional residuals or the so-called Pearson residuals. Therefore, quantile residuals can be seen as a generalization of the Pearson residuals. For more discussion on previous literature on quantile residuals and Pearson residuals, see Kalliovirta (2006).

2.2 Definition - multivariate data

Let \( y_1, ..., y_T \) be vector valued observations, and let the conditional density function \( f_{t-1}(\theta, y_t) \) be defined for every value \( y_t = \left[ y_{1t} \cdots y_{nt} \right]' \). The collection of potential models is denoted by \( \mathcal{P} = \{ f(\theta, y) : \theta \in \Theta \subset \mathbb{R}^k, y \in \mathbb{R}^{nT} \} \).

If the components of \( y_t \) are independent, the quantile residuals are straightforwardly extended to the vector case. Since the conditional cumulative distribution function of \( y_t \) has the product form \( F_{t-1}(\theta, y_t) = \prod_{j=1}^{n} F_{j,t-1}(\theta, y_{jt}) \), where \( F_{j,t-1}(\theta, y_{jt}) \) is the marginal distribution function of the \( j \)th component, the transformation (2) can be done component-wise.

If the components of \( y_t \) are dependent, the quantile residuals are defined as follows. Write the conditional density function of \( y_t \) in the product form

\[
f_{t-1}(\theta, y_t) = \prod_{j=1}^{n} f_{ij,j-1,t-1}(\theta, y_{ij,t})
\]

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 \( \mathcal{A}_{j-1} = \sigma \{ Y_{i1,t}, ..., Y_{ij-1,t} \} \) generated by the component variables. We interpret \( f_{i1,0,t-1}(\theta, y_{i1,t}) = f_{i1,t-1}(\theta, y_{i1,t}) \), and \( F_{ij,j-1,t-1}(\theta, y_{ij,t}) = \int_{-\infty}^{y_{ij,t}} f_{ij,j-1,t-1}(\theta, u) du \). Thus, the vector of theoretical quantile residuals at time point \( t \) takes the form

\[
R_{t,\theta} = \begin{bmatrix}
R_{1t,\theta} \\
R_{2t,\theta} \\
\vdots \\
R_{nt,\theta}
\end{bmatrix} = \begin{bmatrix}
\Phi^{-1}(F_{1,t-1}(\theta, Y_{1,t})) \\
\Phi^{-1}(F_{2,t-1}(\theta, Y_{2,t})) \\
\vdots \\
\Phi^{-1}(F_{n,t-1}(\theta, Y_{n,t}))
\end{bmatrix}.
\]

(4)

This vector is not unique, but can be formed in \( n! \) different ways. However, the results presented in this paper do not depend on the chosen order of conditioning. The vector of observed quantile residuals at time point \( t \) is obtained by replacing \( \theta \) with \( \hat{\theta}_T \), an estimate of \( \theta_0 \), in (4).
Model evaluation can also be based on univariate statistics, congruent with Clements and Smith (2000), we call theoretical joint quantile residuals

\[ Q_{t,\theta} = \Phi^{-1}(Z_{t,\theta}), \]  

where

\[ Z_{t,\theta} = X_{t,\theta} \sum_{k=0}^{n-1} \frac{(-1)^k}{k!} (\ln X_{t,\theta})^k \]  

with \[ X_{t,\theta} = \prod_{j=1}^{n} F_{j,i-1,t-1}(\theta, Y_{i,t}). \]

Clements and Smith (2000) and Clements and Smith (2002) have studied this transformation in a bivariate case and applied it to evaluate forecast densities. The general form of joint quantile residuals has not been previously suggested. The observed joint quantile residuals at time point \( t \) we obtain by replacing \( \theta \) with \( \hat{\theta}_T \), an estimate of \( \theta_0 \), in (5).

2.3 Theoretical properties

It is now shown that, under mild regularity conditions, quantile residuals have properties that make them useful in model evaluation: Lemma 2 gives that observed multivariate quantile residuals are asymptotically independently multnormally distributed, if the estimated model is correctly specified. Lemma 3 yields that the same result holds for the observed joint quantile residuals. The following Condition 1 is both necessary and sufficient for Lemmas 2 and 3 to hold. Unless otherwise stated all limit statements assume that \( T \to \infty \). The symbols \( \xrightarrow{W} \) and \( \xrightarrow{P} \) signify weak convergence and convergence in probability, respectively.

**Condition 1** Let the following assumptions hold.

1. The collection \( \mathcal{P} \) is correctly specified, i.e., \( f(\theta_0, y) \in \mathcal{P} \).
2. \( f_{t-1} : \Theta \times \mathbb{R}^n \to \mathbb{R} \) is a continuous conditional density function for all \( \theta \in \Theta \) and \( t = 1, \ldots, T \).
3. \( \hat{\theta}_T \) is an estimator of \( \theta_0 \) such that \( \hat{\theta}_T \xrightarrow{P} \theta_0 \).

**Lemma 2** Under Condition 1,

a) the distribution of the vector of quantile residuals \( [R_{t,0} \cdots R_{t,0}']' \) is multivariate standard normal, where \( R_{t,0} \) is as in (4) with \( \theta = \theta_0 \).

b) for any \( H \) fixed, the distribution of \( [R_{t,0} \cdots R_{t,0}']' \) is asymptotically multivariate standard normal, where \( R_{t,0} \) is as in (4) with \( \theta = \hat{\theta}_T \), and

c) for any \( s \geq 1 \), \( R_{t+s,0} \) is independent of \( \{Y_1, \ldots, Y_1\} \).

The proof is given in Appendix A.
Lemma 3 Under Condition 1,

a) the distribution of the vector \( \begin{bmatrix} Q_{1,θ_0} & \cdots & Q_{T,θ_0} \end{bmatrix}' \) is multivariate standard normal, where \( Q_{t,θ_0} \) is as in (5) with \( θ = θ_0 \),

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

c) for any \( s \geq 1 \), \( Q_{t+s,θ_0} \) is independent of \( \{ Y_1, ..., Y_t \} \).

The proof is given in Appendix A.

A correct model specification can, therefore, be checked by testing whether the observed quantile residuals are normally and independently distributed. This holds for both the joint and the multivariate versions of the residuals. Especially, all diagnostic checks based on their graphs are available. Note that the joint quantile residuals are always univariate whatever dimension the data has. This can be useful, in particular, when the dimension is large. In the following Sections we develop a framework and some tests based on it in order to have theory that also yields confidence bound for some of these graphs used in the literature.

2.4 Preliminaries on Maximum Likelihood estimation

It is now assumed that conditional on initial values, the log-likelihood function of the sample takes the form

\[
l_T(θ, y) = \sum_{t=1}^{T} l_t(θ, y_t) = \sum_{t=1}^{T} \log f_{t-1}(θ, y_t).
\]

The following Condition 4 is sufficient for the consistency and asymptotic normality of a local maximizer of the conditional likelihood function. These results are needed to derive the limiting distribution of a general statistic which can be used to obtain tests based on quantile residuals. We use \( ||·|| \) to signify the Euclidean norm.

Condition 4 Let the following assumptions hold.

1. \( Θ \subset \mathbb{R}^k \) is an open set.

2. The model is correctly specified, i.e., \( f(θ_0, y) \in \mathcal{P} \).

3. For every \( (θ, x) \in Θ \times D \), where \( D \subset \mathbb{R}^n \), and every \( t = 1, ..., T \), \( f_{t-1}(θ, x) > 0 \) and the second partial derivatives \( \frac{∂^2}{∂θ_i∂θ_j} f_{t-1}(θ, x) \), \( i, j = 1, ..., k \), exist and are continuous.

4. There exist nonrandom \( k \times k \) matrices \( A_T(θ) \), nonsingular and continuous in \( θ \), and such that \( \{A_T(θ)\}^{-1} \to 0 \) and, for all \( c > 0 \),

\[
\sup_{θ \in \mathcal{M}_{T,c}} \| \{A_T(θ_0)\}^{-1} \{A_T(θ)\} - I_k \| \to 0,
\]
and
\[ \sup_{\theta \in M_{T,c}} \left\{ \left\{ A_T(\theta_0) \right\}^{-1} \left[ B_T(\theta) - B_T(\theta_0) \right] \left\{ A_T(\theta_0) \right\}^{-1} \right\}' \overset{P}{\longrightarrow} 0, \]

where \( M_{T,c} = \{ \theta \in \Theta : \| \left\{ A_T(\theta_0) \right\}'(\theta - \theta_0) \| \leq c \} \) and \( B_T(\theta) = -\frac{\partial^2}{\partial \theta \partial \theta'} l_T(\theta, Y) = -\left[ \sum_{t=1}^{T} \frac{\partial^2 l_t(\theta, Y_t)}{\partial \theta \partial \theta'} \right]_{i,j=1}^{k} . \)

(5) Denote with \( S_T(\theta) = \frac{\partial}{\partial \theta} l_T(\theta, Y) = \sum_{t=1}^{T} \frac{\partial}{\partial \theta} l_t(\theta, Y_t) \) the score function and \( W_T(\theta_0) = \left\{ A_T(\theta_0) \right\}^{-1} B_T(\theta_0) \left\{ A_T(\theta_0) \right\}^{-1} \) a scaled Hessian matrix. There exists a (possibly) random matrix \( I(\theta_0) \) such that
\[ \begin{bmatrix} W_T(\theta_0) \\ \left\{ A_T(\theta_0) \right\}^{-1} S_T(\theta_0) \end{bmatrix} \overset{W}{\rightarrow} \begin{bmatrix} I(\theta_0) \\ I(\theta_0)^{1/2} Z \end{bmatrix}, \]

where \( \mathbb{P}_{\theta_0}(I(\theta_0) > 0) = 1 \), i.e., \( I(\theta_0) \) is positive definite \( \mathbb{P}_{\theta_0} \)-a.e., and \( Z \sim N_k(0, I_k) \) is independent of \( I(\theta_0) \).

Condition 4(3) imposes fairly standard regularity conditions on the conditional density functions. Combined with Condition 4(1) it implies the applicability of the Mean-Value Theorem for the score function in any convex set \( A \subset \Theta \). Condition 4(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 \( \theta_0 \). The introduction of matrices \( A_T(\theta) \) in the condition allows the framework to be used also when the model is non-ergodic of the type explained after Condition 7. Condition 4(5) is a high level assumption needed to obtain asymptotic mixed normality of the maximum likelihood estimator (MLE). In standard cases Conditions 4(4) and 4(5) can typically be verified by using an appropriate uniform law of large numbers and a martingale central limit theorem, respectively. Note that Condition 4(1) guarantees the standard assumption that the MLE is an inner point. The correct model specification is necessary for Theorem 5 below and for testing purposes.

In practice, one has to be able to specify how the sequence of nonrandom matrices \( A_T(\theta) \) depends on \( \theta \). Sweeting (1980) suggests that if \( \mathbb{E}_{\theta_0}(I(\theta_0)) < \infty \) then \( \{ A_T(\theta_0) \} = \{ \mathbb{E}_{\theta_0}(I(\theta_0)) \}^{1/2} \) can be used. In the ergodic case, where all components of \( \hat{\theta}_T \) converge at the rate \( \sqrt{T} \), one can use \( A_T(\theta) = \sqrt{T} I_k \) for all \( \theta \). Other examples of choices of \( A_T(\theta) \) are given e.g. by Basawa and Scott (1983).

We define the maximum likelihood estimator \( \hat{\theta}_T \) to be any local maximizer of \( l_T(\theta; y) \) when such a maximum exists, and \( +\infty \) otherwise.

**Theorem 5** Under Condition 4 there exists a sequence of local maximizers \( \hat{\theta}_T \) such that \( \left\{ \{ A_T(\theta_0) \}'(\hat{\theta}_T - \theta_0) \right\}_{T \in \mathbb{N}} \) is bounded in probability and
\[ \left\{ A_T(\theta_0) \right\}^{-1} S_T(\theta_0) - W_T(\theta_0) \left\{ A_T(\theta_0) \right\}'(\hat{\theta}_T - \theta_0) \overset{P}{\longrightarrow} 0. \]
The proof is given in Sweeting (1980).

2.5 Central limit theorem for transformed quantile residuals

Now the general framework for obtaining tests based on multivariate and joint quantile residuals is developed. The function \( g \) below is used to transform the quantile residuals. With different choices of this function one can construct test statistics for different potential departures from the characteristic properties of quantile residuals.

Condition 4 and the following Conditions 6 and 7 together yield the theorems needed to establish asymptotic distributions for our test statistics. As in Condition 4, we denote \( M_{T,c} = \{ \theta \in \Theta : \| \{ A_T(\theta_0) \} (\theta - \theta_0) \| \leq c \} \).

**Condition 6** Let one of the following assumptions hold.

(1a) \( g : \mathbb{R}^{nm} \rightarrow \mathbb{R}^l \) is a continuously differentiable function such that \( \mathbb{E}(g(U_{t,\theta_0})) = 0 \), where \( U_{t,\theta_0} = [R_{t,\theta_0} \cdots R_{t-m+1,\theta_0}]' \in \mathbb{R}^{nm} \) is a vector of quantile residuals defined in (4).

(1b) \( g : \mathbb{R}^{m} \rightarrow \mathbb{R}^l \) is a continuously differentiable function such that \( \mathbb{E}(g(U_{t,\theta_0})) = 0 \), where \( U_{t,\theta_0} = [Q_{t,\theta_0} \cdots Q_{t-m+1,\theta_0}]' \in \mathbb{R}^{m} \) is a vector of joint quantile residuals defined in (5).

**Condition 7** Let the vector \( U_{t,\theta_0} \) and the function \( g \) be as in Condition 6 and let the following assumptions hold.

(1) For all \( c > 0 \)

\[
\sup_{\theta \in M_{T,c}} \left\| \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta'} g(U_{t,\theta}) - G \right\|_F \to 0,
\]

\[
\sup_{\theta \in M_{T,c}} \left\| \frac{1}{T} \sum_{t=1}^{T} g(U_{t,\theta}) g(U_{t,\theta})' - H \right\|_F \to 0,
\]

and

\[
\sup_{\theta \in M_{T,c}} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t,\theta}) \left[ \frac{\partial}{\partial \theta} l_t(\theta,Y_t) \right]' \left[ \{ A_T(\theta_0) \}^{-1} \right]' - \Psi \right\|_F \to 0,
\]

where \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t,\theta_0})) \) and \( H = \mathbb{E}(g(U_{t,\theta_0}) g(U_{t,\theta_0})') \) exist and are finite and \( \Psi \) is a (possibly) random matrix. Moreover, the matrix \( H \) is positive definite.

(2) There exists a nonrandom \( k \times k \) matrix \( J \) such that

\[
\left\| \sqrt{T} \left[ \{ A_T(\theta_0) \}^{-1} \right]' - J \right\| \to 0
\]
and
\[
\begin{bmatrix}
W_T(\theta_0) \\
\vdots \\
\{A_T(\theta_0)\}^{-1} S_T(\theta_0) \\
\frac{1}{T} \sum_{t=1}^{T} g(U_t, \theta_0)
\end{bmatrix}
\xrightarrow{W}
\begin{bmatrix}
I(\theta_0) \\
\vdots \\
\Sigma^{1/2} Z
\end{bmatrix},
\]
where $Z \sim N_{k+l}(0, I_{k+l})$ is independent of $\Sigma = \begin{bmatrix} I(\theta_0) & \Psi' \\ \Psi & H \end{bmatrix}$, a positive definite matrix with (possibly random) elements defined above.

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

Condition 6 allows test statistics to be defined by any continuously differentiable transformation of the multivariate or joint quantile residuals with zero expectation. A large number of different hypotheses can therefore be tested within this framework. Condition 7(1) imposes uniform convergence in probability on special compact sets similar to that in Condition 4(4). Together these two conditions define the covariance matrix $\Sigma$ in Condition 7(2). The joint weak convergence assumption in Condition 7(2) can be verified by using an appropriate central limit theorem. It contains Condition 4(5) as a special case. Condition 7(2) also specifies the type of non-ergodic models that can be used within the framework. Those components of $\hat{\theta}_T$ that converge faster than at the rate $\sqrt{T}$ will have corresponding elements in the matrix $J$ equal to 0. When the framework is applied the matrix $J$ has to be specified. Condition 7(3) complements Condition 4(3).

Now we can state a central limit theorem from which the limiting distributions of our tests are obtained.

**Theorem 8** Under Conditions 4, 6, and 7
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_t, \hat{\theta}_T) \xrightarrow{W} \Omega^{1/2} U,
\]
where $U \sim N(0, I_l)$ is independent of

\[
\Omega = \begin{bmatrix} G J I(\theta_0)^{-1} : I_l \end{bmatrix} \cdot \Sigma \cdot \begin{bmatrix} I(\theta_0)^{-1} J' G' \\ I_l \end{bmatrix}
= G J I(\theta_0)^{-1} J' G' + \Psi I(\theta_0)^{-1} J' G' + G J I(\theta_0)^{-1} \Psi' + H.
\]

The proof is given in Appendix A.

The first three terms in the asymptotic covariance matrix $\Omega$ take the uncertainty caused by parameter estimation into account. Note that Condition 7(2) implies that the random matrix $\Omega$ is positive definite. In the extreme non-ergodic case, when all components of $\hat{\theta}_T$
converge at a faster rate than $\sqrt{T}$, the rank of matrix $J$ is zero, and then $J$ is a null matrix and $\Omega \equiv H$. Another time $\Omega \equiv H$ is obtained, is when the employed model implies that $G = 0$. These two cases have a useful practical implication: Then the uncertainty caused by parameter estimation can be ignored in tests obtained using our framework.

The following lemma provides an estimator for the covariance matrix $\Omega$ needed when a test based on a chosen function $g$ is derived. The estimator might not be consistent, but it always converges weakly to $\Omega$. This lemma is convenient for most nonlinear models for which the components of $\Omega$ are difficult or impossible to obtain analytically.

**Lemma 9** Let Conditions 4, 6, and 7 hold, and define

$$\hat{\Omega}_T = \left[ \hat{G}_T \cdot J \cdot W_T(\hat{\theta}_T)^{-1} : I \right] \cdot \hat{\Sigma}_T \cdot \left[ W_T(\hat{\theta}_T)^{-1} \cdot J' \cdot \hat{G}'_T \right],$$

where $\hat{G}_T = \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \theta_t} g(U_t, \hat{\theta}_T)$, $W_T(\hat{\theta}_T) = \left\{ A_T(\hat{\theta}_T) \right\}^{-1} B_T(\hat{\theta}_T) \left\{ A_T(\hat{\theta}_T) \right\}^{-1}'$, and

$$\hat{\Sigma}_T = \begin{bmatrix} W_T(\hat{\theta}_T) & \hat{\Psi}_T \\ \hat{\Psi}_T' & \hat{H}_T \end{bmatrix}$$

with $\hat{\Psi}_T = \frac{1}{\sqrt{T}} \sum_{t=1}^T g(U_t, \hat{\theta}_T) \left[ \frac{\partial}{\partial \theta_t} l_t(\hat{\theta}_T, Y_t) \right]' \left\{ A_T(\hat{\theta}_T) \right\}^{-1}$ and $\hat{H}_T = \frac{1}{T} \sum_{t=1}^T g(U_t, \hat{\theta}_T) g(U_t, \hat{\theta}_T)'$. Then

$$\hat{\Omega}_T \rightarrow^W \Omega.$$

The proof is given in Appendix A. The numerical value of $\hat{\Omega}_T$ is easily obtained by the employed estimation algorithm, only knowledge of the estimates $\hat{\theta}_T$, the matrix $W_T(\hat{\theta}_T)^{-1}$, the likelihood function $l_t(\hat{\theta}_T, Y_t)$, and the derivatives $\frac{\partial}{\partial \theta_t} g(U_t, \hat{\theta}_T)$ and $\frac{\partial}{\partial \theta_t} l_t(\hat{\theta}_T, Y_t)$ are needed. The needed derivatives are easy to compute numerically if their analytic values are difficult to obtain or not known. Explicit expressions for the derivatives $\frac{\partial}{\partial \theta_t} R_{t, \theta}$ and $\frac{\partial}{\partial \theta_t} Q_{t, \theta}$ are provided in Appendix A, Lemma 11.

Based on the results of Theorem 8 and Lemma 9 we can deduce that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T g(U_t, \hat{\theta}_T)' \cdot \hat{\Omega}_T^{-1} \cdot \frac{1}{\sqrt{T}} \sum_{t=1}^T g(U_t, \hat{\theta}_T) W_t U' \Omega^{1/2} \Omega^{-1} \Omega^{1/2} U = U' U.$$

This yields a general test statistic

$$S = \frac{1}{T - m + 1} \sum_{t=m}^T g(U_t, \hat{\theta}_T)' \cdot \hat{\Omega}_T^{-1} \cdot \sum_{t=m}^T g(U_t, \hat{\theta}_T) \approx \chi^2(l),$$

where $m$ and $l$ are the dimensions defined in Condition 6.
A result often referred to as Delta Method states that if a function $h : \mathbb{R}^l \to \mathbb{R}^r$ is differentiable at $\delta \in \mathbb{R}^l$, \{\(Z_T\)\}_{T \in \mathbb{N}} is a sequence of random vectors taking values in $\mathbb{R}^l$, and $\sqrt{T}(Z_T - \delta) \xrightarrow{W} Z$, then $\sqrt{T}(h(Z_T) - h(\delta)) \xrightarrow{W} \hat{h}(\delta)Z$, where $\hat{h}(\delta)$ is the value of the linear map (matrix) $\hat{h}$ defined by the partial derivatives of $h$ at $\delta$. (See e.g. van der Vaart (1998) for more details.) The application of the Delta Method in conjunction with the central limit theorem (6) constitutes our general framework of obtaining tests.

3 Tests based on Quantile Residuals

In the following sections we illustrate how our general framework can be used to derive misspecification tests. The tests are developed by using a strategy that does not require specification of an alternative hypothesis. Tests of this type were introduced by Cox and Hinkley (1974) who called them pure significance tests. The absence of an alternative hypothesis is a source of both weakness and strength of a pure significance test. Against a given alternative it is usually possible to find a specific test which will outperform a pure significance test. The relatively high power of a specific test may, however, be bought at the price of a lack of sensitivity to other alternatives, so that it may be inferior to a pure significance test when used in an inappropriate situation (Godfrey 1991).

We shall derive separate misspecification tests which can be used to test for non-normality, serial correlation, and conditional heteroscedasticity of multivariate and joint quantile residuals. Instead of these separate tests we could have chosen to employ the approaches, e.g., in Jarque and Bera (1980) or Hong and Li (2005), and use our framework to derive a joint test for these three features. Because the sensitiveness of the individual tests against different misspecifications varies, outcomes of separate tests may give useful hints of the reasons of a potential misspecification. Moreover, separate tests can be used to complement the information provided by graphical methods such as histograms, QQ-plots, autocorrelation and cross correlation functions of quantile residuals and squared quantile residuals. The tests derived here yield confidence bounds for graphs, and thereby justify their use.

A correct model specification is assumed below so that $R_{t, \theta_0} \sim NID(0, I_n)$ and $Q_{t, \theta_0} \sim NID(0, 1)$ hold.

3.1 Multinormality tests

Three multinormality tests are developed under Condition 6(1a), more specifically we choose $U_{t, \theta} = R_{t, \theta}$. Our tests make use of ideas in e.g. Lomnicki (1961), Bowman and Shenton (1975), Kiefer and Salmon (1983), Jarque and Bera (1987), Doornik and Hansen (1994), and Bai and Ng (2005). The null hypothesis employed is based on the first four moments, i.e.,

$$H_0 : \mathbb{E} \left[ R_{jt, \theta_0}^2 \, R_{jt, \theta_0}^3 \, R_{jt, \theta_0}^4 \, \right] = 0 \quad \text{for all} \ j \in \{1, ..., n\} \ \text{and} \ t,$$
which holds if \( R_{jt, \theta_0} \sim NID(0, 1) \). 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 (1994).

Our normality test is based on function \( g : \mathbb{R}^n \rightarrow \mathbb{R}^{4n} \) (see Theorem 8) with

\[
g(u_t, \theta) = \left[ g_1(r_{1t, \theta})' \cdots g_n(r_{nt, \theta})' \right]',
\]

where \( g_j(r_{jt, \theta}) = \left[ r_{jt, \theta}^2 - 1 \ r_{jt, \theta}^3 - 3r_{jt, \theta} \ r_{jt, \theta}^4 - 6r_{jt, \theta}^2 + 3 \right]' \). Using properties of the standard multinormal distribution we see that \( \mathbb{E}(g(U_{t, \theta_0})) = 0 \) with matrix \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t, \theta_0})) \) given in Derivatives section in Appendix A, and

\[
H = \mathbb{E}(g(U_{t, \theta_0})g(U_{t, \theta_0}')) = I_n \otimes diag \left[ 1 \ 2 \ 6 \ 24 \right],
\]

where \( \otimes \) denotes the Kronecker product. Further, the function \( g \) is continuously differentiable. Thus, assuming the conditions of Theorem 8 we get the asymptotic result

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t, \hat{\theta}_T}) \xrightarrow{W} N(0, H).
\]

Define \( \hat{\Omega}_T \), an estimator of \( \Omega \), by replacing the last term in Lemma 9 with the expression of \( H \) given in (8). Note that the matrices \( J \) and \( A_T(\hat{\theta}_T) \) are specified by properties of the estimator \( \hat{\theta}_T \). Using this estimator and the preceding asymptotic result we then obtain the test statistic

\[
N_1 = \frac{1}{T} \cdot \sum_{t=1}^{T} g(u_{t, \hat{\theta}_T})' \cdot \hat{\Omega}_T^{-1} \cdot \sum_{t=1}^{T} g(u_{t, \hat{\theta}_T}) \xrightarrow{H_0} \chi^2(4n).
\]

Applying the Delta Method one can proceed as in Doornik and Hansen (1994) and derive a test based on the sample skewness and kurtosis of quantile residuals. The resulting test statistic, denoted \( N_2 \), nests the test of Doornik and Hansen (1994). If the matrix \( \Omega \) defined in (7) simplifies to the matrix \( H \), the tests are equal. A third normality test, \( N_3 \) derived below, has been found to perform better at least in univariate case (see Kalliovirta (2006)). This test has similarities to the normality test discussed in Bai and Ng (2005). Define the function \( h = \begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix} : \mathbb{R}^{4n} \rightarrow \mathbb{R}^{2n} \), where components are \( h_j : \mathbb{R}^4 \rightarrow \mathbb{R}^2 \),

\[
h_j(x_{j1}, x_{j2}, x_{j3}, x_{j4}) = \left[ x_{j3} + 3x_{j1} \ x_{j4} + 3x_{j2} \right]', \quad j = 1, \ldots, n.
\]

Clearly, each \( h_j \) is continuously differentiable in a neighborhood of \((0, 0, 0, 0)\) and straightfor-

\[\text{H}^1\text{Compared to earlier normality tests based on Pearson residuals we have included the term } r_{jt, \theta}. \text{ The addition of this term has improved small sample properties of the test for nonlinear models. It has to be removed, if the mean of quantile residuals of the estimated model is automatically zero. In that case the matrix defined in (8) is not positive definite and the asymptotic result does not hold. This happens e.g. when models can be estimated using ordinary least squares.} \]
ward calculations give
\[
\dot{h}_j(0,0,0,0) = \frac{\partial}{\partial x_j} h_j(0,0,0,0) = \begin{bmatrix} 3 & 0 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix}
\]
and \( h(0) = 0 \).

Therefore,
\[
\dot{h} = \frac{\partial}{\partial x} h(0) = I_n \otimes \begin{bmatrix} 3 & 0 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{bmatrix}.
\]

Using (9) and the Delta Method, one obtains
\[
\sqrt{T} \cdot h \left( \frac{1}{T} \sum_{t=1}^{T} g(U_t, \tilde{\theta}_T) \right) \xrightarrow{W} N(0, \dot{h} \Omega \dot{h}'),
\]
where \( \dot{h} \) has rank \( 2n \) and, therefore, \( \dot{h} \Omega \dot{h}' \) is of rank \( 2n \) as long as \( \Omega \) is of full rank. Our third normality test statistic is defined as
\[
N_3 = T \cdot h \left( \frac{1}{T} \sum_{t=1}^{T} g(U_t, \tilde{\theta}_T) \right) \cdot \left[ \dot{h} \Omega_T \dot{h}' \right]^{-1} \cdot h \left( \frac{1}{T} \sum_{t=1}^{T} g(U_t, \tilde{\theta}_T) \right) \overset{H_0}{\sim} \chi^2(2n),
\]
where \( h \left( \frac{1}{T} \sum_{t=1}^{T} g(U_t, \tilde{\theta}_T) \right) = \left[ h_1 \left( \frac{1}{T} \sum_{t=1}^{T} g_1(r_{1t}, \tilde{\theta}_T) \right)' \cdots h_n \left( \frac{1}{T} \sum_{t=1}^{T} g_n(r_{nt}, \tilde{\theta}_T) \right)' \right]' \) with
\[
h_j \left( \frac{1}{T} \sum_{t=1}^{T} g_j(r_{jt}, \tilde{\theta}_T) \right) = \left[ \frac{1}{T} \sum_{t=1}^{T} r_{jt, \tilde{\theta}_T}^3 - 3 \frac{1}{T} \sum_{t=1}^{T} r_{jt, \tilde{\theta}_T}^2 \right].
\]

Test statistics \( N_1, N_2, \) and \( N_3 \) can also be based on the sample estimate \( \hat{H}_T = \frac{1}{T} \sum_{t=1}^{T} g(U_t, \tilde{\theta}_T) g(U_t, \tilde{\theta}_T)' \) of the matrix \( H \) (see Theorem 8 and Lemma 9). These versions of the test statistics are denoted by \( N_1^*, N_2^*, \) and \( N_3^* \).

If we want to test normality of the joint quantile residuals, we choose the null hypothesis
\[
H_0 : E \left[ Q_{t, \theta_0}^2 Q_{t, \theta_0}^2 - 1 \ Q_{t, \theta_0}^3 Q_{t, \theta_0}^4 - 3 \right] = 0 \quad \text{for all} \ t.
\]
This holds if \( Q_{t, \theta_0} \sim NID(0,1) \). Assuming Condition 6(1b) we construct univariate forms of the normality tests obtained above. Thus, we set \( U_{t, \theta} = Q_{t, \theta} \), and make appropriate changes in the matrices \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(Q_{t, \theta_0})) \), \( \Psi \), and \( H \) as well as in their empirical counterparts, \( \hat{G}_T \), \( \hat{\Psi}_T \), and \( \hat{H}_T \). The resulting tests are denoted by \( N_1^J, N_2^J, N_3^J, N_1^{J*}, N_2^{J*}, \) and \( N_3^{J*} \).
3.2 Test for Autocorrelation

The autocorrelation test is first developed under Condition 6(1a), i.e., we test for potential serial correlation in multivariate quantile residuals. Therefore, it is assumed that $U_{t,\theta} = \begin{bmatrix} R_{t,\theta}^r & \cdots & R_{t-K_1,\theta}^r \end{bmatrix}'$ for some $K_1$, and the general null hypothesis

$$H_0 : \mathbb{E}(R_{t,\theta_0}R_{t-s,\theta_0}') = 0 \quad \text{for all } t \text{ and } s > 0$$

is considered. The test is based on the statistics

$$\hat{C}_s = \frac{1}{T-s} \sum_{t=1+s}^{T} r_{t,\theta_T} r_{t-s,\theta_T}' \quad s = 1, \ldots, K_1, \quad K_1 << T,$$

i.e., uncentered sample autocovariance matrices of quantile residuals. These are reasonable estimators because theoretically $\mathbb{E}(R_{t,\theta_0}) = 0$, even though in general $\mathbb{E}(\Omega_T) = \frac{1}{T} \sum_{t=1}^{T} r_{t,\theta_T} \neq 0$. The potential inadequacy in the model is assumed to be reflected by the first $K_1$ autocovariance matrices. A similar test statistic formulated in terms of autocorrelations has been used e.g. in Chitturi (1974).

The continuously differentiable function $g : \mathbb{R}^{n(K_1+1)} \to \mathbb{R}^{n^2K_1}$ is defined as

$$g(u_{t,\theta}) = vec \begin{bmatrix} r_{t,\theta} r_{t-1,\theta}' & \cdots & r_{t,\theta} r_{t-K_1,\theta}' \end{bmatrix}.$$ 

Then clearly $\mathbb{E}(g(U_{t,\theta_0})) = 0$ with matrix $G = \mathbb{E}(\frac{\partial}{\partial \theta_0} g(U_{t,\theta_0}))$ given in Derivatives section in Appendix A. Properties of the standard multinormal distribution yield

$$H = \mathbb{E}(g(U_{t,\theta_0})g(U_{t,\theta_0})') = \mathbb{I}_{n^2K_1}.$$

Using Theorem 8 and the estimator for $\Omega$ given in Lemma 9 with the last term replaced with $\mathbb{I}_{n^2K_1}$, gives the test statistic

$$A_{K_1} = \frac{1}{T-K_1} \sum_{t=1+K_1}^{T} g(u_{t,\theta_T})' \cdot \hat{\Omega}_T^{-1} \cdot \sum_{t=1+K_1}^{T} g(u_{t,\theta_T}) \overset{H_0}{\sim} \chi^2(n^2K_1).$$

An alternative version of this test statistic, denoted by $A_{K_1}'$, is formed by using the sample estimate $\hat{\Omega}_T = \frac{1}{T} \sum_{t=1}^{T} g(u_{t,\theta_T})g(u_{t,\theta_T})'$ for $H$.

In addition to the overall test statistic $A_{K_1}$ or $A_{K_1}'$ it may also be useful to consider individual autocovariance and cross covariance estimates $\hat{c}_{ij}$. A large value of $\hat{c}_{ij}$ compared to its approximate standard error obtained from the relevant diagonal element of the matrix $T^{-1}\hat{\Omega}_T$ suggests model inadequacy. Therefore, a useful model criticism procedure is to plot $\hat{c}_{ij1}, \ldots, \hat{c}_{ijr}$ divided by their standard errors for each $j$ and some value $r$, and compare them with their approximate 95% critical bounds, as already suggested in McLeod (1978). 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.

The autocorrelation test can also be based on joint quantile residuals, i.e., we assume Condition 6(1b). Choose \( U_{t, \theta} = \left[ Q_{t, \theta} \cdots Q_{t-K_1, \theta} \right]' \) and test the null hypothesis

\[
H_0 : \mathbb{E}(Q_{t, \theta}Q_{t-s, \theta}) = 0 \quad \text{for all } t \text{ and } s > 0,
\]

by computing the above autocorrelation test in univariate form. Thus, replace \( r_{t, \theta}, \ldots, r_{t-K_1, \theta} \) with \( q_{t, \theta}, \ldots, q_{t-K_1, \theta} \), and make the appropriate changes in the matrices \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t, \theta_0})) \), \( \Psi \), and \( H \) as well as their empirical counterparts, \( \hat{G}_T \), \( \hat{\Psi}_T \), and \( \hat{H}_T \). Tests obtained in this way are denoted by \( A_{K_1}^J \) and \( A_{K_1}^J \).

### 3.3 Test for Conditional Heteroscedasticity

The test of potential conditional heteroscedasticity is first developed using multivariate quantile residuals, i.e., under Condition 6(1a). Therefore, the general null hypothesis

\[
H_0 : \mathbb{E}(R_{i,t, \theta_0}^2, R_{j,t-s, \theta_0}^2) = 0 \quad \text{for } i, j \in \{1, \ldots, n\}, \text{ all } t, \text{ and } s > 0
\]

is considered. Modifying of the ideas suggested in McLeod and Li (1983) and Ling and Li (1997), the test is based on the statistics

\[
\hat{d}_{ijs} = \frac{1}{T-s} \sum_{t=1+s}^T \left( r_{i,t, \theta_T}^2 - 1 \right) \left( r_{j,t-s, \theta_T}^2 - 1 \right) \quad i, j \in \{1, \ldots, n\}, \ s = 1, \ldots, K_2, \ K_2 << T,
\]

i.e., sample autocovariances of squared multivariate quantile residuals. Note that theoretically \( \mathbb{E}(R_{i,t, \theta_0}^2) = 1 \), even though \( \frac{1}{T} \sum_{t=1}^T r_{i,t, \theta_T}^2 \neq 1 \). As in the previous section, a relatively small number of autocovariances is assumed to sufficiently reflect the potential inadequacy in the model.

We assume that \( U_{t, \theta} = \left[ R_{t, \theta}^I \cdots R_{t-K_2, \theta}^I \right]' \), and according to the preceding discussion we define the continuously differentiable function \( g : \mathbb{R}^{n(K_2+1)} \rightarrow \mathbb{R}^{n^2K_2} \),

\[
g(u_{t, \theta}) = vec \left[ v_{t, \theta}v_{t-1, \theta}' \cdots v_{t, \theta}v_{t-K_2, \theta}' \right]
\]

with \( v_{t-s, \theta} = \left[ r_{1,t-s, \theta}^2 - 1 \cdots r_{n,t-s, \theta}^2 - 1 \right]' \), \( s = 0, \ldots, K_2 \). Then \( \mathbb{E}(g(U_{t, \theta_0})) = 0 \) with the matrix \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t, \theta_0})) \) given in Derivatives section in Appendix A. The properties of the standard multinormal distribution give

\[
H = \mathbb{E}(g(U_{t, \theta_0})g(U_{t, \theta_0}')) = 4I_{n^2K_2}.
\]

Using the estimator \( \hat{\Omega}_T \) given in Lemma 9 with the last term replaced with \( 4I_{n^2K_2} \) yields the
test statistic

\[ H_{K_2} = \frac{1}{T - K_2} \sum_{t=1+K_2}^{T} g(u_t, \hat{\theta}_T) \cdot \hat{\Omega}_T^{-1} \cdot \sum_{t=1+K_2}^{T} g(u_t, \hat{\theta}_T) \approx \chi^2(n^2 K_2). \] (10)

An alternative test statistic, denoted by \( H_{K_2}^* \), can again be based on the sample estimate

\[ \hat{H}_T = \frac{1}{T} \sum_{t=1}^{T} g(u_t, \hat{\theta}_T) g(u_t, \hat{\theta}_T)' \] for \( H \).

It is also useful to supplement the overall test statistic \( H_{K_2} \) or \( H_{K_2}^* \) by plotting individual autocovariance estimates \( \hat{d}_{ij,s} \) of the squared quantile residuals. Again, approximate standard errors can be obtained from the square roots of the diagonal elements of matrix \( T^{-1} \hat{\Omega}_T \).

The heteroscedasticity test can also be based on joint quantile residuals. Under Condition 6(1b) the null hypothesis

\[ H_0 : \mathbb{E}(Q_{t, \theta_0}^2, Q_{t-s, \theta_0}^2) = 0 \] for all \( t \) and \( s > 0 \),

can be tested with the above heteroscedasticity test computed in the following form. Choose \( U_{t, \theta} = [Q_{t, \theta}, \ldots, Q_{t-K_2, \theta}]' \), thus \( r_{t, \theta}, \ldots, r_{t-K_2, \theta} \) are replaced with \( q_{t, \theta}, \ldots, q_{t-K_2, \theta} \). Further, make the appropriate changes in the matrices \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t, \theta_0})) \), \( \Psi \), and \( H \) as well as in their empirical counterparts, \( \hat{G}_T, \hat{\Phi}_T \), and \( \hat{H}_T \). Tests obtained in this way are denoted by \( H_{K_1}^J \) and \( H_{K_1}^{J*} \).

4 Empirical example on Multivariate Generalized Orthogonal Factor GARCH 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 find a relatively small number of factors that can depict the multivariate conditional variance structure of the data adequately. We illustrate how the multivariate and joint quantile residuals can be used to compare the properties of candidate models.

4.1 The Model

The Multivariate Generalized Orthogonal Factor GARCH model of Lanne and Saikkonen (2007) is slightly generalized here. Let \( y_t \) be a \( n \) dimensional process that has conditional density function

\[ f_{t-1}(y_t) = p(2\pi)^{-\frac{n}{2}} \det(WH_1t \Psi_1^{-1} W')^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} y_t' (WH_1t \Psi_1^{-1} W')^{-1} y_t \right\} \]  
\[ + (1-p)(2\pi)^{-\frac{n}{2}} \det(WH_2t \Psi_2^{-1} W')^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} y_t' (WH_2t \Psi_2^{-1} W')^{-1} y_t \right\}, \] (11)
where \( p \in (0, 1) \), \( \mathbf{W} \) is a nonsingular parameter matrix, \( \Psi_1 = p \mathbf{I}_n + (1 - p) \Psi \), \( \Psi_2 = \Psi_1 \Psi^{-1} \), and \( \Psi = \text{diag}[\psi_1, \ldots, \psi_n] \) are parameter matrices, and \( \mathbf{H}_{1t} \) and \( \mathbf{H}_{2t} \) are stochastic diagonal matrices defined below. The matrix \( \Psi \) is assumed to have positive elements, so \( \psi_i > 0 \) for all \( i \in \{1, \ldots, n\} \). The stochastic diagonal matrices have the form
\[
\mathbf{H}_{jt} = \text{diag}[\mathbf{V}_{jt} : \mathbf{I}_{n-t}] \quad \text{with} \quad \mathbf{V}_{jt} = \text{diag}[v_{it}^{(j)} \ldots v_{nt}^{(j)}],
\]
where
\[
v_{it}^{(j)} = (1 - \alpha_{ji} - \beta_{ji}) + \beta_{ji} v_{i,t-1}^{(j)} + \alpha_{ji} (b'_i y_{t-1})^2, \quad i = 1, \ldots, r, \quad \text{and} \quad j = 1, 2. \tag{12}
\]
The parameters \( \alpha_{1i}, \alpha_{2i}, \beta_{1i}, \beta_{2i} \) in (12) satisfy \( \alpha_{ji} > 0, \beta_{ji} \geq 0, \) and \( \alpha_{ji} + \beta_{ji} < 1 \) for all \( i \) and \( j \). These parameter restrictions imply that, under some further assumptions, the process \( y_t \) is strictly stationary and ergodic and also second order stationary (see Lanne and Saikkonen (2007) and the references therein). The parameter vector \( b'_i \) is the \( i \)th row of the parameter matrix \( \mathbf{B}' = \mathbf{W}^{-1} \). Note that the intercept terms in (12) are normalized in such a way that that the components of \( \mathbf{B}' y_t \) have unit unconditional variance. Thus, the conditional distribution of \( y_t \) is a mixture of normal distributions with \( \mathbb{E}(y_t) = 0 \) and \( \text{cov}_{t-1}(y_t) = p \mathbf{WH}_{1t} \Psi^{-1}_1 \mathbf{W}' + (1 - p) \mathbf{WH}_{2t} \Psi^{-1}_2 \mathbf{W}' \).

The model has an alternative representation as a function of parameters and two unobservable random variables
\[
y_t = \mathbf{W} \left( I(s_t = 0) \cdot \mathbf{H}_{1t}^{1/2} \mathbf{Psi}^{-1/2}_1 + I(s_t = 1) \cdot \mathbf{H}_{2t}^{1/2} \mathbf{Psi}^{-1/2}_2 \right) \varepsilon_t, \tag{13}
\]
where \( I(\cdot) \) is the indicator function, \( \varepsilon_t \sim \text{NID}(0, \mathbf{I}_n) \) and \( s_t \) is an i.i.d. random variable with \( \text{Pr}(s_t = 0) = p \) and \( \text{Pr}(s_t = 1) = 1 - p \). The processes \( \{\varepsilon_t\} \) and \( \{s_t\} \) are independent. This model clearly yields the conditional density (11) for \( y_t \). This representation is easily compared with the model in Lanne and Saikkonen (2007), where
\[
y_t = \mathbf{W} \mathbf{H}_{t}^{1/2} \left( I(s_t = 0) + I(s_t = 1) \cdot \Psi^{1/2} \right) \Psi^{-1/2} \varepsilon_t.
\]
The model is identified up to multiplying the columns of \( \mathbf{B} \) by minus one. It is assumed that the above process is stationary under the assumptions made on the parameters. Therefore, it is also reasonable to assume that the high level conditions imposed in Section 2 are satisfied with the choice \( A_T(\theta) = \sqrt{T} \mathbf{I}_k \) for all \( \theta \) and \( \mathbf{J} = \mathbf{I}_k \).

We analyze the same data as in Lanne and Saikkonen (2007): 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 years 1984–1997. That makes 782 observations. The maximum likelihood estimation of the parameters was carried out using Gauss 5.0 and the cmll-package. The initial values of \( \hat{\mathbf{H}}_{11} \) and \( \hat{\mathbf{H}}_{21} \) were calculated using the sample variances of

\[
\text{...}
\]
Therefore the initial values of $\hat{H}_{k1}$ are different for each iteration. Before estimation the series is centered in order the mean to be zero.

### 4.2 Comparison of the estimated models

Four different Multivariate Generalized Orthogonal Factor GARCH models are compared with each other using quantile residuals. The estimation results of the two models by Lanne and Saikkonen (2007) can be found in their paper, a two factor model under normality (Model 1) and a one factor mixture-normal model (Model 2). The estimation results of a one factor mixture-normal model (Model 3) and a two factor mixture-normal model (Model 4) are given in the Appendix B (Tables 2-3). These new models were build in order to find out whether the previous models could be improved upon. Two factors were also used in these models 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.

The quantile residuals are computed as proposed in equations (4) and (5). We have chosen $(i_1, i_2, i_3, i_4) = (1, 2, 3, 4)$ in equation (4), therefore, our observed multivariate quantile residuals are

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

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 mixture of normal distributions the marginal and conditional distributions belong to the same family of distributions. Thus, the residuals for each observation could be solved iteratively by solving the parameters of one marginal and one conditional distribution at a time. See Appendix B for more details on this.

Table 1 below gives the test statistics 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 $BIC = k \cdot \log(T - n) - 2 \cdot l_T$, 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 $n$ is the number of needed initial values. Tests based on an estimated covariance matrix $\hat{H}_T$ are only given because, according to earlier experience, they are the more reliable versions of the tests. As can be seen from Table 1, tests based on joint quantile residuals are not very critical on any of the models. The same is observed by looking at graphs based on joint quantile residuals. This is illustrated in Figure 1 in the Appendix B that gives autocovariance graphs of joint quantile residuals and squared joint quantile residuals of the Model 4. The reverse is true for tests based on multivariate quantile residuals that do not seem to accept any of the models. Model 4 is favoured by the information criteria.
Table 1: P-values of the test statistics and the values of the information criteria computed for the Models 1-4.

<table>
<thead>
<tr>
<th>Model</th>
<th>$N_3^*$</th>
<th>$A_3^*$</th>
<th>$H_3^*$</th>
<th>$N_3^*$</th>
<th>$A_3^*$</th>
<th>$H_3^*$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>0.0773</td>
<td>0.343</td>
<td>0.674</td>
<td>0.0667</td>
<td>0.0002</td>
<td>0.0025</td>
<td>2789</td>
<td>2882</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.0886</td>
<td>0.655</td>
<td>0.386</td>
<td>0.0694</td>
<td>0.0001</td>
<td>0</td>
<td>2495</td>
<td>2602</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.0982</td>
<td>0.644</td>
<td>0.494</td>
<td>0.0336</td>
<td>0.0008</td>
<td>0</td>
<td>2483</td>
<td>2599</td>
</tr>
<tr>
<td>Model 4</td>
<td>0.0264</td>
<td>0.587</td>
<td>0.638</td>
<td>0.0500</td>
<td>0</td>
<td>0.0001</td>
<td>2328</td>
<td>2453</td>
</tr>
</tbody>
</table>

The tests are computed using the estimated matrix $\hat{H}_T$, thus, denoted with *. Tests with denotation $J$ are computed using the joint quantile residuals. P-value 0 means a value < 0.00005.

In all of the four models the third multivariate quantile residual $r_{3t,\theta_T}$ is negatively auto-correlated at lag one. The absolute value of the autocorrelation is around 0.20. The Figure 1 depicts this for the Model 4 along with the 99% confidence bounds. This explains why the autocorrelation tests $A_3^*$ reject all the models. This indicates that the mean might not be constant. 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 seems to lessen when a mixture distribution is used. But even for Model 4 there is some autocorrelation left in the series of $r_{4t,\theta_T}^2$, therefore, $H_3^*$ rejects. See Figure 1 for the autocovariance functions of $r_{4t,\theta_T}$ and $r_{4t,\theta_T}^2$.

The multivariate quantile residual series (Figures 2 and 3 in the Appendix B) show that Model 4 has been able to capture the outliers quite well. The normality tests used here are not very sensitive towards that kind of misspecification, thus, they do not notice that in Model 1. The same holds for Models 2 and 3, but the graphs are not given here. An inspection of the distributional fit by other methods like histograms and normal probability plots based on the multivariate quantile residuals (not reported) are in favour of the Models 3 and 4.

The time series of $b'_t y_t$ in Model 4 are depicted in Figure 4 in the Appendix B. Comparison of the actual rate series given in Lanne and Saikkonen (2007) and Figure 4 shows that the first factor series $b'_1 y_t$ is the scaled NLG/DEM rate and the second factor series $b'_2 y_t$ is depicting the common structure in FRF/DEM and FRF/NLG rates. This can also be seen in the first and second columns of the matrix B.

---

Footnote: These 99% confidence bounds are derived in the way explained in the Section 3. Since we actually test several tests at the same time, the Bonferroni correction should be made. If we use 99% confidence bounds for 5 tests at the same time, we are, according to the correction, actually basing our inference on 95% confidence bounds.
5 Conclusion

Residual diagnostics are very useful in model evaluation in general. Excellent graphical tools are available as soon as appropriate residuals can be obtained, that is, as soon as the residuals reflect the theoretical properties of the assumed model. As pointed out in the paper, this is not the case if Pearson residuals are used with models based on mixture distributions. Since mixture models are already being used in practice, there is a need for residuals that can be used with them.

In this paper, we study multivariate quantile residuals and joint quantile residuals that can be seen as generalizations of traditional residuals. Under regularity conditions, their theoretical properties are stated, and a general framework is developed that can be used to obtain tests based on them. The framework derived takes estimation uncertainty into account. This was implemented via a standard Taylor expansion of the likelihood function and a continuously differentiable function of quantile residuals.

To illustrate how our framework can be used to obtain misspecification tests, we derived tests for non-normality, autocorrelation, cross correlation, and conditional heteroscedasticity in quantile residuals. The test statistics are simple to compute once the parameters of the model are estimated, and their application only requires the use of a conventional $\chi^2$ criterion. These tests are applicable for all models for which quantile residuals are suited. This includes models for which also traditional residuals work. In the paper, we focused on mixtures of Multivariate Generalized Orthogonal Factor GARCH models that are examples of models for which traditional residuals are not well suited.

A useful aspect of the theory provided in the paper is that it enables the use of traditional graphical diagnostics. Normal probability plots and $\chi^2$—goodness–of–fit tests are not theoretically studied in the paper, and form a topic for future research.
A Appendix: Proofs

For the sake of completeness, the usual framework of a parametric model is stated and assumed hereafter. Let \((\Omega, \mathfrak{A}, \mathbb{P})\) be a fixed probability space with a complete measure \(\mathbb{P}\) and \(Y_\theta : \Omega \to \mathbb{R}^{nT}\) a family of random variables indexed by the parameter \(\theta\) belonging to the set \(\Theta \subset \mathbb{R}^k\). Let \((\mathbb{R}^{nT}, \mathfrak{B}^{nT}, \mathbb{P}_\theta)\) be the probability space induced by \(Y_\theta\). Then \(\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}\) 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 \(f(\theta, y), \mathcal{P} = \{f(\theta, y) : \theta \in \Theta, y \in \mathbb{R}^{nT}\}\), which is the definition in the main text.

Proof of Lemma 2. Following the proof of Rosenblatt (1952) and the notation in the main text, we write \(Z_{jt} = F_{i,j-1,t-1}(\theta_0, Y_{i,t})\) for each \(j = 1, ..., n\) and \(t = 1, ..., T\). We fix the point \((z_1, ..., z_T) \in (0, 1)^{nT}\), with \(z_t = (z_{1t}, ..., z_{nt})\). Then for each \(z_{jt}\) there exists unique \(y_{i,j,t}\) such that \(z_{jt} = F_{i,j-1,t-1}(\theta_0, y_{i,j,t})\) for all \(j\) and \(t\). This follows from the fact that the distributions \(F_{i,j-1,t-1}\) are absolutely continuous w.r.t. Lebesgue measure. We denote

\[
\Theta = \left\{ \mathbb{P}(z_{1j} \leq z_1, ..., z_{Tj} \leq z_T) \mid j = 1, ..., n\right\}
\]

and

\[
\mathfrak{A} = \left\{ \prod_{t=1}^{T} \prod_{j=1}^{n} (0, z_{jt}) \mid j = 1, ..., n\right\} \subset (0, 1)^{nT}.
\]

Now,

\[
F_{(z_1, ..., z_T)|Y_0}(z_1, ..., z_T|G_0) = \mathbb{P}(Z_1 \leq z_1, ..., Z_T \leq z_T|G_0) = \mathbb{P}(Y_{i,j,t} \leq F_{i,j-1,t-1}^{-1}(\theta_0, z_{jt}) \text{ for all } j \text{ and } t|G_0)
\]

\[
= \int_{A} \prod_{t=1}^{T} \prod_{j=1}^{n} f_{i,j-1,t-1}(\theta_0, u_{i,j,t}) du_{i,j,t}
\]

\[
= \int_{B} \prod_{t=1}^{T} \prod_{j=1}^{n} dv_{jt} = \prod_{t=1}^{T} \prod_{j=1}^{n} z_{jt}.
\]

The second equality follows from absolute continuity of \(F_{i,t-1}\). The third equality uses equations (1) and (3) to rewrite the joint density. The fourth equality follows by change of variable \(v_{jt} = F_{i,j-1,t-1}(\theta_0, u_{i,j,t})\), and the fifth by integration. This proves that \(Z_{11}, ..., Z_{nT}\) are independent (conditional on \(Y_0\))^3 and each \(Z_{jt} \sim \text{Uniform}(0, 1)\). Since \(\Phi^{-1}\) is continuous, it is measurable. Then \(R_{1,\theta_0}, ..., R_{T,\theta_0}\), where \(R_{t,\theta_0} = [\Phi^{-1}(Z_{1t}) \cdots \Phi^{-1}(Z_{nt})]'\), are independent as measurable mappings of independent random variables. Clearly, \(R_{jt,\theta_0} \sim N(0, 1)\) for each \(j\) and \(t\), and therefore,

\[
[R'_{1,\theta_0} \cdots R'_{T,\theta_0}]' = [R_{11,\theta_0} \cdots R_{nT,\theta_0}]' \sim N(0, I_{nT}).
\]

^3This remark holds for every independence proven in this paper and is hereafter omitted.
Since the mapping \( F_{ij,j-1,t-1} : \Theta \times \mathbb{R} \to (0, 1) \) is continuous with respect to \( \theta \), the Continuous Mapping Theorem (see for example van der Vaart (1998), page 7) and Condition 1(3) together imply that \( F_{ij,j-1,t-1}(\theta_T, y_{ij,t}) \xrightarrow{P} F_{ij,j-1,t-1}(\theta_0, y_{ij,t}) \) whereas the continuity of \( \Phi^{-1} : (0, 1) \to \mathbb{R} \) yields

\[
R_{jt, \theta} = \Phi^{-1} \left( F_{ij,j-1,t-1}(\theta_T, y_{ij,t}) \right) \xrightarrow{P} \Phi^{-1} \left( F_{ij,j-1,t-1}(\theta_0, y_{ij,t}) \right) = R_{jt, \theta_0}
\]

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

The independence of \( R_{t+s, \theta_0} \) and \( \{Y_1, ..., Y_t\} \) (again conditional on \( Y_0 \)) for \( s \ge 1 \) follows easily using the results above: \( R_{t+s, \theta_0} \) is independent of \( \{R_{1, \theta_0}, ..., R_{t, \theta_0}\} \), and \( \{Y_1, ..., Y_t\} \) is a measurable mapping of \( \{R_{1, \theta_0}, ..., R_{t, \theta_0}\} \), since \( Y_{ij,t} = F_{ij,j-1,t+s-1}^{-1}(\theta_0, \Phi(R_{jt, \theta_0})) \).

**A.1 Proof of Lemma 3**

The proof of Lemma 3 is given as follows. First Lemma 10 is stated and proven, then it is used to prove Lemma 3.

**Lemma 10** Let \( X_1, ..., X_n \) be independently uniformly distributed random variables on \( (0, 1) \) and \( X = \prod_{i=1}^n X_i \), then \( f_n(X) \sim Uniform(0, 1) \), where \( f_n(x) = x^{\sum_{i=0}^{n-1} (-1)^i / i!} \).

**Proof.** Let \( n = 2 \), and denote \( Z_1 = X_1X_2 \) and \( Z_2 = X_2 \). Applying the determinant of the Jacobian for the inverse transformation 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 range \((z_1, 1)\) we get the density function \( f_{Z_1}(z_1) = -\ln z_1 \), and the cumulative distribution function \( F_{Z_1}(z_1) = z_1 - z_1 \ln z_1 \). Proof of Lemma 2 yields \( F_{Z_1}(Z_1) \sim Uniform(0, 1) \), as required.

We make an induction assumption that the result holds for \( n = k - 1 \), and we now show that it holds for \( n = k \).

Denote \( Z = \prod_{i=1}^{k-1} X_i \). Induction assumption gives \( F_Z(z) = z \sum_{i=0}^{k-2} (-1)^i / i! \ln z \). Therefore, derivation with respect to \( z \) yields the density function of the variable \( Z \)

\[
f_Z(z) = \frac{(-1)^{k-2}}{(k-2)!} (\ln z)^{k-2}.
\]

Denote \( V_1 = ZX_k \) and \( V_2 = X_k \).

Since \( Z \) and \( X_k \) are independent, the joint density function of \( V_1 \) and \( V_2 \) is obtained by applying the determinant of the Jacobian for the inverse transformation

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

22
when \(0 < v_1 < v_2 < 1\), and \(f_{V_1,V_2}(v_1,v_2) = 0\) otherwise. Since
\[
\frac{d}{dv_2} \left( \frac{\ln v_1}{v_2} \right)^{k-1} = (-1)(k-1) \left( \frac{\ln v_1}{v_2} \right)^{k-2} \frac{1}{v_2}
\]
and \(\ln 1 = 0\), the density function of \(V_1\) is
\[
f_{V_1}(v_1) = \frac{(-1)^{k-2}}{(k-2)!} \int_{v_1}^{1} \left( \frac{\ln v_1}{v_2} \right)^{k-2} \frac{1}{v_2} \, dv_2 = \frac{(-1)^{k-1}}{(k-1)!} (\ln v_1)^{k-1}.
\]

Integrating by parts we get the distribution function of \(V_1\)
\[
F_{V_1}(v_1) = \int_{0}^{v_1} \frac{(-1)^{k-1}}{(k-1)!} (\ln x)^{k-1} \, dx = \frac{v_1(-1)^{k-1}}{(k-1)!} (\ln v_1)^{k-1} - \lim_{x \to 0} \frac{(-1)^{k-2}}{(k-1)!} (\ln x)^{k-1} + \int_{0}^{v_1} \frac{(-1)^{k-2}}{(k-2)!} (\ln x)^{k-2} \, dx
\]
for \(0 < v_1 < 1\).

We see using (14), that \(\int_{0}^{v_1} \frac{(-1)^{k-2}}{(k-2)!} (\ln x)^{k-2} \, dx = F_Z(v_1) = v_1 \sum_{i=0}^{k-2} (-1)^i \frac{1}{i!} (\ln v_1)^i\). Application of L’Hospital’s Rule \((k-1)\) times yields \(\lim_{x \to 0} \frac{(-1)^{k-1}}{(k-1)!} (\ln x)^{k-1} = \lim_{x \to 0} x = 0\). Therefore,
\[
F_{V_1}(v_1) = v_1 \sum_{i=0}^{k-2} \frac{(-1)^i}{i!} (\ln v_1)^i,
\]
and proof of Lemma 2 yields \(F_{V_1}(v_1) \sim Uniform(0,1)\). Since \(V_1 = \prod_{i=1}^{k} X_i\), the induction principle completes the proof. ■

**Proof of Lemma 3.** Write \(X_{t,\theta} = \prod_{j=1}^{n} F_{i,j-j-1,t-1}(\theta, y_{i,j,t})\) using (3). Lemma 2 gives that \(F_{i,j-j-1,t-1}(\theta_0, Y_{i,j,t})\) are independently uniformly distributed, then Lemma 10 implies that
\[
Z_{t,\theta_0} \sim Uniform(0,1),
\]
where \(Z_{t,\theta_0} = X_{t,\theta} \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} (\ln X_{t,\theta})^i\). Clearly, \(Q_{t,\theta_0} = \Phi^{-1}(Z_{t,\theta_0}) \sim N(0,1)\).

Since \(\{X_{1,\theta_0}, ..., X_{T,\theta_0}\}\) are independent by the proof of Lemma 2, then \(Z_{t,\theta_0}\)’s as well as \(Q_{t,\theta_0}\)’s are independent as measurable transformations of independent variables. Therefore, part a) of the Lemma follows.

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

Lemma 2 c) yields: For \(s \geq 1\) \(X_{t+s,\theta_0} = \prod_{j=1}^{n} F_{i,j,j-1,t+s-1}(\theta_0, Y_{i,j,t})\) and \(\{Y_1, ..., Y_t\}\) are independent. Thus, \(Q_{t,\theta_0}\) and \(\{Y_1, ..., Y_t\}\) (again conditional on \(Y_0\)) are independent for
s \geq 1. \quad \blacksquare

**Proof of Theorem 8.** Since by Theorem 5 \( \lim_{T \to \infty} \mathbb{P}(\hat{\theta}_T \neq \infty) = 1 \), it is assumed that \( \hat{\theta}_T \neq \infty \), where \( \mathbb{P} = \mathbb{P}_{\theta_0} \) is the probability measure induced by the true parameter value \( \theta_0 \).

Again by Theorem 5 for every \( \varepsilon > 0 \) there exists \( c_0 \) and \( T_0 \) such that \( \mathbb{P}(\hat{\theta}_T \in M_{T,c_0}) > 1 - \varepsilon \) for all \( T > T_0 \). Since \( \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}_T}) \xrightarrow{P} G \) by Condition 7(1) for all \( \hat{\theta}_T \in M_{T,c} \) and \( c > 0 \), then especially \( \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}_T}) \xrightarrow{P} G \).

The Mean-Value Theorem and Conditions 6 and 7(3) together imply that

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t,\hat{\theta}_T}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}})(\hat{\theta}_T - \theta_0) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t,\theta_0}), \quad (15)
\]

where

\[
\frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}}) = \begin{bmatrix} \frac{\partial}{\partial \theta^T} g_1(U_{t,\hat{\theta}^{(1)}}) & \cdots & \frac{\partial}{\partial \theta^T} g_n(U_{t,\hat{\theta}^{(n)}}) \end{bmatrix}^T
\]

is a \((n \times k)\) Jacobian-matrix with \( U_{t,\hat{\theta}^{(j)}} = \begin{bmatrix} R'_{t,\hat{\theta}^{(j)}} & \cdots & R'_{t-m+1,\hat{\theta}^{(j)}} \end{bmatrix}^T \) (or \( U_{t,\hat{\theta}^{(j)}} = \begin{bmatrix} Q_{t,\hat{\theta}^{(j)}} & \cdots & Q_{t-m+1,\hat{\theta}^{(j)}} \end{bmatrix}^T \) depending on the choice in Condition 6), \( \hat{\theta} = (\hat{\theta}^{(1)}, \ldots, \hat{\theta}^{(n)}) \), and \( \|\hat{\theta}^{(j)} - \theta_0\| < \|\hat{\theta}_T - \theta_0\| \) for each \( j = 1, \ldots, n \).

Theorem 5 gives

\[
\sqrt{T}(\hat{\theta}_T - \theta_0) = \sqrt{T} \left[ \{A_T(\theta_0)\}^{-1} W_T(\theta_0)^{-1} \{A_T(\theta_0)\}^{-1} S_T(\theta_0) + o_P(1) \right], \quad (16)
\]

because we have by Condition 7(2)

\[
\sqrt{T} \left[ \{A_T(\theta_0)\}^{-1} W_T(\theta_0)^{-1} \cdot o_P(1) = o_P(1). \right.
\]

Since (Condition 7(1))

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}}) \cdot o_P(1) = o_P(1)
\]

equation (15) can be written using (16)

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t,\hat{\theta}_T}) = \left[ \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta^T} g(U_{t,\hat{\theta}}) \cdot \sqrt{T} \left[ \{A_T(\theta_0)\}^{-1} W_T(\theta_0)^{-1} \right. \right] + o_P(1).
\]

24
Conditions 7(1), 7(2), and Slutsky’s Lemma ensure that

\[
\left[ \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} g(U_{t, \hat{\theta}}) \cdot \sqrt{T} \left[ \{ A_T(\theta_0) \} \right]^{-1} W_T(\theta_0)^{-1} : I \right] \xrightarrow{\text{W}} \left[ GJI(\theta_0)^{-1} : I \right].
\]

(17)

Finally, using (17), Condition 7(2), and the Continuous Mapping Theorem

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} g(U_{t, \hat{\theta}_T}) \xrightarrow{\text{W}} \left[ GJI(\theta_0)^{-1} : I \right] \cdot \Sigma^{1/2} Z,
\]

where \( Z \sim N_{k+1}(0, I_{k+1}) \) and

\[
\Omega = \left[ GJI(\theta_0)^{-1} : I \right] \Sigma \left[ I(\theta_0)^{-1} J' G' \right] \left[ I_{k+1} \right] = GJI(\theta_0)^{-1} J' G' + \Psi I(\theta_0)^{-1} J' G' + GJI(\theta_0)^{-1} \Psi + H.
\]

Then we can write

\[
\left[ GJI(\theta_0)^{-1} : I \right] \Sigma^{1/2} Z = \Omega^{1/2} U,
\]

where \( U \sim N_k(0, I_k) \).

Independence of \( U \) and \( \Omega \) follows from that of \( Z \) and \( \left[ GJI(\theta_0)^{-1} : I \right] \Sigma^{1/2} \). ■

**Proof of Lemma 9.** We can write

\[
\left\| W_T(\hat{\theta}_T) - W_T(\theta_0) \right\| \leq \left\| C_T \ : I_k \right\| \left\| \begin{bmatrix} D_T & 0 \\ 0 & D_T \end{bmatrix} \right\| \left\| C_T \ : I_k \right\|,
\]

where \( C_T = \{ A_T(\hat{\theta}_T) \}^{-1} \{ A_T(\theta_0) \} \) and \( D_T = \{ A_T(\theta_0) \}^{-1} \left[ B_T(\hat{\theta}_T) - B_T(\theta_0) \right] \left[ \{ A_T(\theta_0) \}^{-1} \right]' \).

Theorem 5 and Condition 4(4) imply that \( C_T \xrightarrow{P} I_k \), \( D_T \xrightarrow{P} 0 \), and moreover

\[
\left\| \begin{bmatrix} D_T & 0 \\ 0 & D_T \end{bmatrix} \right\| = \sqrt{2} \left\| D_T \right\| \xrightarrow{P} 0.
\]

Therefore, \( \left\| W_T(\hat{\theta}_T) - W_T(\theta_0) \right\| \xrightarrow{P} 0 \). Since \( W_T(\theta_0) \xrightarrow{W} I(\theta_0) \) by Condition 4(5), then \( W_T(\hat{\theta}_T) \xrightarrow{W} I(\theta_0) \). The Continuous Mapping Theorem yields \( W_T(\hat{\theta}_T)^{-1} \xrightarrow{W} I(\theta_0)^{-1} \). Finally, the lemma follows from an application of the Continuous Mapping Theorem and Slutsky’s Lemma. ■

**Derivatives**

**Lemma 11**

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

25
and
\[ \frac{\partial}{\partial \theta} Q_{t, \theta} = \left[ \phi(Q_{t, \theta}) \right]^{-1} \frac{\partial}{\partial \theta} Z_{t, \theta}, \]

where
\[ \frac{\partial}{\partial \theta} Z_{t, \theta} = \frac{(-1)^{n-1}}{(n-1)!} (\log X_{t, \theta})^{n-1} \cdot \frac{\partial}{\partial \theta} X_{t, \theta} \]

and
\[ \frac{\partial}{\partial \theta} X_{t, \theta} = \frac{\partial}{\partial \theta} \left( \prod_{j=1}^{n} F_{i, j-1, t-1}(\theta, Y_{i, j, t}) \right). \]

Here \( \phi \) is the density of the standard normal distribution.

**Proof.** Let \( r_{jt, \theta} = \Phi^{-1}(F_{i, j-1, t-1}(\theta, y_{i, j, t})) \). The fact that \( \phi(x) > 0 \) for all \( x \in \mathbb{R} \) ensures that \( \frac{d}{dy} \Phi^{-1}(y) = \frac{1}{\phi(y)} = \frac{1}{\phi(x)} \), where \( x = \Phi^{-1}(y) \), exists for each \( y \in (0, 1) \). This and Condition 7(3) give
\[
\frac{\partial}{\partial \theta} r_{jt, \theta} = \frac{\partial}{\partial \theta} \Phi^{-1}(F_{i, j-1, t-1}(\theta, y_{i, j, t})) = \left[ (\Phi^{-1})' (F_{i, j-1, t-1}(\theta, y_{i, j, t})) \right] \frac{\partial}{\partial \theta} (F_{i, j-1, t-1}(\theta, y_{i, j, t}))
= \Phi' \left[ \Phi^{-1}(F_{i, j-1, t-1}(\theta, y_{i, j, t})) \right]^{-1} \frac{\partial}{\partial \theta} (F_{i, j-1, t-1}(\theta, y_{i, j, t}))
= \left[ \phi(r_{jt, \theta}) \right]^{-1} \cdot \frac{\partial}{\partial \theta} (F_{i, j-1, t-1}(\theta, y_{i, j, t}))
\]
for all \( s = 1, \ldots, k \). Since \( \frac{\partial}{\partial \theta} r_{jt, \theta} \) is continuous, \( \frac{\partial}{\partial \theta} R_{jt, \theta} \) is a well defined random variable.

Similarly, the above and Condition 7(3) imply \( \frac{\partial}{\partial \theta} \varphi_{t, \theta} = \left[ \phi(Q_{t, \theta}) \right]^{-1} \frac{\partial}{\partial \theta} z_{t, \theta} \). Since
\[
\frac{d}{dx} f_n(x) = \frac{d}{dx} \left( x \sum_{i=0}^{n-1} \frac{(-1)^i}{i!} (\log x)^i \right) = \frac{(-1)^{n-1}}{(n-1)!} (\log x)^{n-1},
\]
then
\[
\frac{\partial}{\partial \theta} z_{t, \theta} = \frac{(-1)^{n-1}}{(n-1)!} (\log x_{t, \theta})^{n-1} \cdot \frac{\partial}{\partial \theta} x_{t, \theta},
\]
and
\[
\frac{\partial}{\partial \theta} x_{t, \theta} = \frac{\partial}{\partial \theta} \left( \prod_{j=1}^{n} F_{i, j-1, t-1}(\theta, y_{i, j, t}) \right)
\]
Because \( \frac{\partial}{\partial \theta} \varphi_{t, \theta} \) is continuous \( \frac{\partial}{\partial \theta} Q_{t, \theta} \) is a well defined random variable. \( \blacksquare \)

Multinormality test with \( U_{t, \theta} = R_{t, \theta} \) has
\[
G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t, \theta})) = \left[ \mathbb{E}(\frac{\partial}{\partial \theta} g_1(R_{1t, \theta})) \ldots \mathbb{E}(\frac{\partial}{\partial \theta} g_n(R_{nt, \theta})) \right]^t,
\]

26
where
\[ \mathbb{E}\left( \frac{\partial}{\partial \theta} g_j(R_{jt,\theta_0}) \right) = \mathbb{E}\left[ \frac{\partial}{\partial \theta} R_{jt,\theta_0} + 2R_{jt,\theta_0} \frac{\partial}{\partial \theta} R_{jt,\theta_0} + 3 \left( R_{jt,\theta_0}^2 - 1 \right) \frac{\partial}{\partial \theta} R_{jt,\theta_0} + 4 \left( R_{jt,\theta_0}^3 - 3R_{jt,\theta_0} \right) \frac{\partial}{\partial \theta} R_{jt,\theta_0} \right] \]
and \( \frac{\partial}{\partial \theta} R_{jt,\theta_0} \) is given in Lemma 11.

If we have \( U_{t,\theta} = Q_{t,\theta} \), then the normality test has
\[
G = \mathbb{E}(\frac{\partial}{\partial \theta} g(Q_{t,\theta_0})) = \mathbb{E}\left[ \frac{\partial}{\partial \theta} Q_{t,\theta_0} + 2Q_{t,\theta_0} \frac{\partial}{\partial \theta} Q_{t,\theta_0} + 3 \left( Q_{t,\theta_0}^2 - 1 \right) \frac{\partial}{\partial \theta} Q_{t,\theta_0} + 4 \left( Q_{t,\theta_0}^3 - 3Q_{t,\theta_0} \right) \frac{\partial}{\partial \theta} Q_{t,\theta_0} \right]
\]
with \( \frac{\partial}{\partial \theta} Q_{t,\theta} \) given in Lemma 11.

**Remark 12** The random variables

1. \( \frac{\partial}{\partial \theta} R_{i,t-s,\theta_0} \) and \( R_{jt,\theta_0} \) are independent for all \( i, j \in \{1, ..., n\} \) and \( s \geq 1 \), and
2. \( \frac{\partial}{\partial \theta} Q_{t-s,\theta_0} \) and \( Q_{t,\theta_0} \) are independent for all \( s \geq 1 \).

**Proof.** According to Lemma 11
\[
\frac{\partial}{\partial \theta} R_{i,t-s,\theta_0} = \left[ \phi(R_{i,t-s,\theta_0}) \right]^{-1} \frac{\partial}{\partial \theta}(F_{m_i,t-s-i}(\theta_0, Y_{m_i,t-s}))
\]
is measurable, and especially a measurable function of random variables \( \{Y_0, Y_1, ..., Y_{t-s}\} \). Lemma 2 c) gives the independence of \( R_{jt,\theta_0} \) and \( \{Y_0, Y_1, ..., Y_{t-s}\} \) for all \( s \geq 1 \), which implies the stated result (1).

Likewise, Lemma 3 c) yields the independence of \( Q_{t,\theta_0} \) and \( \frac{\partial}{\partial \theta} Q_{t-s,\theta_0} \) for all \( s \geq 1 \). \( \blacksquare \)

Using Remark 12 it is seen that a typical row of the matrix \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t,\theta_0})) \) in the autocorrelation test with \( U_{t,\theta} = \left[ R'_{t,\theta} \cdots R'_{t-K_1,\theta} \right]' \) is
\[
\mathbb{E}\left( \frac{\partial}{\partial \theta} (R_{i,t-s,\theta_0} R_{jt,\theta_0}) \right) = \mathbb{E}(R_{i,t-s,\theta_0} \frac{\partial}{\partial \theta} R_{jt,\theta_0}) + \mathbb{E}(R_{jt,\theta_0} \frac{\partial}{\partial \theta} R_{i,t-s,\theta_0})
\]
\[
= \mathbb{E}(R_{i,t-s,\theta_0} \frac{\partial}{\partial \theta} R_{jt,\theta_0}),
\]
where \( \frac{\partial}{\partial \theta} R_{jt,\theta_0} \) is the vector of derivatives given in Lemma 11, \( i, j \in \{1, ..., n\} \), and \( s = 1, ..., K_1 \).

If we have \( U_{t,\theta} = \left[ Q_{t,\theta} \cdots Q_{t-K_1,\theta} \right]' \) in the autocorrelation test, then using Remark 12 we see that the \( s \)th row of the matrix \( G = \mathbb{E}(\frac{\partial}{\partial \theta} g(U_{t,\theta_0})) \) is
\[
\mathbb{E}\left( \frac{\partial}{\partial \theta} (Q_{t-s,\theta_0} Q_{t,\theta_0}) \right) = \mathbb{E}(Q_{t-s,\theta_0} \frac{\partial}{\partial \theta} Q_{t,\theta_0}) + \mathbb{E}(Q_{t,\theta_0} \frac{\partial}{\partial \theta} Q_{t-s,\theta_0})
\]
\[
= \mathbb{E}(Q_{t-s,\theta_0} \frac{\partial}{\partial \theta} Q_{t,\theta_0}),
\]
where \( \frac{\partial}{\partial \theta} Q_{t,\theta} \) is given in Lemma 11.
Remark 13. The random variables
(1) $R_{i,t,\theta_0}^2$ and $R_{j,t-s,\theta_0} \frac{\partial}{\partial \theta_0} R_{j,t-s,\theta_0}$ are independent for all $s \geq 1$,
and
(2) $Q_{i,t,\theta_0}^2$ and $Q_{t-s,\theta_0} \frac{\partial}{\partial \theta_0} Q_{t-s,\theta_0}$ are independent for all $s \geq 1$.

Proof. $R_{i,t,\theta_0}^2$ is a measurable function of $R_{i,t,\theta_0}$, and $\frac{\partial}{\partial \theta_0} (F_{m,j-t-s-1}(\theta_0, Y_{m,t-s}))$, $[\phi (R_{j,t-s,\theta_0})]^{-1}$ and $R_{j,t-s,\theta_0}$ are measurable functions of $\{Y_0, Y_1, ..., Y_{t-s}\}$. The independence follows using Lemma 2 c).
Similarly, Lemma 3 c) yields the independence of $Q_{i,t,\theta_0}^2$ and $Q_{t-s,\theta_0} \frac{\partial}{\partial \theta_0} Q_{t-s,\theta_0}$ for all $s \geq 1$. ■

Using Remark 13 a typical row of the matrix $G = \mathbb{E}(\frac{\partial}{\partial \theta_0} g(U_{t,\theta_0}))$ in the heteroscedasticity test based on $U_{t,\theta} = \left[ R_{i,t} \cdots R_{i-K,\theta} \right]'$ is

\[
\mathbb{E} \left[ \frac{\partial}{\partial \theta_0} (R_{i,t-s,\theta_0}^2 - 1) \left( R_{j,t,\theta_0}^2 - 1 \right) \right] = 2\mathbb{E} \left[ \left( R_{i,t-s,\theta_0}^2 - 1 \right) R_{j,t,\theta_0} \frac{\partial}{\partial \theta_0} R_{j,t,\theta_0} + \left( R_{j,t,\theta_0}^2 - 1 \right) R_{i,t-s,\theta_0} \frac{\partial}{\partial \theta_0} R_{i,t-s,\theta_0} \right]
\]

\[
= 2\mathbb{E} \left[ \left( R_{i,t-s,\theta_0}^2 - 1 \right) R_{j,t,\theta_0} \frac{\partial}{\partial \theta_0} R_{j,t,\theta_0} \right],
\]

where $\frac{\partial}{\partial \theta_0} R_{j,t,\theta_0}$ is the vector of derivatives given in Lemma 11, $i, j \in \{1, \ldots, n\}$, and $s = 1, \ldots, K_2$.

If we have $U_{t,\theta} = \left[ Q_{i,t} \cdots Q_{i-K,\theta} \right]'$ in the heteroscedasticity test, then Remark 13 yields

\[
\mathbb{E}(\frac{\partial}{\partial \theta_0} (Q_{i,t-s,\theta_0}^2 - 1) \left( Q_{i,t,\theta_0}^2 - 1 \right)) = 2\mathbb{E}((Q_{i,t-s,\theta_0}^2 - 1)Q_{i,t,\theta_0} \frac{\partial}{\partial \theta_0} Q_{i,t,\theta_0}) + \mathbb{E}((Q_{i,t,\theta_0}^2 - 1)Q_{t-s,\theta_0} \frac{\partial}{\partial \theta_0} Q_{t-s,\theta_0})
\]

\[
= 2\mathbb{E}((Q_{i,t-s,\theta_0}^2 - 1)Q_{i,t,\theta_0} \frac{\partial}{\partial \theta_0} Q_{i,t,\theta_0}),
\]

as a $s$th row of matrix $G = \mathbb{E}(\frac{\partial}{\partial \theta_0} g(U_{t,\theta_0}))$ and $\frac{\partial}{\partial \theta_0} Q_{i,t,\theta}$ is given in Lemma 11.
Appendix

B.1 Multivariate quantile residuals

It is shown below that for the family of mixture of normal distributions the marginal and conditional distributions belong to the same family of distributions.

Denote with \( X \) \((n \times 1)\) a random variable that follows a mixture of two normal distributions. The density of \( X \) is

\[
f_X(x) = p (2\pi)^{-n/2} \det(\Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\} \\
+ (1 - p) (2\pi)^{-n/2} \det(\Omega)^{-1/2} \exp \left\{ -\frac{1}{2} (x - \nu)' \Omega^{-1} (x - \nu) \right\}
\]

\[
= p \cdot MN_n(\mu, \Sigma) + (1 - p) \cdot MN_n(\nu, \Omega),
\]

where \( MN_n(\mu, \Sigma) \) and \( MN_n(\nu, \Omega) \) denote the densities of multinormal distribution with expectations \( \mu \) and \( \nu \), and covariance matrices \( \Sigma \) and \( \Omega \), respectively. Make a partition on \( X = [X^{(1)}' \ X^{(2)}']' \) and analogous partitions on expectations \( \mu = [\mu_1' \ \mu_2'] \) and \( \nu = [\nu_1' \ \nu_2'] \) and on covariance matrices

\[
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad \text{and} \quad \Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.
\]

If the dimensions of variables \( X^{(1)} \) and \( X^{(2)} \) are \( k \) and \( n - k \), respectively, then the marginal distribution of \( X^{(2)} \) is a mixture of two normal distributions with density

\[
f_{X^{(2)}}(x^{(2)}) = p (2\pi)^{-(n-k)/2} \det(\Sigma_{22})^{-1/2} \exp \left\{ -\frac{1}{2} (x^{(2)} - \mu_2)' \Sigma_{22}^{-1} (x^{(2)} - \mu_2) \right\} \\
+ (1 - p) (2\pi)^{-(n-k)/2} \det(\Omega_{22})^{-1/2} \exp \left\{ -\frac{1}{2} (x^{(2)} - \nu_2)' \Omega_{22}^{-1} (x^{(2)} - \nu_2) \right\}
\]

\[
= p \cdot MN_{n-k}(\mu_2, \Sigma_{22}) + (1 - p) \cdot MN_{n-k}(\nu_2, \Omega_{22}).
\]

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

In order to obtain the conditional distribution of \( X^{(1)} \) conditional on \( X^{(2)} = x^{(2)} \) we introduce following notation. Define \( \Sigma_{11:2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \) and \( \Omega_{11:2} \) analogously. Form the so-called Schur decomposition

\[
\begin{bmatrix} I_k & -\Sigma_{12} \Sigma_{22}^{-1} \\ 0 & I_l \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_k & 0 \\ -\Sigma_{22}^{-1} \Sigma_{21} & I_l \end{bmatrix} = \begin{bmatrix} \Sigma_{11:2} & 0 \\ 0 & \Sigma_{22} \end{bmatrix},
\]

29
which yields

\[
\Sigma^{-1} = \begin{bmatrix}
    I_k & 0 \\
    -\Sigma_{11}^{-1} \Sigma_{21} & I_l
\end{bmatrix}
\begin{bmatrix}
    \Sigma_{11}^{-1} & 0 \\
    0 & \Sigma_{22}^{-1}
\end{bmatrix}
\begin{bmatrix}
    I_k & -\Sigma_{12} \Sigma_{22}^{-1} \\
    0 & I_l
\end{bmatrix}.
\]

Then \( \det (\Sigma) = \det (\Sigma_{11}) \det (\Sigma_{22}) \). This and the notation \( x^{(1)} - \mu_1 - \Sigma_{12} \Sigma_{22}^{-1} (x^{(2)} - \mu_2) = x^{(1)} - a (x^{(2)}) \) together give

\[
(x - \mu)' \Sigma^{-1} (x - \mu) = (x^{(1)} - a (x^{(2)}))' \Sigma_{11}^{-1} (x^{(1)} - a (x^{(2)})) + (x^{(2)} - \mu_2)' \Sigma_{22}^{-1} (x^{(2)} - \mu_2).
\]

The same holds when \( \Sigma \) and \( (x^{(1)} - a (x^{(2)})) \) are replaced by \( \Omega \) and \( x^{(1)} - b (x^{(2)}) = x^{(1)} - \nu_1 - \Omega_{12} \Omega_{22}^{-1} (x^{(2)} - \nu_2) \). The joint density function of \( X \) can therefore be written

\[
f_X (x) = p \cdot MN_k (x^{(1)} - a (x^{(2)}), \Sigma_{11}) \cdot MN_n-k (\mu_2, \Sigma_{22}) + (1 - p) \cdot MN_k (x^{(1)} - b (x^{(2)}), \Omega_{11}) \cdot MN_n-k (\nu_2, \Omega_{22}).
\]

The conditional distribution of \( X^{(1)} \) on condition \( X^{(2)} = x^{(2)} \) is

\[
f_{X^{(1)}|X^{(2)}} (x^{(1)}|X^{(2)} = x^{(2)}) = f_X (x) / f_X (x^{(2)}) =
\]

\[
p (x^{(2)}) \cdot MN_k (x^{(1)} - a (x^{(2)}), \Sigma_{11}) + (1 - p) (x^{(2)}) \cdot MN_k (x^{(1)} - b (x^{(2)}), \Omega_{11}),
\]

where

\[
p (x^{(2)}) = p \cdot MN_n-k (\mu_2, \Sigma_{22}) / (p \cdot MN_n-k (\mu_2, \Sigma_{22}) + (1 - p) \cdot MN_n-k (\nu_2, \Omega_{22})
\]

is a function of \( x^{(2)} \) and the parameters \( p, \mu_2, \nu_2, \Sigma_{22}, \text{and } \Omega_{22} \).

Thus, the residuals for each observation can be solved 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 \( a (x^{(2)}) \) and \( b (x^{(2)}) \), covariance matrices \( \Sigma_{11} \) and \( \Omega_{11} \) and the mixing proportion \( p (x^{(2)}) \) that form the set of parameters for the new conditional distribution. At the same time one marginal distribution is solved, which then is integrated in order to solve a desired component of the multivariate quantile residual vector at a fixed time point. This procedure can be used for the models in our empirical example whatever order of conditioning is chosen for the multivariate quantile residuals.

In general multivariate quantile residuals can always be computed with numerical integration. This task becomes very burdensome along with the growth of the dimension of the data. Therefore, any theory that yields analytical results on the solution of the marginal and conditional distributions is very useful. Similar results, than presented here, can be obtained
within the families of elliptical and spherical distributions. See Fang, Kotz, and Ng (1990) and references therein for the general results on these families.

B.2 Estimated models

Table 2: One-factor mixture-normal model (Model 3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\Psi$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0108</td>
<td>0.3111</td>
<td>0.9864</td>
<td>0.6159</td>
<td>0.2660</td>
<td>0.1295</td>
</tr>
<tr>
<td></td>
<td>(0.0074)</td>
<td>(0.0828)</td>
<td>(0.0066)</td>
<td>(0.0817)</td>
<td>(0.0589)</td>
<td>(0.0207)</td>
</tr>
</tbody>
</table>

$B$ $\begin{pmatrix} 0.0249 & 3.2706 & 0.4356 & -0.0898 \\ (0.2053) & (0.2095) & (0.3533) & (0.2856) \end{pmatrix}$

$\begin{pmatrix} -9.3219 & -1.3286 & 1.0471 & -1.6022 \\ (3.8093) & (0.4053) & (2.2683) & (0.6551) \end{pmatrix}$

$\begin{pmatrix} 9.2943 & -1.8667 & -0.1021 & 0.4369 \\ (3.7970) & (0.4225) & (2.1488) & (0.6192) \end{pmatrix}$

$\begin{pmatrix} -0.0718 & 0.0097 & -0.7762 & 1.5093 \\ (0.1053) & (0.0555) & (0.5740) & (0.3169) \end{pmatrix}$

$\begin{pmatrix} 0.2660 & 0.0497 & 0.3067 & 0.4709 \\ (0.0589) & (0.0084) & (0.0568) & (0.1016) \end{pmatrix}$

The estimated standard errors are in the parentheses. Estimates are computed using the cross-product of the first derivatives of the log-likelihood function.
### Table 3: Two-factor mixture-normal model (Model 4)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\alpha_{11}$</th>
<th>$\alpha_{21}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\alpha_{22}$</th>
<th>$\beta_{22}$</th>
<th>$B$</th>
<th>$\Psi$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0202</td>
<td>0.2726</td>
<td>0.9789</td>
<td>0.6441</td>
<td>0.1684</td>
<td>0.8239</td>
<td>0.1901</td>
<td>0.1106</td>
<td>0.0588</td>
</tr>
<tr>
<td></td>
<td>(0.0121)</td>
<td>(0.0608)</td>
<td>(0.0136)</td>
<td>(0.0546)</td>
<td>(0.0216)</td>
<td>(0.0221)</td>
<td>(0.1226)</td>
<td>(0.0371)</td>
<td>(0.0118)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1684</td>
<td>0.8239</td>
<td></td>
<td></td>
<td>-8.9138</td>
<td>0.1906</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0216)</td>
<td>(0.0221)</td>
<td>(0.3075)</td>
<td>(0.0366)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.1854)</td>
<td>(0.0366)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.2512)</td>
<td>(0.0473)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-8.9138</td>
<td>(0.0371)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.2523)</td>
<td>(0.0473)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.8504)</td>
<td>(0.0473)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.8000)</td>
<td>(0.0949)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-8.9138</td>
<td>(2.9147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.523)</td>
<td>(2.9147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.8504)</td>
<td>(2.9147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.8000)</td>
<td>(2.9147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0924</td>
<td>(0.0612)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0216)</td>
<td>(0.0612)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.3728)</td>
<td>(0.0612)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.2204)</td>
<td>(0.0612)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.0924</td>
<td>(0.0346)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.0372)</td>
<td>(0.0346)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.2204)</td>
<td>(0.0346)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.7480</td>
<td>(2.8563)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2.925)</td>
<td>(2.8563)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.6920)</td>
<td>(2.8563)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.7183)</td>
<td>(2.8563)</td>
<td></td>
</tr>
<tr>
<td>$\Psi$</td>
<td>0.1106</td>
<td>0.1390</td>
<td>0.2005</td>
<td>0.3879</td>
<td></td>
<td></td>
<td>0.0588</td>
<td>0.0588</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0371)</td>
<td>(0.0366)</td>
<td>(0.0473)</td>
<td>(0.0949)</td>
<td></td>
<td></td>
<td>(0.0118)</td>
<td>(0.0118)</td>
<td></td>
</tr>
</tbody>
</table>

The estimated standard errors are in the parentheses. Estimates are computed using the cross-product of the first derivatives of the log-likelihood function.
Figure 1: Autocovariance functions of joint or multivariate quantile residuals and squared joint or multivariate quantile residuals of the Model 4 divided by their approximate standard errors. The standard errors are obtained from the estimated covariance matrix $T^{-1}\hat{\Omega}_T$ with estimated $\hat{H}_T$ as described in Section 3. 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)
Figure 4: Time series of $b_i y_t$, $i = 1, 2, 3, 4$ according to the two factor mixture-normal model, i.e., Model 4.
References


BOWMAN, K. O., AND L. R. SHENTON (1975): “Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and $b_2$,” Biometrika, 62(2), 243–250.


