

**A note on nonparametric testing  
for Gaussian innovations  
in AR-ARCH models**

N. NEUMEYER, L. SELK

**Preprint–No. 2012-02      April 2012**

FACHBEREICH MATHEMATIK  
SCHWERPUNKT MATHEMATISCHE STATISTIK  
UND STOCHASTISCHE PROZESSE

# A note on nonparametric testing for Gaussian innovations in AR-ARCH models

Natalie Neumeyer and Leonie Selk\*

University of Hamburg, Department of Mathematics

Bundesstrasse 55, 20146 Hamburg, Germany

April 18, 2012

## Abstract

In this paper we consider autoregressive models with conditional autoregressive variance, including the case of homoscedastic AR-models and the case of ARCH models. Our aim is to test the hypothesis of normality for the innovations in a completely nonparametric way, i. e. without imposing parametric assumptions on the conditional mean and volatility functions. To this end the Cramér-von Mises test based on the empirical distribution function of nonparametrically estimated residuals is shown to be asymptotically distribution-free. We demonstrate its good performance for finite sample sizes in a simulation study.

AMS 2010 Classification: Primary 62M10, Secondary 62G10

Keywords and Phrases: autoregression, conditional heteroscedasticity, empirical distribution function, kernel estimation, nonparametric CHARN model, time series

---

\*corresponding author, e-mail: leonie.selk@math.uni-hamburg.de

# 1 Introduction

Nonlinear AR-ARCH models, i. e. models with an autoregressive conditional mean function and an autoregressive conditional variance function which are both not assumed linear, have become increasingly popular. They are also called CHARN (conditional heteroscedastic autoregressive nonlinear) models. In this paper we assume an AR-ARCH model of order one. Our aim is to test for Gaussian distribution of the innovations, which constitutes a typical assumption in the modelling of time series data. Under the normality assumption, asymptotic results often simplify. For instance, then no fourth moments appear in the asymptotic variance matrix of the empirical autocovariances of linear processes, see e. g. Brockwell & Davis (2006), Proposition 7.3.3. Further, the lack-of-fit test for ARCH models by Horváth, Kokozka & Teyssière (2004) strongly depends on the assumption of Gaussian innovations. Araveporn (2011) uses the assumption of Gaussian innovations in order to estimate the conditional mean and volatility functions. Furthermore, estimation of conditional quantiles is of great importance in the context of financial time series. Starting from a nonparametric AR-ARCH model, however, the quantiles of the innovation distribution need to be estimated, or better be known (see Franke, Kreiß & Mammen (2009), section 4). Moreover, under Gaussianity of the innovations one can derive asymptotically distribution-free versions of other specification tests; see the discussion below.

We suggest a completely nonparametric test, which does not assume any parametric assumption on either mean or volatility function, but applies kernel estimators for those functions (see Doukhan & Ghindès (1983), Robinson (1983), Masry & Tjøstheim (1995), Härdle & Tsybakov (1997), among others, for estimation procedures in nonparametric AR-ARCH models). Relatedly, in a homoscedastic nonparametric AR model Müller, Schick & Wefelmeyer (2009), who prove an asymptotic expansion of the empirical process of estimated innovations, mention the possibility to use their result for testing goodness-of-fit of the innovation distributions. They do not present the asymptotic distribution of the test statistics, nor finite sample properties. On the other hand, goodness-of-fit tests for the innovation distribution in parametric time series are suggested by Koul & Ling (2006) in AR-ARCH models and by Klar, Lindner & Meintanis (2011) for GARCH models, for instance. Ducharme & Lafaye de Micheaux (2004) test for normality of the innovations in standard ARMA models.

Our test statistic for Gaussianity of the innovations is the Cramér-von Mises distance of a weighted empirical distribution function of estimated innovations and the standard normal distribution. Though the mean and volatility functions are not specified in any way, the test statistic is shown to be asymptotically distribution-free. The test and its asymptotic distribution are presented in section 2.1. We treat the special cases of nonparametric AR and nonparametric ARCH models in detail in sections 2.2 and 2.3. In a small simulation study we demonstrate the good performance of the test for moderate sample sizes in section 3. We further discuss briefly how the test can be generalized to AR-ARCH models of higher

order or to models with additional covariates.

As already mentioned, under Gaussianity of the innovations other testing procedures based on the residual empirical process will be asymptotically distribution-free. Then bootstrap procedures (for which asymptotic validity often is not investigated rigorously in the literature) can be avoided. As example for an asymptotically distribution-free specification test under the normality assumption we present a lack-of-fit test for standard AR(1) models in section 4. Further we reconsider the test for multiplicative structure by Dette, Pardo-Fernández & Van Keilegom (2009) under the normality assumption.

Finally, technical assumptions are listed in appendix A.

## 2 Main results

### 2.1 AR-ARCH model

Assume we have observed  $X_0, \dots, X_n$ , where  $(X_t)_{t \in \mathbb{Z}}$  is a real valued stationary  $\alpha$ -mixing stochastic process following the AR-ARCH model of order one, i. e.

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t. \quad (2.1)$$

Here the innovations  $\varepsilon_t$ ,  $t \in \mathbb{Z}$ , are assumed independent and identically distributed with unknown distribution function  $F$ . Moreover, the innovations are centered with unit variance and  $\varepsilon_t$  is independent of the past  $X_s$ ,  $s \leq t-1$ ,  $\forall t$ .

Our aim is to test the null hypothesis  $H_0$  of standard normal innovations against a general alternative  $H_1$ . To this end we define kernel estimators for the conditional mean and conditional variance function as

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right) X_i}{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right)}, \quad \hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right) (X_i - \hat{m}(x))^2}{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right)} \quad (2.2)$$

where  $K$  denotes a kernel function and  $c_n$  a sequence of positive bandwidths. Technical assumptions are listed in the appendix. Now we estimate the innovations as residuals

$$\hat{\varepsilon}_t = \frac{X_t - \hat{m}(X_{t-1})}{\hat{\sigma}(X_{t-1})}$$

and consider a weighted empirical distribution function, i. e.

$$\hat{F}_n(y) = \sum_{t=1}^n v_{n,t} I\{\hat{\varepsilon}_t \leq y\}, \quad (2.3)$$

as estimator for the innovation distribution. Here we define

$$v_{n,t} = \frac{w_n(X_{t-1})}{\sum_{s=1}^n w_n(X_{s-1})}$$

while  $w_n$  denotes some weight function fulfilling assumption (W) in the appendix. Selk & Neumeier (2012) showed (see (A.1), (A.3) and the arguments following in the proof of Th. 3.1 in that paper) that under the assumptions stated in the appendix,

$$\begin{aligned}\hat{F}_n(y) &= \sum_{t=1}^n v_{n,t} I \left\{ \varepsilon_t \leq \frac{(\hat{m} - m)(X_{t-1})}{\sigma(X_{t-1})} + y \frac{\hat{\sigma}(X_{t-1})}{\sigma(X_{t-1})} \right\} \\ &= \frac{1}{n} \sum_{t=1}^n I \{ \varepsilon_t \leq y \} + f(y) \sum_{t=1}^n v_{n,t} \left( \frac{(\hat{m} - m)(X_{t-1})}{\sigma(X_{t-1})} + y \frac{(\hat{\sigma}^2 - \sigma^2)(X_{t-1})}{2\sigma^2(X_{t-1})} \right) \\ &\quad + o_P\left(\frac{1}{\sqrt{n}}\right)\end{aligned}\quad (2.4)$$

uniformly with respect to  $y \in \mathbb{R}$ , where  $f$  denotes the innovation density. Further in the aforementioned paper it is shown that

$$\sum_{t=1}^n v_{n,t} \frac{(\hat{m} - m)(X_{t-1})}{\sigma(X_{t-1})} = \frac{1}{n} \sum_{t=1}^n \varepsilon_t + o_P\left(\frac{1}{\sqrt{n}}\right) \quad (2.5)$$

$$\sum_{t=1}^n v_{n,t} \frac{(\hat{\sigma}^2 - \sigma^2)(X_{t-1})}{2\sigma^2(X_{t-1})} = \frac{1}{2n} \sum_{t=1}^n (\varepsilon_t^2 - 1) + o_P\left(\frac{1}{\sqrt{n}}\right) \quad (2.6)$$

(see (A.5)–(A.7) in the cited paper). Thus,

$$\hat{F}_n(y) = \frac{1}{n} \sum_{t=1}^n \left( I \{ \varepsilon_t \leq y \} + f(y) \varepsilon_t + \frac{y f(y)}{2} (\varepsilon_t^2 - 1) \right) + o_P\left(\frac{1}{\sqrt{n}}\right) \quad (2.7)$$

and the stochastic process

$$\sqrt{n} \left( \hat{F}_n(y) - F(y) \right), \quad y \in \mathbb{R},$$

converges weakly to a centered Gaussian process  $(\mathbb{K}(y))_{y \in \mathbb{R}}$  with

$$\begin{aligned}\text{Cov}(\mathbb{K}(y), \mathbb{K}(z)) &= F(y \wedge z) - F(y)F(z) \\ &\quad + f(y) \left( E[\varepsilon_1 I \{ \varepsilon_1 \leq z \}] + y E[(\varepsilon_1^2 - 1) I \{ \varepsilon_1 \leq z \}] \right) \\ &\quad + f(z) \left( E[\varepsilon_1 I \{ \varepsilon_1 \leq y \}] + z E[(\varepsilon_1^2 - 1) I \{ \varepsilon_1 \leq y \}] \right) \\ &\quad + f(y)f(z) \left( 1 + (y + z) E[\varepsilon_1^3] + yz (E[\varepsilon_1^4] - 1) \right).\end{aligned}$$

Now let  $\Phi$  and  $\varphi$  denote distribution and density function of the standard normal distribution, respectively, and denote by  $(\mathbb{G}(u))_{u \in [0,1]}$  a centered Gaussian process with covariance structure

$$\text{Cov}(\mathbb{G}(u), \mathbb{G}(v)) = u \wedge v - uv - \varphi(\Phi^{-1}(u))\varphi(\Phi^{-1}(v)).$$

Then we have the following result for the Cramér-von Mises test statistic.

**Theorem 2.1** *Under model (2.1) and the assumptions stated in the appendix under the null hypothesis  $H_0$  of Gaussian innovations the test statistic*

$$T_n = n \int_{\mathbb{R}} (\hat{F}_n(y) - \Phi(y))^2 \varphi(y) dy$$

*converges in distribution to  $T = \int_0^1 \mathbb{G}^2(u) du$ .*

**Proof.** A calculation of the covariance of  $\mathbb{K}$  in the case  $F = \Phi$  gives

$$\begin{aligned} \text{Cov}(\mathbb{K}(y), \mathbb{K}(z)) &= \Phi(y \wedge z) - \Phi(y)\Phi(z) \\ &\quad - 2\varphi(y)\varphi(z) - 2yz\varphi(y)\varphi(z) + \varphi(y)\varphi(z)(1 + 2yz) \\ &= \Phi(y \wedge z) - \Phi(y)\Phi(z) - \varphi(y)\varphi(z). \end{aligned}$$

because for  $\varepsilon_1$  standard normally distributed one easily calculates  $E[\varepsilon_1 I\{\varepsilon_1 \leq y\}] = -\varphi(y)$ ,  $E[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \leq y\}] = -y\varphi(y)$  and one has  $E[\varepsilon_1^3] = 0$ ,  $E[\varepsilon_1^4] = 3$ . From the continuous mapping theorem it follows that  $T_n$  converges in distribution to

$$\int_{\mathbb{R}} \mathbb{K}(y)^2 \varphi(y) dy = \int_0^1 (\mathbb{K}(\Phi^{-1}(u)))^2 du$$

while  $\mathbb{K} \circ \Phi^{-1}$  has the same distribution as  $\mathbb{G}$ . This finishes the proof.  $\square$

**Remark 2.2** It follows from Stephens (1976) that  $\mathbb{G}$  is also the weak limit of some process  $(Y(u))_{u \in [0,1]}$  such that  $\tilde{T} = \int_0^1 Y^2(u) du$  is the limit of

$$\tilde{T}_n = n \int \left( \frac{1}{n} \sum_{j=1}^n I\{Z_j \leq \cdot\} - \Phi_{\hat{\mu}, \tau^2} \right)^2 d\Phi_{\hat{\mu}, \tau^2},$$

where  $Z_1, \dots, Z_n$  are iid with known variance  $\tau^2$  and unknown expectation  $\mu$  and where  $\Phi_{\hat{\mu}, \tau^2}$  denotes the normal distribution function with expectation  $\hat{\mu} = n^{-1} \sum_{j=1}^n Z_j$  and variance  $\tau^2$ . That the limits of  $T_n$  and  $\tilde{T}_n$  coincide in distribution might be surprising because in our AR-ARCH model the variance is unknown and has to be estimated. However, as can be seen from the proof, in the asymptotic covariance of  $\sqrt{n}(\hat{F}_n - \Phi)$  under  $H_0$  exactly those terms cancel that arise from the estimation of the variance function  $\hat{\sigma}^2$  (cf. (2.6)). Quantiles of  $\tilde{T}$  and thus critical values for  $T$  are tabled in Stephens (1976) and restated in Table 1 for convenience. We obtain an asymptotically distribution-free test by rejecting  $H_0$  for asymptotic level  $\alpha$  whenever  $T_n$  is larger than the critical value  $c_\alpha$ . Consistency can be deduced from uniform convergence of  $\hat{F}_n$  to  $F$  in probability, which follows from (2.7).  $\blacksquare$

nominal level $\alpha$	0.15	0.1	0.05	0.025	0.01
critical value $c_\alpha$	0.118	0.135	0.165	0.196	0.237

Table 1: *Asymptotic critical values for the Cramér-von Mises test for normality of the innovations in the AR-ARCH model.*

**Remark 2.3** The results for the nonparametric AR-ARCH model (2.1) can be extended to models of the form  $X_t = m(Z_t) + \sigma(Z_t)\varepsilon_t$  where  $(X_t, Z_t)$  is a stationary time series and  $Z_t$  may include covariates as well as a finite number of past values  $X_{t-1}, \dots, X_{t-p}$  while  $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0$ ,  $\text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = 1$ . Here  $\mathcal{F}_{t-1}$  denotes the sigma-field generated by  $Z_t, (X_{t-1}, Z_{t-1}), (X_{t-2}, Z_{t-2}), \dots$ . We conjecture that applying local polynomial estimators for  $m$  and  $\sigma^2$  and assuming enough smoothness of those functions the expansion (2.7) stays valid. A thorough treatment is beyond the scope of the paper. Note also that asymptotic properties of estimators for the innovation distribution in such nonparametric AR( $p$ )/regression models have not yet been treated in the literature. However, in the case of independent observations Neumeyer & Van Keilegom (2010) showed validity of an expansion like (2.7) for the empirical distribution of residuals in multiple nonparametric regression models (they obtain the same expansion as in the case of one-dimensional covariates, see Akritas & Van Keilegom (2001)). Thus we believe that Theorem 2.1 stays valid in the more general model under suitable regularity conditions and the same test for Gaussianity of the innovation distribution can be applied. ■

## 2.2 AR model

In this section we consider a nonparametric AR model of order one, i. e.

$$X_t = m(X_{t-1}) + \eta_t, \quad (2.8)$$

where the innovations  $\eta_t$ ,  $t \in \mathbb{Z}$ , are iid and centered and  $\eta_t$  is independent of the past  $X_s$ ,  $s \leq t-1$ . Thus we have model (2.1) with  $\eta_t = \sigma\varepsilon_t$  for the unknown (constant) variance  $\sigma^2 = \text{Var}(\eta_t)$ . Our aim is to test the null hypothesis of normal innovations, i. e.

$$H_0 : \exists \sigma^2 > 0 \text{ s. t. } \eta_t \sim N(0, \sigma^2).$$

The constant variance is estimated by

$$\hat{\sigma}^2 = \sum_{t=1}^n v_{n,t} (X_t - \hat{m}(X_{t-1}))^2,$$

where  $\hat{m}$  is the kernel estimator defined in (2.2). In this case we define the residuals as

$$\hat{\varepsilon}_t = \frac{X_t - \hat{m}(X_{t-1})}{\hat{\sigma}}$$

and consider  $\hat{F}_n$  as defined in (2.3) as estimator for the distribution of the standardized innovations  $\varepsilon_t$ . Let again  $F$  and  $f$  denote distribution and density function of  $\varepsilon_t$ , respectively. Then  $H_0$  is equivalent to  $F = \Phi$  and we have the following results.

**Lemma 2.4** *Under the assumptions stated in the appendix we have the expansion*

$$\hat{F}_n(y) = \frac{1}{n} \sum_{t=1}^n \left( I\{\varepsilon_t \leq y\} + f(y)\varepsilon_t + \frac{yf(y)}{2}(\varepsilon_t^2 - 1) \right) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

uniformly with respect to  $y \in \mathbb{R}$ .

**Proof.** First we consider the variance estimator, for which one obtains

$$\begin{aligned} \hat{\sigma}^2 - \sigma^2 &= \frac{1}{n} \sum_{t=1}^n v_{n,t}(\eta_t - (\hat{m} - m)(X_{t-1}))^2 - \sigma^2 \\ &= \sigma^2 \frac{1}{n} \sum_{t=1}^n v_{n,t}(\varepsilon_t^2 - 1) - 2\sigma \sum_{t=1}^n v_{n,t}\varepsilon_t(\hat{m} - m)(X_{t-1}) + \sum_{t=1}^n v_{n,t}(\hat{m} - m)^2(X_{t-1}) \\ &= \sigma^2 \sum_{t=1}^n v_{n,t}(\varepsilon_t^2 - 1) + o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

by Lemmata B.2 and B.3 in Selk & Neumeier (2012). Further

$$\frac{1}{n} \sum_{t=1}^n (w_n(X_{t-1}) - 1)(\varepsilon_t^2 - 1) = o_P\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\frac{1}{n} \sum_{t=1}^n w_n(X_{t-1}) = 1 + o_P\left(\frac{1}{\sqrt{n}}\right)$$

(see also (A.4) in Selk & Neumeier (2012)) and thus

$$\hat{\sigma}^2 - \sigma^2 = \sigma^2 \frac{1}{n} \sum_{t=1}^n (\varepsilon_t^2 - 1) + o_P\left(\frac{1}{\sqrt{n}}\right). \quad (2.9)$$

Now note that

$$\hat{F}_n(y) = \sum_{t=1}^n v_{n,t} I\left\{ \varepsilon_t \leq \frac{(\hat{m} - m)(X_{t-1})}{\sigma} + y \frac{\hat{\sigma}}{\sigma} \right\}$$

and with arguments analogous to the proof of Theorem 3.1 in Selk & Neumeier (2012) which leads to (2.4) (see also the proof of Lemma 1 in Dette, Pardo-Fernández & Van Keilegom (2009) or the proof of Theorem 3.1 in Müller, Schick & Wefelmeyer (2009)) one obtains that

$$\begin{aligned} \hat{F}_n(y) &= \frac{1}{n} \sum_{t=1}^n I\{\varepsilon_t \leq y\} + f(y) \sum_{t=1}^n v_{n,t} \left( \frac{(\hat{m} - m)(X_{t-1})}{\sigma} + y \frac{\hat{\sigma}^2 - \sigma^2}{2\sigma^2} \right) \\ &\quad + o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

uniformly with respect to  $y \in \mathbb{R}$ . The assertion follows from (2.5) and (2.9).  $\square$

**Corollary 2.5** *Under model (2.8) and the assumptions stated in the appendix under the null hypothesis  $H_0$  of normal innovations*

$$T_n = n \int_{\mathbb{R}} (\hat{F}_n(y) - \Phi(y))^2 \varphi(y) dy$$

*converges in distribution to  $T$  defined in Theorem 2.1.*

**Proof.** The result immediately follows from Lemma 2.4 and Theorem 2.1.  $\square$

Thus using the critical values from Table 1 we obtain a completely nonparametric consistent distribution-free asymptotic level  $\alpha$  test for Gaussianity of the innovations in AR models (see also Remark 2.3 which holds here analogously).

## 2.3 ARCH model

In this section we consider a nonparametric ARCH model of order one, i. e.

$$X_t = \sigma(X_{t-1})\varepsilon_t, \tag{2.10}$$

where the innovations  $\varepsilon_t$ ,  $t \in \mathbb{Z}$ , are iid and centered with unit variance and  $\varepsilon_t$  is independent of the past  $X_s$ ,  $s \leq t-1$ . Thus we have model (2.1) with conditional mean  $m \equiv 0$ . Our aim is to test the null hypothesis  $H_0$  of normal innovations. To this end let  $\sigma^2$  be estimated by the kernel estimator

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right) X_i^2}{\sum_{i=1}^n K\left(\frac{x-X_{i-1}}{c_n}\right)}$$

and the residuals be defined as

$$\hat{\varepsilon}_t = \frac{X_t}{\hat{\sigma}(X_{t-1})},$$

whereas  $F_n$  is as in (2.3). Let again  $\Phi$  denote the standard normal distribution function and let  $\mathbb{B}$  denote a standard Brownian bridge on  $[0, 1]$ . Then we obtain the following result.

**Theorem 2.6** *Under the assumptions stated in the appendix under the null hypothesis of Gaussian innovations in the ARCH model (2.10) the test statistic*

$$T_n = n \int_{\mathbb{R}} (\hat{F}_n(y) - \Phi(y))^2 \varphi(y) dy$$

*converges in distribution to  $T = \int_0^1 \mathbb{B}^2(u) du$ .*

**Proof.** In the expansion (2.4) the estimation of  $m$  is not present while (2.6) stays valid for the new estimator  $\hat{\sigma}$ . Thus it follows that

$$\hat{F}_n(y) = \frac{1}{n} \sum_{t=1}^n \left( I \{ \varepsilon_t \leq y \} + \frac{yf(y)}{2} (\varepsilon_t^2 - 1) \right) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

and the stochastic process

$$\sqrt{n} \left( \hat{F}_n(y) - F(y) \right), \quad y \in \mathbb{R},$$

converges weakly to a centered Gaussian process  $(\mathbb{K}(y))_{y \in \mathbb{R}}$  with

$$\begin{aligned} \text{Cov}(\mathbb{K}(y), \mathbb{K}(z)) &= F(y \wedge z) - F(y)F(z) \\ &\quad + f(y)yE[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \leq z\}] + f(z)zE[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \leq y\}] \\ &\quad + f(y)f(z)yz(E[\varepsilon_1^4] - 1). \end{aligned}$$

As the innovations are standard normally distributed we have  $E[(\varepsilon_1^2 - 1)I\{\varepsilon_1 \leq z\}] = -z\varphi(z)$ ,  $f = \varphi$  and  $E[\varepsilon_1^4] = 3$ . Hence,

$$\text{Cov}(\mathbb{K}(y), \mathbb{K}(z)) = \Phi(y \wedge z) - \Phi(y)\Phi(z).$$

The assertion follows by the continuous mapping theorem noting that  $\mathbb{K} \circ \Phi^{-1}$  is a standard Brownian bridge.  $\square$

For convenient reference we state the critical values for the test in Table 2. Here  $c_\alpha$  is the  $(1 - \alpha)$ -quantile of  $T = \int_0^1 \mathbb{B}^2(u) du$ , see Shorack & Wellner (1986), p. 147.

nominal level $\alpha$	0.15	0.1	0.05	0.02	0.01
critical value $c_\alpha$	0.284	0.347	0.461	0.6198	0.743

Table 2: *Asymptotic critical values for the Cramér-von Mises test for normality of the innovations in the ARCH model.*

### 3 Simulations

To examine the performance of the test on small samples we consider AR(1) models and ARCH(1) models, for which we compare the results under the assumption of an AR-ARCH model like (2.1) and under the assumption of an AR model like (2.8) (respectively ARCH like (2.10)).

For the AR(1) case we consider the models

$$X_t = 0.5 \cdot X_{t-1} + \varepsilon_t, \quad \varepsilon_1, \dots, \varepsilon_n \sim \tilde{F}_\zeta,$$

where  $\tilde{F}_\zeta$  denotes the skew-normal distribution with location parameter

$$-\sqrt{\frac{2\pi((5\zeta)^2 + (5\zeta)^4)}{\pi^2 + (2\pi^2 - 2\pi) \cdot (5\zeta)^2 + (\pi^2 - 2\pi) \cdot (5\zeta)^4}},$$

scale parameter  $(\pi(1 + (5\zeta)^2)^{1/2}/(\pi + (\pi - 2)(5\zeta)^2)^{1/2})$  and shape parameter  $5\zeta$  for different values of  $\zeta$ . For  $\zeta = 0$  this is the standard normal distribution. The rejection probabilities for 500 repetitions and level 5% are displayed in Table 3 and Figure 1 for the AR-ARCH model (2.1) and the AR model (2.8) respectively. It can be seen that the level is approximated well and the power increases for increasing parameter  $\zeta$  as well as for increasing sample size  $n$ .

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
AR-ARCH								
$n = 100$	4.8	5.4	8.6	12.2	26.2	48.8	62	75.4
$n = 200$	5	6	8.8	18.6	44	83.6	94	96.2
AR								
$n = 100$	5	7.4	9	15.6	27	53.2	67.4	76.6
$n = 200$	5	6.8	8.8	19	41.8	77.8	93.6	98

Table 3: Rejection probabilities obtained from AR(1) models with skew-normally distributed innovations

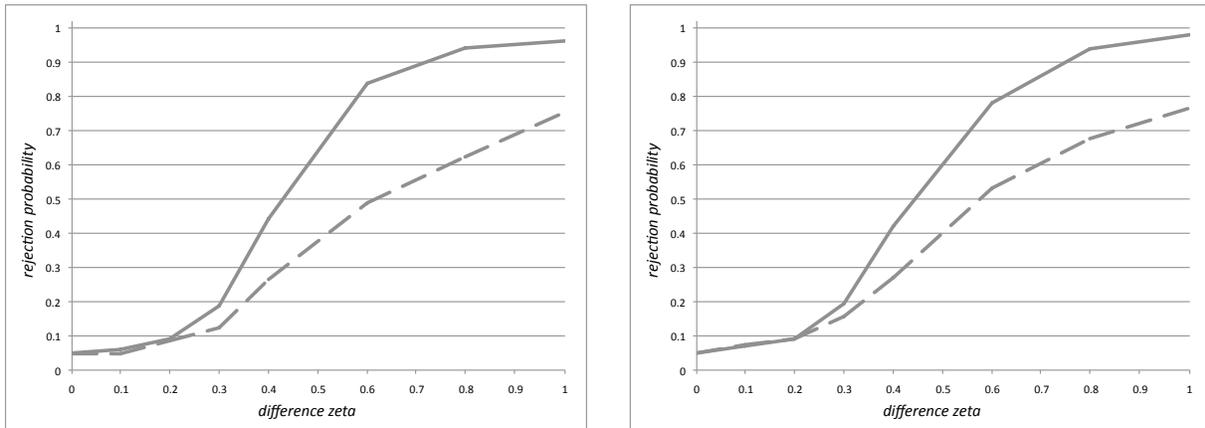


Figure 1: Rejection probabilities obtained from AR(1) models with skew-normally distributed innovations for  $n = 100$  (dashed curve) and  $n = 200$  (solid curve). On the left panel the results for the AR-ARCH model are shown and on the right the ones for the AR model.

We also examine ARCH(1) models with the same innovation distribution,

$$X_t = \sqrt{0.75 + 0.25X_{t-1}^2} \cdot \varepsilon_t, \quad \varepsilon_1, \dots, \varepsilon_n \sim \tilde{F}_\zeta$$

for different values of  $\zeta$ . The rejection probabilities for 500 repetitions and level 5% are shown in Table 4 and Figure 2 for the AR-ARCH model (2.1) and the ARCH model (2.10)

respectively. The asymptotic level is approximated well and the power increases with increasing  $\zeta$  as well as with increasing  $n$ . For the ARCH model the increase with  $\zeta$  for small  $n$  is not as pronounced as for the models considered before, therefore we additionally examined this model with sample size  $n = 500$  for which an increase comparable to those before can be observed.

%	$\zeta = 0$	$\zeta = 0.1$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0.6$	$\zeta = 0.8$	$\zeta = 1$
AR-ARCH								
$n = 100$	5.2	6.6	8.2	15.8	33	61.8	72.8	82.6
$n = 200$	4.8	6.2	7.8	22	47.4	82.2	94.8	98.8
ARCH								
$n = 100$	5	6.2	6.4	8	16.2	24.4	31.2	39.6
$n = 200$	5.2	6	7.2	10.4	15	34	48.8	55.8
$n = 500$	5	5.6	6.8	14.2	31.8	68.6	87	91.2

Table 4: Rejection probabilities obtained from ARCH(1) models with skew-normally distributed innovations

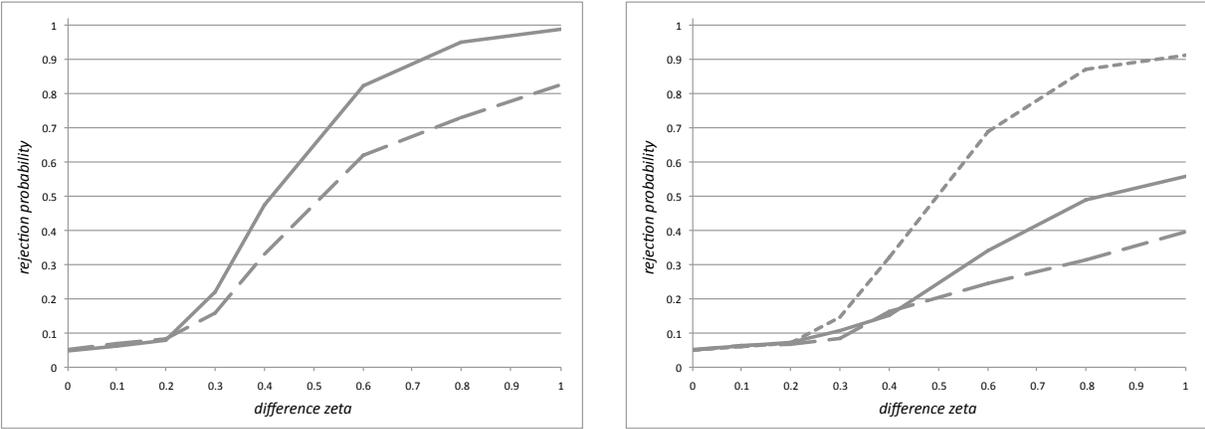


Figure 2: Rejection probabilities obtained from ARCH(1) models with skew-normally distributed innovations for  $n = 100$  (dashed curve),  $n = 200$  (solid curve) and  $n = 500$  (dotted curve). On the left panel the results for the AR-ARCH model are shown and on the right panel the ones for the ARCH model.

To further study the power of the testing procedure, we examine the same models with Student-t distributed innovations with different degrees of freedom. Due to the fact that  $\text{Var}(\varepsilon_t)$  has to be one for all  $t$ , the Student-t distribution was standardized. The models are

AR(1)

$$X_t = 0.5 \cdot X_{t-1} + \varepsilon_t, \quad \varepsilon_1, \dots, \varepsilon_n \sim \mathcal{St}(\zeta)$$

and ARCH(1)

$$X_t = \sqrt{0.75 + 0.25X_{t-1}^2} \cdot \varepsilon_t, \quad \varepsilon_1, \dots, \varepsilon_n \sim \mathcal{St}(\zeta)$$

for different values of  $\zeta$ . The rejection probabilities for 500 repetitions and level 5% are displayed in Table 5, Figure 3 (AR(1) model) and Figure 4 (ARCH(1) model). As before we compare the results under the assumption of an AR-ARCH model like (2.1) and under the assumption of an AR model like (2.8) (respectively ARCH like (2.10)). It can be seen that the power is good and increases for increasing sample size  $n$  while it decreases for increasing parameter  $\zeta$ , because the Student-t distribution converges to the standard normal distribution for increasing degree of freedom.

%	$\zeta = 3$	$\zeta = 4$	$\zeta = 5$	$\zeta = 6$	$\zeta = 7$	$\zeta = 3$	$\zeta = 4$	$\zeta = 5$	$\zeta = 6$	$\zeta = 7$
AR-ARCH										
$n = 100$	61.4	49.8	35.2	27.8	23.6	60.4	45.2	35.8	26.4	22.6
$n = 200$	90	71.8	52.2	45.4	34.6	91	72.2	52.4	37.2	31
$n = 500$	100	98.8	89.6	77.6	62	100	98.4	86.6	75	59.2
AR/ARCH										
$n = 100$	23.4	15.8	15.4	11	9.8	18	9.8	7.6	7.2	5.2
$n = 200$	54.8	25.6	10.8	9.8	6.4	47.2	22.4	12	6.6	6.2
$n = 500$	99	74.6	41.8	22.4	18.6	96.8	70.2	37	19.6	13.4

Table 5: Rejection probabilities obtained from AR(1) models (left) and ARCH(1) models (right) with  $\mathcal{St}(\zeta)$  distributed innovations

**Simulation setting.** For each simulation  $10 \cdot n$  observations  $X_t$  were generated. For the test the last  $n$  observations were used. This was done to ensure that the process is in balance.

The empirical processes were built with weight function  $w_n = I_{[-\log(n), \log(n)]}$ . The Nadaraya-Watson estimators  $\hat{m}$  and  $\hat{\sigma}$  were calculated with Gaussian kernel. This is not compatible with all assumptions, e.g. the support of the kernel is not compact. However this has negligible effect on the simulations because the Gaussian kernel decreases exponentially fast at the tails. The bandwidth was chosen according to the assumptions by a rule of thumb as  $\hat{\sigma}^2 n^{-\frac{2}{6+\sqrt{3}}}$  with  $\hat{\sigma}^2 = \sum_{t=1}^n v_{n,t} (X_t - \hat{m}(X_{t-1}))^2$ .

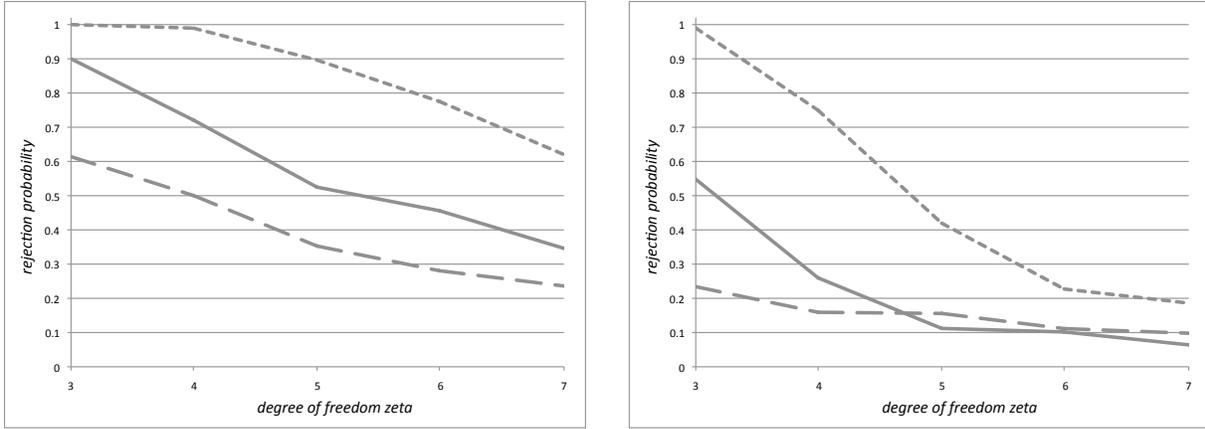


Figure 3: Rejection probabilities obtained from  $AR(1)$  models with  $St(\zeta)$  distributed innovations for  $n = 100$  (dashed curve),  $n = 200$  (solid curve) and  $n = 500$  (dotted curve). On the left panel the results for the AR-ARCH model are shown and on the right panel the ones for the AR model.

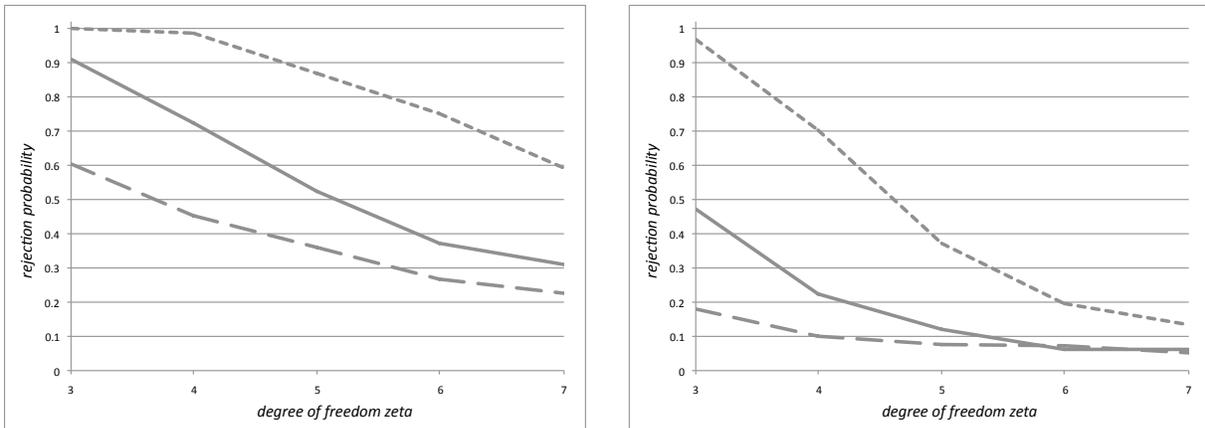


Figure 4: Rejection probabilities obtained from  $ARCH(1)$  models with  $St(\zeta)$  distributed innovations for  $n = 100$  (dashed curve),  $n = 200$  (solid curve) and  $n = 500$  (dotted curve). On the left panel the results for the AR-ARCH model are shown and on the right panel the ones for the ARCH model.

## 4 Further examples

### 4.1 Testing for linear $AR(1)$

As was mentioned in the introduction other distribution-free specification tests for the AR-ARCH model can be derived once the Gaussianity of the errors has been established. In this section we will study in detail a lack-of-fit test for the linear  $AR(1)$  model. For the method

compare Van Keilegom, González Manteiga, & Sánchez Sellero (2008) in a nonparametric regression model with independent observations. Similarly one can derive tests, e. g., for parametric ARCH models.

For simplicity here we assume that  $E[X_0] = 0$  as this is a typical assumption in AR models. We consider the model

$$X_t = m(X_{t-1}) + \sigma\varepsilon_t, \quad (4.1)$$

where the innovations  $\varepsilon_t$ ,  $t \in \mathbb{Z}$ , are iid standard normal and  $\sigma$  is an unknown positive constant. Further,  $\varepsilon_t$  is independent of the past  $X_s$ ,  $s \leq t-1$ . Our aim is to test the null hypothesis

$$\tilde{H}_0 : \exists \vartheta \in (-1, 1) \text{ s. t. } m(x) = \vartheta x.$$

See Hong-zhi & Bing (1991) or Koul & Stute (1999) for other procedures to test for  $\tilde{H}_0$ .

Let  $\hat{\vartheta}$  denote any (under  $H_0$ )  $\sqrt{n}$ -consistent estimator for  $\vartheta$ , e. g. the Yule-Walker or ordinary least squares estimator under suitable regularity assumptions. Now define residuals under the null as

$$\hat{\varepsilon}_{0,t} = \frac{X_t - \hat{\vartheta}X_{t-1}}{\hat{\sigma}}$$

with  $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n (X_t - \hat{\vartheta}X_{t-1})^2$ , and

$$\hat{F}_{0,n}(y) = \sum_{t=1}^n I\{\hat{\varepsilon}_{0,t} \leq y\}$$

Let further  $\hat{F}_n$  be defined as in section 2.2. Then we have the following result.

**Theorem 4.1** *Under the assumptions stated in the appendix for model (4.1) with Gaussian innovations we have under  $\tilde{H}_0$  that*

$$T_n = 2\sqrt{\pi}n \int_{\mathbb{R}} (\hat{F}_n(y) - \hat{F}_{0,n}(y))^2 dy$$

*converges in distribution to a  $\chi_1^2$ -distributed random variable.*

**Proof.** Let  $\vartheta_0$  denote the ‘true’ parameter under  $H_0$ . Similar to the proof of Lemma 2.4 we have for standard Gaussian  $\varepsilon_t$  that

$$\begin{aligned} \hat{F}_{0,n}(y) &= \sum_{t=1}^n I\left\{\varepsilon_t \leq \frac{(\hat{\vartheta} - \vartheta_0)(X_{t-1})}{\sigma} + y \frac{\hat{\sigma}}{\sigma}\right\} \\ &= \frac{1}{n} \sum_{t=1}^n I\{\varepsilon_t \leq y\} + \varphi(y) \sum_{t=1}^n \left(\frac{(\hat{\vartheta} - \vartheta_0)X_{t-1}}{\sigma} + y \frac{\hat{\sigma}^2 - \sigma^2}{2\sigma^2}\right) + o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

uniformly with respect to  $y \in \mathbb{R}$ . Note that for the  $\hat{\sigma}^2$  defined here the last equality in (2.6) also holds. Further,

$$\frac{1}{n} \sum_{t=1}^n X_{t-1} = E[X_0] + o_P(1) = o_P(1)$$

because  $(X_t)_{t \in \mathbb{Z}}$  is ergodic due to its mixing property (see e.g. Doukhan (1994)). Thus with the  $\sqrt{n}$ -consistency of  $\hat{\vartheta}$  we obtain

$$\hat{F}_{0,n}(y) = \frac{1}{n} \sum_{t=1}^n \left( I\{\varepsilon_t \leq y\} + \frac{y\varphi(y)}{2}(\varepsilon_t^2 - 1) \right) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

uniformly with respect to  $y \in \mathbb{R}$ . Now from Lemma 2.4 we have

$$\sqrt{n}(\hat{F}_n(y) - \hat{F}_{0,n}(y)) = \varphi(y) \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t + o_P(1)$$

which converges in distribution to  $\varphi(y)Z$  with a standard normally distributed  $Z$ . Thus  $T_n$  converges in distribution to

$$2\sqrt{\pi} \int_{\mathbb{R}} (\varphi(y))^2 dy Z^2 = Z^2,$$

which is  $\chi_1^2$ -distributed. □

An asymptotically distribution-free level- $\alpha$  test is obtained by rejecting the null hypothesis  $\tilde{H}_0$  of a standard AR(1)-model whenever  $T_n$  is larger than the  $(1 - \alpha)$ -quantile of the  $\chi_1^2$ -distribution.

## 4.2 Testing for multiplicative structure

Under the assumption of Gaussian innovations the test for multiplicative models by Dette, Pardo-Fernández & Van Keilegom (2009) simplifies. Here the null hypothesis to be tested for model (2.1) is

$$\bar{H}_0 : \exists c \text{ s. t. } m = c\sigma.$$

This condition connects ARCH models of the form  $Z_t = s(Z_{t-1})\varepsilon_t$  to model (2.1) by setting  $X_t = Z_t^2 = m(X_{t-1}) + \sigma(X_{t-1})\varepsilon_t$  with  $\varepsilon_t = \varepsilon_t^2 - 1$  and  $c = (E[\varepsilon_1^4] - 1)^{-1/2}$ .

Note that under  $\bar{H}_0$  the constant  $c$  can be estimated  $\sqrt{n}$ -consistently by a least-squares estimator  $\hat{c}$  defined by Dette, Pardo-Fernández & Van Keilegom (2009). Its asymptotic expansion under  $\bar{H}_0$  and under our assumptions is

$$\hat{c} - c = \frac{1}{n} \sum_{t=1}^n \left( -\frac{1}{2}c\varepsilon_t^2 + \varepsilon_t + \frac{1}{2}c \right) \frac{\sigma^4(X_{t-1})}{E[\sigma^4(X_0)]} + o_P\left(\frac{1}{\sqrt{n}}\right), \quad (4.2)$$

see Theorem 5 in the aforementioned paper, but note that under our assumptions given in the appendix the influence of the weight function vanishes asymptotically. Now define

$$\hat{F}_{0,n}(y) = \sum_{t=1}^n v_{n,t} I\{\hat{\varepsilon}_{0,t} \leq y\}$$

with residuals

$$\hat{\varepsilon}_{0,t} = \frac{X_t - \hat{c}\hat{\sigma}(X_{t-1})}{\hat{\sigma}(X_{t-1})},$$

where  $\hat{\sigma}$  is as in (2.2). Let for  $k \in \{4, 8\}$ ,

$$s_k = \sum_{t=1}^n v_{n,t} \hat{\sigma}^k(X_{t-1})$$

and  $\hat{\tau}^2 = s_8/s_4^2 - 1$ . Finally, let  $\hat{F}_n$  be defined as in (2.3). Then we have the following result.

**Theorem 4.2** *Under the assumptions stated in the appendix for model (2.1) with Gaussian innovations we have under  $\bar{H}_0$  that*

$$T_n = \frac{2\sqrt{\pi}n \int_{\mathbb{R}} (\hat{F}_n(y) - \hat{F}_{0,n}(y))^2 dy}{(\frac{3}{4}c^2 + 1)^2 \hat{\tau}^2}$$

converges in distribution a  $\chi_1^2$ -distributed random variable.

**Proof.** We only sketch the main differences to the proof in Dette, Pardo-Fernández & Van Keilegom (2009) due to slightly different assumptions (under which in particular the influence of the weight function is asymptotically negligible). We have

$$\begin{aligned} \hat{F}_{0,n}(y) &= \sum_{t=1}^n v_{n,t} I\left\{\varepsilon_t \leq \frac{(\hat{c}\hat{\sigma} - c\sigma)(X_{t-1})}{\sigma(X_{t-1})} + y \frac{\hat{\sigma}(X_{t-1})}{\sigma(X_{t-1})}\right\} \\ &= \frac{1}{n} \sum_{t=1}^n I\{\varepsilon_t \leq y\} + \varphi(y) \sum_{t=1}^n v_{n,t} \left(\frac{(\hat{c}\hat{\sigma} - c\sigma)X_{t-1}}{\sigma(X_{t-1})} + y \frac{\hat{\sigma}^2(X_{t-1}) - \sigma^2(X_{t-1})}{2\sigma^2(X_{t-1})}\right) \\ &\quad + o_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

uniformly with respect to  $y \in \mathbb{R}$  (compare to (2.4)). From this and (4.2), (2.4), (2.6) one obtains

$$\sqrt{n}(\hat{F}_n(y) - \hat{F}_{0,n}(y)) = \varphi(y) \frac{1}{\sqrt{n}} \sum_{t=1}^n \left(\frac{1}{2}c(\varepsilon_t^2 - 1) - \varepsilon_t\right) \left(1 - \frac{\sigma^4(X_{t-1})}{E[\sigma^4(X_0)]}\right) + o_P(1)$$

which by Th. 2.21 in Fan & Yao (2003) converges in distribution to  $\varphi(y)Z$ , where  $Z$  is centered normally distributed with variance  $(\frac{3}{4}c^2 + 1)\tau^2$  with

$$\tau^2 = E\left[\left(1 - \frac{\sigma^4(X_{t-1})}{E[\sigma^4(X_0)]}\right)^2\right] = \frac{E[\sigma^8(X_0)]}{(E[\sigma^4(X_0)])^2} - 1.$$

Finally the assertion follows because  $\hat{\tau}^2$  consistently estimates  $\tau^2$  because  $\hat{\sigma}$  consistently estimates  $\sigma$  and  $(\sigma(X_t))_{t \in \mathbb{Z}}$  inherits the mixing property of  $(X_t)_{t \in \mathbb{Z}}$  and is therefore ergodic as well. Thus  $T_n$  converges in distribution to

$$2\sqrt{\pi} \int_{\mathbb{R}} (\varphi(y))^2 dy \frac{Z^2}{(\frac{3}{4}c^2 + 1)\tau^2} = \frac{Z^2}{(\frac{3}{4}c^2 + 1)\tau^2},$$

which is  $\chi_1^2$ -distributed. □

One obtains an asymptotically distribution-free test for  $\bar{H}_0$  and thus avoids to implement bootstrap procedures. Note that it is not obvious which kind of bootstrap procedure should be applied here in the context of arbitrary innovation distributions.

## A Technical assumptions

The assumptions are similar to those in Selk & Neumeier (2012) and required for their Theorem 3.1 which we use.

**(K)** The kernel  $K$  is a three times differentiable density with compact support  $[-C, C]$  and  $\sup_{u \in [-C, C]} |K^{(\mu)}(u)| \leq \bar{K} < \infty$ ,  $\mu = 0, 1, 2, 3$ . Moreover  $K(C) = K(-C) = K'(C) = K'(-C) = 0$  and  $\int K(u)udu = 0$ .

**(C)** The sequence of bandwidths  $c_n$  fulfills

$$nc_n^4(\log n)^\eta \rightarrow 0, \quad \frac{(\log n)^\eta}{nc_n^{2+\sqrt{3}}} \rightarrow 0 \text{ for all } \eta > 0.$$

**(I)** For the interval  $I_n = [a_n, b_n]$  some  $r_I < \infty$  exists, such that  $(b_n - a_n) = O(\log(n)^{r_I})$ . Moreover  $\left( \int_{-\infty}^{a_n+\kappa} f_{X_0}(x)dx + \int_{b_n-\kappa}^{\infty} f_{X_0}(x)dx \right) = o((\log n)^{-1})$ , where  $f_{X_0}$  denotes the density of  $X_0$ .

**(W)** The weight function  $w_n : \mathbb{R} \rightarrow [0, 1]$  fulfills  $w_n(x) = 1$  for  $x \in [a_n + \kappa, b_n - \kappa]$  and  $w_n(x) = 0$  for  $x \notin [a_n, b_n]$  for some  $\kappa > 0$  independent of  $n$  and is three times differentiable such that  $\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}} |w_n^{(\mu)}(x)| < \infty$  for  $\mu = 1, 2, 3$ .

**(F)** The innovations  $\varepsilon_j$ ,  $j \in \mathbb{Z}$ , are independent and standard normally distributed.

**(E)** Some  $b > 1 + \sqrt{3}$  exists such that  $E[|X_0|^{2b}] < \infty$ .

**(X)** The observation process  $(X_j)_{j \in \mathbb{Z}}$  is  $\alpha$ -mixing with mixing-coefficient  $\alpha(n) = O(n^{-\beta})$  for some

$$\beta > \max \left( 2 \frac{(3 + \sqrt{3})b + 2 + \sqrt{3}}{(1 + \sqrt{3})b - 2(2 + \sqrt{3})}, 7 \right).$$

Their density  $f_{X_0}$  is bounded and four times differentiable with bounded derivatives. The density is also bounded away from zero on compact intervals and some  $r_f < \infty$  exists, such that  $\frac{1}{\inf_{x \in I_n} f_{X_0}(x)} = O((\log n)^{r_f})$ .

(Z) It holds that

$$\sup_{x \in J_n} \left( (|m(x)| + |\sigma(x)|)^{2k} f_{X_0}(x) \right) = O(1)$$

and there exists some  $1 \leq j^* < \infty$  such that

$$\sup_{x, x' \in J_n} \left( (|m(x)| + |\sigma(x)|)^k (|m(x')| + |\sigma(x')|)^k f_{X_0, X_{j-1}}(x, x') \right) = O(1)$$

is valid for all  $j > j^* + 1$ , for  $k = 1, 2$ ,  $n \rightarrow \infty$  with  $J_n = [a_n - (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n, b_n + (C + c_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}) c_n]$ .

(M) The regression function  $m$  and the scale function  $\sigma$  are four times differentiable and there exist some  $r_q, r_s < \infty$  and  $q_n, q_n^\sigma$  with  $\sup_{x \in [a_n - Cc_n, b_n + Cc_n]} |m^{(\mu)}(x)| = O(q_n)$ ,  $\sup_{x \in [a_n - Cc_n, b_n + Cc_n]} |\sigma^{(\mu)}(x)| = O(q_n)$ ,  $\mu = 0, 1, 2, 3, 4$  and  $\frac{1}{\inf_{x \in I_n} |\sigma(x)|} = O(q_n^\sigma)$ , where  $q_n = O((\log n)^{r_q})$ ,  $q_n^\sigma = O((\log n)^{r_s})$ ,  $(q_n)^{-1} = O(1)$ ,  $(q_n^\sigma)^{-1} = O(1)$ .

**Remark A.1** Note that the mixing condition in (X) is weaker than the one in Selk & Neumeyer (2012). This is due to the nonsequential case that is examined here for which the proof of Lemma B.3 in the aforementioned paper can be simplified. Further note that the assumptions above are formulated under the null hypothesis  $H_0$  of Gaussian innovations. To obtain consistency of the testing procedures one needs to replace assumption (F) by

(F') The innovations  $\varepsilon_j, j \in \mathbb{Z}$ , are independent and identically distributed with distribution function  $F$ . Their density  $f$  is continuously differentiable and  $\sup_{t \in \mathbb{R}} |f(t)t| < \infty$  as well as  $\sup_{t \in \mathbb{R}} |f'(t)t^2| < \infty$ . Further,  $E[|\varepsilon_1|^{2b}] < \infty$  for  $b$  from assumption (E).

For AR model (2.8) with  $\eta_t = \sigma \varepsilon_t$  some conditions in (Z) and (M) are redundant because  $\sigma$  is a constant function. A similar remark holds for the ARCH model (2.10) where  $m \equiv 0$ . ■

## References

- Akritis, M. & Van Keilegom, I. (2001). *Nonparametric estimation of the residual distribution*. Scand. J. Statist. 28, 549-567.
- Araveeporn, A. (2011). *Developing Nonparametric Conditional Heteroscedastic Autoregressive Nonlinear Model by Using Maximum Likelihood Method*. Chiang Mai J. Sci. 38, 331-345.

- Brockwell, P. J. & Davis, R. A. (2006). *Time Series: Theory and Methods*. Springer, New York.
- Dette, H., Pardo-Fernández, J. C. & Van Keilegom, I. (2009). *Goodness-of-Fit Tests for Multiplicative Models with Dependent Data*. *Scand. J. Statist.* 36, 782-799.
- Doukhan, P. (1994). *Mixing, Properties and Examples*. Springer, New York.
- Doukhan, P. & Ghindès, M. (1983). *Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1*. *Stochastic Process. Appl.* 15, 271-293.
- Ducharme, G. & Lafaye de Micheaux, P. (2004). *Goodness-of-fit tests of normality for the innovations in ARMA models*. *J. Time Ser. Anal.* 25, 373-395
- Franke, J., Kreiß, J.-P. & Mammen, E. (2009). *Nonparametric modelling of financial time series*. T. G. Andersen (ed) et al, *Handbook of Financial Time Series*. Springer, Berlin, 927-952.
- Härdle, W. & Tsybakov, A. (1997). *Local polynomial estimators of the volatility function in nonparametric autoregression*. *J. Econometrics* 81, 223-242.
- Härdle, W. & Vieu P. (1991). *Kernel regression smoothing of time series*. *J. Time Ser. Anal.* 13, 209-232.
- Hong-zhi, A. & Bing, C. (1991). *A Kolmogorov-Smirnov type statistic with application to test for nonlinearity in time series*. *Internat. Statist. Rev.* 59, 287-307.
- Horváth, L., Kokoszka, P. & Teyssière, G. (2004). *Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals*. *J. Stat. Comput. Simul.* 74 , 469-485.
- Klar, B., Lindner, F. & Meintanis, S. G. (2011). *Specification tests for the error distribution in GARCH models*. *Comput. Statist. Data Anal.*, to appear. preprint available at <http://www.math.kit.edu/stoch/klar/seite/veroeffentlichungen/de>
- Koul, H. L. & Ling, S. (2006). *Fitting an error distribution in some heteroscedastic time series models*. *Ann. Statist.* 34, 994-1012.
- Koul, H. L. & Stute, W. (1999). *Nonparametric model checks for time series*. *Ann. Statist.* 27, 204-236.
- Masry, E. & Tjøstheim, D. (1995). *Nonparametric estimation and identification of nonlinear ARCH time series*. *Econometric Theory* 11, 258-289.

- Müller, U. U., Schick, A. & Wefelmeyer, W. (2009). *Estimating the innovation distribution in nonparametric autoregression*. Probab. Theory Relat. Fields 144, 53-77.
- Neumeier, N. & Van Keilegom, I. (2010). *Estimating the error distribution in nonparametric multiple regression with applications to model testing*. J. Multiv. Anal. 101, 1067-1078.
- Robinson, P. M. (1983). *Nonparametric estimators for time series*. J. Time Ser. Anal. 4, 185-207.
- Selk, L. & Neumeier, N. (2012). *Testing for a change of the innovation distribution in nonparametric autoregression - the sequential empirical process approach*. Preprint available at <http://preprint.math.uni-hamburg.de/public/ims.html>
- Shorack, G. R. & Wellner, J. A. (1986). *Empirical Processes with Applications of Statistics*. Wiley, New York.
- Stephens, M. A. (1976). *Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters*. Ann. Statist. 4, 357-369.
- Van Keilegom, I., González Manteiga, W. & Sánchez Sellero, C. (2008). *Goodness-of-fit tests in parametric regression based on the estimation of the error distribution*. TEST, 17, 401-415.