Goodness-of-fit tests for multiplicative models with dependent data

Holger Dette[*]   Juan Carlos Pardo-Fernández[†]

Fakultät für Mathematik   *Departamento de Estatística e IO*

Ruhr-Universität Bochum   *Universidade de Vigo*

Ingrid Van Keilegom[‡]

Institute of Statistics

Université catholique de Louvain

September 19, 2007

Abstract

Several classical time series models can be written as a regression model of the form $Y_t = m(X_t) + \sigma(X_t)\varepsilon_t$, where (X_t, Y_t), $t = 0, \pm 1, \pm 2, \ldots$, is a bivariate strictly stationary process. Some of those models, such as ARCH or GARCH models, share the property of proportionality of the regression function, m, and the scale function, σ. In this article, we present a procedure to test for this feature in a nonparametric context, which is a preliminary step to identify certain time series models. The test is based on the difference between two nonparametric estimators of the distribution of the regression error. Asymptotic results are proved and some simulations are shown in the paper in order to illustrate the finite sample properties of the procedure.

Key Words: Dependent data; error distribution; kernel smoothing; location-scale model; multiplicative model; mixing sequences; nonparametric regression; bootstrap.

1 Introduction and motivation of the test

Let (X_t, Y_t), $t = 0, \pm 1, \pm 2, \ldots$, be a bivariate strictly stationary discrete time process, and assume that there exists a nonparametric relationship of the form

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t, \tag{1}$$

where $m(x) = E(Y_t|X_t = x)$ is an unknown regression function, $\sigma^2(x) = \mathrm{Var}(Y_t|X_t = x)$ is an unknown conditional variance function, and ε_t are the unobservable errors satisfying $E(\varepsilon_t|X_t) = 0$ and $\mathrm{Var}(\varepsilon_t|X_t) = 1$.

This general nonparametric framework includes typical time series models, where X_t represents lagged variables of Y_t (for instance $X_t = Y_{t-1}$). In particular, consider the ARCH(1) model [see, for example, Fan and Yao (2003), page 143]

$$Z_t = (a_0 + a_1 Z_{t-1}^2)^{1/2}\epsilon_t,$$

for some constants $a_0, a_1 \geq 0$, $a_1 < 1$, where ϵ_t has mean 0 and variance 1 and is independent of Z_{t-1} for all t. Straightforward manipulations allow us to write the above model as

$$Z_t^2 = (a_0 + a_1 Z_{t-1}^2) + c^{-1}(a_0 + a_1 Z_{t-1}^2)\varepsilon_t, \tag{2}$$

where $\varepsilon_t = c(\epsilon_t^2 - 1)$ and c is a positive scaling factor given by $c^2 = [E(\epsilon_t^4) - 1]^{-1}$. Clearly model (2) can be identified as a particular case of the general model (1) by simply taking $Y_t = Z_t^2$, $X_t = Z_{t-1}^2$, $m(X_t) = a_0 + a_1 X_t$ and $\sigma(X_t) = c^{-1}(a_0 + a_1 X_t)$. Note that the new errors verify $E(\varepsilon_t) = cE(\epsilon_t^2 - 1) = 0$ and $\mathrm{Var}(\varepsilon_t) = c^2[E(\epsilon_t^4) - 1] = 1$. We have therefore seen that the ARCH(1) model can be written in the form (1) with the peculiarity that the regression function is proportional to the square root of the variance function, that is $m(\cdot) = c\sigma(\cdot)$, where the constant c only depends on the error distribution.

This feature is not exclusive for ARCH models, but it holds for other time series models with a multiplicative structure of the form $Z_t = \sigma_t \epsilon_t$. Different choices of the variance function lead to different models: see, for instance, GARCH, exponential GARCH (EGARCH) and fractionally integrated ARCH (FIARCH) in Fan and Yao (2003), autoregressive conditional duration models in Engle and Russell (1998), or the extension of GARCH proposed by Carroll, Härdle and Mammen (2002).

In other contexts, several authors discussed the problem of estimating and testing the regression function under the assumption of a constant coefficient of variation, which also corresponds to the situation described above. For example McCullagh and Nelder (1989) considered generalized linear models, Carroll and Ruppert (1988) investigated a parametric model with a constant coefficient of variation, while Eagleson and Müller (1997) considered the problem of nonparametric estimation of the regression function in a model where the standard deviation function is proportional to the regression function.

In this paper, we derive a test for the null hypothesis

$$H_0 : m(\cdot) = c\sigma(\cdot), \tag{3}$$

where c is a fixed positive value, in general unknown, versus the general alternative $H_1 : m(\cdot) \neq c\sigma(\cdot)$. In time series analysis, this hypothesis is a preliminary step to be tested before applying other procedures, such as specific tests for ARCH or GARCH models.

The problem of specification testing for nonparametric regression models for stationary time series has found considerable interest in the recent literature. Most authors investigate test procedures for parametric hypotheses regarding the mean effect $m(x)$ [see e.g. Masry and Tjøstheim (1995), Hjellvik, Yao and Tjøstheim (1998), Fan and Li (1999) or Dette and Spreckelsen (2004) among many others]. On the other hand – to the knowledge of the authors – the problem of testing the hypothesis of a constant coefficient of variation has not been considered in the literature, despite the fact that this characterizes time series models defined by a multiplicative structure.

The paper is organized as follows. In Section 2 we describe the proposed testing procedure, which is based on a comparison of two empirical processes of the standardized nonparametric residuals calculated under the hypothesis of a multiplicative structure and the alternative of a general nonparametric regression model. Some asymptotic results establishing weak convergence of the (appropriately standardized) difference of the processes are stated in Section 3. For the sake of simplicity we consider a bivariate time series, while extensions to more general models are briefly indicated in Section 4. Section 5 presents the results of a small simulation study, which illustrates the finite sample properties of a bootstrap version of the test. The proofs of the main results are complicated

3

and therefore deferred to the Appendix.

2 Testing for multiplicative structure

Our testing procedure is based on the comparison of two estimators of the error distribution, and it can be justified as follows. First, consider the errors of regression model (1) :

$$\varepsilon_t = \frac{Y_t - m(X_t)}{\sigma(X_t)},$$

with distribution function $F_\varepsilon(y) = P(\varepsilon_t \leq y)$. Note that the stationarity of the process ensures that the distribution of ε_t is the same for any value of the index t. The same happens for the following random variables

$$\varepsilon_{t0} = \frac{Y_t - c\sigma(X_t)}{\sigma(X_t)},$$

with distribution function $F_{\varepsilon 0}(y) = P(\varepsilon_{t0} \leq y)$.

Under the null hypothesis H_0, the random variables ε_t and ε_{t0} are equal, and consequently they have the same distribution. On the other hand, if ε_t and ε_{t0} have the same distribution then necessarily $m(\cdot) = c\sigma(\cdot)$. This idea is stated in the following theorem, whose proof can be found in the Appendix.

Theorem 1 *Let m and σ be continuous functions. The hypothesis $H_0 : m(\cdot) = c\sigma(\cdot)$ (for some $c > 0$ fixed) is valid if and only if the random variables ε_t and ε_{t0} have the same distribution.*

In practice the regression errors are estimated from observations $(X_1, Y_1), \ldots, (X_T, Y_T)$ generated from model (1). For this purpose we consider the following nonparametric estimators of the regression and variance functions :

$$\hat{m}(x) = \sum_{t=1}^{T} W_t(x, h) Y_t \quad \text{and} \quad \hat{\sigma}^2(x) = \sum_{t=1}^{T} W_t(x, h) Y_t^2 - \hat{m}^2(x),$$

where $W_t(x, h) = K((x - X_t)h^{-1})/[\sum_{t'=1}^{T} K((x - X_{t'})h^{-1})]$ are Nadaraya-Watson type weights, K is a known kernel function (typically, a symmetric density) and $h = h_T$ is an appropriate bandwidth sequence converging to 0 with increasing sample size. Also, let $\hat{c}$

be any root-T weakly consistent estimator of the scaling factor c. An obvious example is the statistic

$$\hat{c}_{ls}^2 = \frac{\sum_{t=1}^{T} \hat{m}^2(X_t)(Y_t - \hat{m}(X_t))^2}{\sum_{t=1}^{T} \hat{\sigma}^4(X_t)} \, , \tag{4}$$

which arises from the least squares problem

$$\min_{c^2} \sum_{t=1}^{T} \left(m^2(X_t) - c^2 \sigma^2(X_t) \right)^2 \, .$$

Note that the minimum is attained for

$$c_{\min}^2 = \frac{\sum_{t=1}^{n} m^2(X_t)\sigma^2(X_t)}{\sum_{t=1}^{n} \sigma^4(X_t)} \, .$$

For the construction of the estimator $\hat{c}_{ls}^2$ we replace $\sigma^2(X_t)$ in the numerator by its residual $(Y_t - \hat{m}(X_t))^2$ and in the denominator by $\hat{\sigma}^2(X_t)$. Similarly, $m(X_t)$ is estimated by $\hat{m}(X_t)$. By interchanging the role of $\hat{\sigma}^2(X_t)$ and $(Y_t - \hat{m}(X_t))^2$ alternative estimates can be obtained, but we restrict ourselves to $\hat{c}_{ls}^2$ for the sake of brevity. A structural different estimate can be obtained from the method of moments which yields

$$\hat{c}_{mom}^2 = \left\{ \frac{1}{T} \sum_{t=1}^{T} \left(\frac{Y_t}{\hat{m}(X_t)} - 1 \right)^2 \right\}^{-1} \tag{5}$$

as an estimate of c^2, since $E[(Y_t/m(X_t) - 1)^2] = c^{-2}$ when H_0 holds. Under appropriate assumptions on the stationary process it follows that these estimates are root-T consistent (see Theorems 5 and 6 below).

In the general nonparametric model (1) the error distribution is estimated by the empirical distribution of the estimated residuals, that is

$$\hat{F}_{\varepsilon}(y) = \frac{1}{T} \sum_{t=1}^{T} I \left(\frac{Y_t - \hat{m}(X_t)}{\hat{\sigma}(X_t)} \le y \right) . \tag{6}$$

On the other hand, under the null hypothesis H_0 of a multiplicative model, we can also estimate the error distribution by the empirical versions of the random variables ε_{t0}, i.e.

$$\hat{F}_{\varepsilon 0}(y) = \frac{1}{T} \sum_{t=1}^{T} I \left(\frac{Y_t - \hat{c}\hat{\sigma}(X_t)}{\hat{\sigma}(X_t)} \le y \right) . \tag{7}$$

As seen in Theorem 1, any difference between the two estimators of the error distribution in (6) and (7) gives evidence against the null hypothesis. A typical example is depicted

in Figure 1, where we show the empirical distribution functions $\hat{F}_\varepsilon$ and $\hat{F}_{\varepsilon 0}$ corresponding to the cases (a) $m(x) = \sigma(x) = 1 + 0.1\,x$, and (b) $m(x) = 1 + 0.1\,x$, $\sigma(x) = 0.5\sqrt{|x|}$. The statistical comparison of the two distributions is now performed through the empirical process

$$\hat{W}(y) = T^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y)), \quad -\infty < y < \infty. \tag{8}$$

More precisely, we consider Kolmogorov-Smirnov and Cramér-von Mises type statistics defined over the process (8) :

$$T_{KS} = \sup_y |\hat{W}(y)| \quad \text{and} \quad T_{CM} = \int \hat{W}^2(y) d\hat{F}_\varepsilon(y).$$

The null hypothesis is rejected for large values of the test statistics. In the following section we study the asymptotic properties of the process $\hat{W}(y)$ and – as a corollary – derive the asymptotic limit of the statistics T_{KS} and T_{CM}.
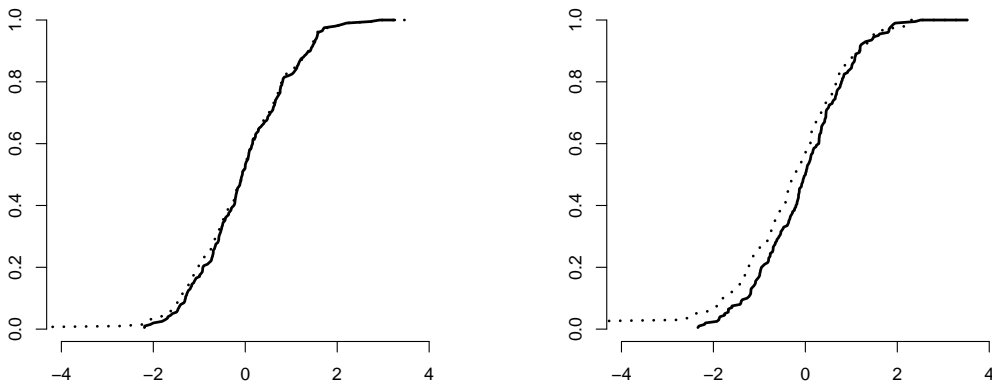

Figure 1: *The empirical processes $\hat{F}_\varepsilon$ (solid line) and $\hat{F}_{\varepsilon 0}$ (dotted line) corresponding to the testing problem (3). The sample size is $T = 200$, $m(x) = 1 + 0.1x$. The left panel corresponds to the null hypothesis of a multiplicative model $H_0 : m(x) = c\sigma(x)$, where both processes are visually non distinguishable. The right panel shows the two processes for the alternative $\sigma(x) = 0.5\sqrt{|x|}$.*

3 Asymptotic results

Let us first introduce some notation. Throughout this paper $F_X(x) = P(X_t \le x)$ denotes the distribution function of the random variable X_t, $F(x, y) = P(X_t \le x, Y_t \le y)$ the joint distribution function of (X_t, Y_t), $F_\varepsilon(y) = P(\varepsilon_t \le y)$ the distribution function of the error, and $F_\varepsilon(y|x) = P(\varepsilon_t \le y | X_t = x)$ the conditional distribution function of the error given $X_t = x$. Note that the distributions of these random variables do not depend on t, because of the strict stationarity of the process (X_t, Y_t), $t \in \mathbb{Z}$. Lower case letters are used for the corresponding densities. Some regularity assumptions are needed in order to prove our main results:

(A1) The process (X_t, Y_t), $t = 0, \pm 1, \pm 2, \ldots$, is strictly stationary and absolutely regular [β-mixing – see Doukham (1994)], with mixing coefficients satisfying $\beta_t = O(t^{-\beta})$, for some $\beta > 2$.

(A2)
(i) X_t is absolutely continuous with density f_X. The support of f_X, which we denote by R_X, is a compact interval of $\mathbb{R}$.
(ii) The functions f_X, m and σ^2 are twice continuously differentiable, $\inf_{x \in R_X} f_X(x) > 0$ and $\inf_{x \in R_X} \sigma^2(x) > 0$.

(A3)
(i) $E(|Y_0|^s) < \infty$ and $\sup_{x \in R_X} E(|Y_0|^s \mid X_0 = x) < \infty$ for some $s > 2 + 2/(\beta - 2)$.
(ii) There exists some j' such that for all $j \ge j'$,

$$\sup_{x_0, x_j \in R_X} E(|Y_0 Y_j|^2 \mid X_0 = x_0, X_j = x_j) f_j(x_0, x_j) < \infty,$$

where $f_j(x_0, x_j)$ denotes the joint density of (X_0, X_j).
(iii) The errors of the regression model satisfy

$$E(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}(X, Y)) = E(\varepsilon_t | X_t) = 0 \quad \text{and} \quad \mathrm{Var}(\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}(X, Y)) = E(\varepsilon_t^2 | X_t) = 1,$$

where $\mathcal{F}_{-\infty}^{t-1}(X, Y)$ denotes the σ-algebra generated by the sequence $\{(X_j, Y_j), j = -\infty, \ldots, t-1\}$.

(A4) The function $F(x, y)$ is continuous in (x, y), differentiable with respect to x and y, and the corresponding density $f(x, y)$ is also continuous in (x, y) and satisfies $\sup_y |y^2 f(x, y)| < \infty$. The same holds for all other partial derivatives of $F(x, y)$ with respect to x and y up to order two.

(A5)

(i) The bandwidth sequence h_T satisfies the following three conditions

$$(\log T)^{-1} T^\theta h_T \to \infty \text{ for } \theta = \frac{\beta - 2 - (1+\beta)/(s-1)}{\beta + 3 - (1+\beta)/(s-1)},$$

$$(\log h_T^{-1})^{-1} T h_T^{3+\delta} \to \infty \text{ for some } \delta > 2,$$

and $(\log T)^{-1} T h_T^5 = O(1)$.

(ii) The kernel K is a symmetric density function with compact support and is twice continuously differentiable.

(A6) The estimator $\hat{c}$ has a stochastic expansion of the form

$$\hat{c} - c = \frac{1}{T} \sum_{t=1}^{T} s(X_t, \varepsilon_t) + o_P(T^{-1/2}),$$

where the function $s(x, e)$ is twice continuously differentiable in (x, e), $E[s(X_t, \varepsilon_t)] = 0$ and $E[s^{2+\delta}(X_t, \varepsilon_t)] < \infty$ for some $\delta > 0$.

As an additional remark, note in the case of independence the assumptions above can be relaxed in the following sense: (A1) disappears since the mixing coefficients are zero; in (A3) it suffices to take $s = 2$ as in Akritas and Van Keilegom (2001), so the assumption is redundant with the model itself; finally, $\theta = 1$ in (A5-i) and hence the first condition on the bandwidth is redundant with the second one.

The asymptotic results can now be stated. In Theorem 2, a stochastic expansion for the difference $\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y)$ is obtained. The weak convergence of the corresponding empirical process is stated in Theorem 3 and the asymptotic distributions of the test statistics under the null hypothesis are presented in Corollary 4. The proofs are complicated and therefore deferred to the Appendix.

Theorem 2 *Assume that conditions (A1)-(A6) are satisfied. Then, under the null hypothesis H_0 of a multiplicative model, the following representation holds:*

$$\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y) = \frac{1}{T}\sum_{t=1}^{T} f_{\varepsilon}(y|X_t)W_t + o_P(T^{-1/2}),$$

uniformly in $-\infty < y < \infty$, where $W_t = 0.5c\varepsilon_t^2 - \varepsilon_t - 0.5c + s(X_t, \varepsilon_t)$, $t = 1, \ldots, T$.

Theorem 3 *Assume that conditions (A1)-(A6) are satisfied. Then, under the null hypothesis H_0 of a multiplicative model, the process $T^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y))$, $-\infty < y < \infty$, converges weakly to a centered Gaussian process $W(y)$ with covariance structure given by $Cov(W(y), W(y')) = \sum_{t=1}^{\infty} Cov(f_{\varepsilon}(y|X_1)W_1, f_{\varepsilon}(y'|X_t)W_t)$.*

Corollary 4 *Assume that conditions (A1)-(A6) are satisfied. Then, under the null hypothesis H_0 of a multiplicative model,*

$$T_{KS} \xrightarrow{d} \sup_{y} |W(y)| \quad and \quad T_{CM} \xrightarrow{d} \int W^2(y)dF_{\varepsilon}(y).$$

To conclude this section, we show that the moment estimator $\hat{c}_{mom}$ and the least squares estimator $\hat{c}_{ls}$, defined in (4) and (5), satisfy condition (A6).

Theorem 5 *Assume that conditions (A1)-(A5) are satisfied. Then, under the null hypothesis H_0 of a multiplicative model,*

$$\hat{c}_{mom} - c = \frac{1}{T}\sum_{t=1}^{T}\big\{ -0.5\,c\,\varepsilon_t^2 + \varepsilon_t + 0.5\,c\big\} + o_P(T^{-1/2}).$$

Note that Theorem 5 implies that the main term in the representation in Theorem 2 equals zero when the moment estimator $\hat{c}_{mom}$ is used. As a consequence, the limiting distribution in Theorem 3 is degenerate in that case.

Theorem 6 *Assume that conditions (A1)-(A5) are satisfied. Then, under the null hypothesis H_0 of a multiplicative model,*

$$\hat{c}_{ls} - c = \frac{1}{T}\sum_{t=1}^{T}\frac{\sigma^4(X_i)}{E[\sigma^4(X_1)]}\{-0.5\,c\,\varepsilon_t^2 + \varepsilon_t + 0.5\,c\} + o_P(T^{-1/2}).$$

9

Note that the representations in Theorem 5 and 6 have a very similar structure, but there appear additional factors $\sigma^4(X_i) \ / \ E[\sigma^4(X_i)]$ in the stochastic expansion of the least squares estimate $\hat{c}_{ls}$ because it is based on the estimate of the regression and variance function. Although the expansion in Theorem 5 appears to be simpler, it turns out that the least squares estimate $\hat{c}_{ls}$ yields better results in the simulation study presented in Section 5.

4   Extensions

Extensions to models with more than one covariate are interesting in practice. Let $\mathbf{X}_t = (X_{t1}, \ldots, X_{td})$ denote now a d-dimensional covariate and let $(\mathbf{X}_t, Y_t), t = 0, \pm 1, \pm 2, \ldots$, be a strictly stationary process. A completely nonparametric model of the form $Y_t = m(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\varepsilon_t$ can be considered again. Unfortunately, the so-called 'curse of dimensionality' not only makes the estimation of the regression and variance function difficult, but also causes some additional problems in the estimation of the error distribution.

For this reason many authors have suggested imposing some structure on the components of the covariate, such as an additive or a multiplicative structure. In generalized additive models each component of the covariate vector has an additive effect on the response and then all of them are combined through a known link function g:

$$m(\mathbf{x}) = m(x_1, \ldots, x_d) = g\left(m_0 + \sum_{j=1}^{d} m_j(x_j)\right),$$

where the partial functions m_j are unknown and m_0 is a constant. Several procedures have been proposed in the literature in order to estimate the functions m_j nonparametrically: backfitting, marginal integration, etc. [see e.g. Hastie and Tibshirani (1990), Linton and Nielsen (1995), Nielsen and Sperlich (2005) among many others]. A more delicate issue, which has not been sufficiently addressed in the literature yet, is the appropriate modeling and estimation of the variance function $\sigma^2(\mathbf{x})$ in a multidimensional setting.

Consider, for instance, the ARCH(q) model. As in (2), this model can be written as

$$Z_t^2 = (a_0 + a_1 Z_{t-1}^2 + \cdots + a_q Z_{t-q}^2) + c^{-1}(a_0 + a_1 Z_{t-1}^2 + \cdots + a_q Z_{t-q}^2)\varepsilon_t.$$

Thus, if we consider the multidimensional covariate $\mathbf{X}_t = (Z_{t-1}^2, \ldots, Z_{t-q}^2)$ and $m(\mathbf{x}) = c\sigma(\mathbf{x}) = a_0 + a_1 x_1 + \cdots + a_q x_q$, we can consider the ARCH(q) model as a special case of a nonparametric regression model where the regression and the standard deviation are proportional and have additive structure.

The results given in this paper for the unidimensional case are still valid in the multidimensional case as long as the estimators of the regression and variance function satisfy certain uniform convergence rates. Some details regarding these rates can be found in the proof of Lemma 7 in the Appendix.

5   Simulation study

In this section we study the finite sample properties of the proposed test based on the Cramér-von Mises statistic T_{CM} in two AR(1) models and one ARCH(1). Note that by Corollary 4 the asymptotic distribution of the statistic T_{CM} depends on several features of the data generating process, which are not known by the statistician. Because the covariance structure in Theorem 3 is difficult to estimate in practice, we have implemented a (smooth) bootstrap test.

To be precise we have estimated the regression function by the local linear estimate $\hat{m}$, while the variance function was estimated by the Nadaraya-Watson estimate defined in Section 2. The local linear estimate in used for the estimation of the regression function in order to address for boundary effects, which would have a substantial influence on the residual based (smooth) bootstrap. The two bandwidths for the estimation of the regression and variance function have been chosen separately by least squares cross validation. The Cramér-von Mises statistic T_{CM} has been calculated from these data in order to compare the distributions of the residuals. For the generation of the bootstrap data we have estimated the constant c in the hypothesis H_0 by the least squares estimate defined in (4), where only data corresponding to the [10%, 90%] range of the explanatory variables X_t was considered for the estimate, in order to make the estimate $\hat{c}_{ls}$ less sensitive with respect to outliers in the residuals caused by boundary effects. In a next step we have

generated bootstrap data

$$Y_t^* = \hat{c}\hat{\sigma}(X_t) + \hat{\sigma}(X_t)\varepsilon_t^* ; \qquad t = 1, \ldots, T , \tag{9}$$

where

$$\varepsilon_t^* = \hat{\varepsilon}_t^* + vZ_t , \tag{10}$$

$\hat{\varepsilon}_1^*, \ldots, \hat{\varepsilon}_T^*$ is an i.i.d. sample from the empirical distribution function

$$\hat{F}_\varepsilon(y) = \frac{1}{T}\sum_{t=1}^T I(\hat{\varepsilon}_t \le y) , \quad \hat{\varepsilon}_t = (Y_t - \hat{m}(X_t))/\hat{\sigma}(X_t)$$

$(t = 1, \ldots, T)$, $Z_1, \ldots, Z_T$ are i.i.d. standard normal random variables and v is a sufficiently small constant, in our case $v = 0.1$. If B bootstrap replications have been performed with ordered outcomes $T_{CM}^{(1)*} \le \cdots \le T_{CM}^{(B)*}$, then the null hypothesis of a multiplicative model is rejected if

$$T_{CM} > T_{CM}^{(\lfloor B(1-\alpha)\rfloor)*} \tag{11}$$

($\lfloor u \rfloor$ denotes the integer part of u). In each scenario 1000 simulation runs with $B = 100$ bootstrap replications have been performed to estimate the empirical level of the bootstrap test.

Example 5.1. We consider a classical (heteroscedastic) AR(1)-model

$$X_t = c(1 + 0.1\ X_{t-1}) + (1 + 0.1\ X_{t-1})\ \varepsilon_t, \quad t \in \mathbb{Z} , \tag{12}$$

where the innovations ε_t are i.i.d. and standard normally distributed. In the first part of Table 1 we show the simulated level of the bootstrap test for the scaling factors $c = 0.5, 1, 1.5$ and sample sizes $T = 50, 100$ and 200. We observe that the level is very well approximated in nearly all cases.

In order to study the power of the test we consider the non-multiplicative model

$$X_t = c(1 + 0.1\ X_{t-1}) + 0.5\sqrt{|X_{t-1}|}\ \varepsilon_t , \ t \in \mathbb{Z} \tag{13}$$

and display the corresponding rejection probabilities in the second part of Table 1. The alternative of a non-constant coefficient of variation is clearly detected with reasonable

12

power. The empirical distribution functions $\hat{F}_\varepsilon$ and $\hat{F}_{\varepsilon 0}$ corresponding to the null hypothesis and alternative have been depicted in Figure 1. Note that the parameter c in this table represents the factor in the null hypothesis and does not correspond to deviations from the null hypothesis.

In order to demonstrate that these results are – in some sense – representative, we consider a second example, namely the autoregressive model

$$X_t = c \cdot \sin(1 + 0.5\ X_{t-1}) + \sin(1 + 0.5\ X_{t-1})\ \varepsilon_t\ ; \quad t \in \mathbb{Z} \tag{14}$$

with alternative

$$X_t = c \cdot \sin(1 + 0.5\ X_{t-1}) + \cos(1 + 0.5\ X_{t-1})\ \varepsilon_t\ . \tag{15}$$

Note that this example corresponds to a more oscillating regression and variance function.

The corresponding results are shown in Table 2 and yield a similar picture. We observe a good approximation of the nominal level and reasonable rejection probabilities under the alternative.

Table 1: *Simulated rejection probabilities of the bootstrap test (11) under the null hypothesis of a multiplicative structure H_0 [model (12)] and the alternative of non multiplicative model [model (13)].*

	T	50			100			200		
	α c	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
(12)	0.5	0.026	0.039	0.084	0.041	0.058	0.107	0.038	0.061	0.106
	1.0	0.024	0.037	0.084	0.039	0.062	0.109	0.037	0.058	0.105
	1.5	0.036	0.052	0.094	0.040	0.057	0.112	0.034	0.053	0.102
(13)	0.5	0.244	0.328	0.416	0.287	0.363	0.491	0.351	0.434	0.570
	1.0	0.176	0.236	0.320	0.185	0.249	0.371	0.203	0.281	0.393
	1.5	0.244	0.288	0.364	0.254	0.301	0.389	0.282	0.312	0.401

Table 2: *Simulated rejection probabilities of the bootstrap test (11) under the null hypothesis of a multiplicative structure H_0 [model (14)] and the alternative of non multiplicative model [model (15)].*

	T	50			100			200		
	α c	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
	0.5	0.025	0.038	0.079	0.026	0.047	0.086	0.033	0.052	0.097
(14)	1.0	0.023	0.034	0.081	0.028	0.041	0.089	0.029	0.043	0.094
	1.5	0.032	0.052	0.100	0.037	0.055	0.106	0.039	0.057	0.108
	0.5	0.232	0.312	0.428	0.356	0.445	0.548	0.593	0.641	0.713
(15)	1.0	0.220	0.266	0.376	0.369	0.420	0.554	0.586	0.664	0.776
	1.5	0.148	0.204	0.312	0.229	0.305	0.394	0.382	0.458	0.602

Example 5.2. We will conclude this section discussing the application of the methodology for testing ARCH-structures. For this purpose we generated data from the ARCH(1)-model

$$Z_t = \sqrt{0.75 + 0.25\, Z_{t-1}^2}\, \epsilon_t\,; \quad t \in \mathbb{Z}, \tag{16}$$

where the random variables ϵ_t are i.i.d. and standard normally distributed. We have applied the bootstrap test to the 'data' $(X_t, Y_t) = (Z_{t-1}^2, Z_t^2)$, where the scaling factor is estimated by the least squares method (4). The corresponding results for sample sizes $T = 50, 100$ and 200 are depicted in Table 3. We observe that the nominal level is rather well approximated for sample sizes $T \geq 100$, whereas the level is over-estimated for the sample size $T = 50$. Next we study the power of the bootstrap test under the alternative

$$Z_t = \sqrt{0.75 + 0.25\, |Z_{t-1}|}\, \epsilon_t\,; \quad t \in \mathbb{Z}, \tag{17}$$

where the random variables ϵ_t are i.i.d. and standard normally distributed. The corresponding results are depicted in the second row of Table 3 and show that the test clearly

14

Table 3: *Simulated rejection probabilities of the bootstrap test (11) for an ARCH(1) structure. The equation (16) corresponds to the "null hypothesis" of an ARCH(1) model while equations (17) and (18) correspond to two alternatives.*

T	50			100			200		
α	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.010
(16)	0.065	0.097	0.137	0.039	0.052	0.102	0.024	0.041	0.082
(17)	0.118	0.170	0.285	0.203	0.274	0.401	0.350	0.441	0.608
(18)	0.138	0.200	0.308	0.164	0.228	0.334	0.265	0.347	0.464

detects the alternative of a non constant coefficient of variation

$$E[Z_t^2 \mid Z_{t-1}^2]/\sqrt{\text{Var}(Z_t^2 \mid Z_{t-1}^2)} \, .$$

As a further alternative we have considered the model

$$Z_t = \sqrt{0.75 + 0.25 \sin(Z_{t-1})} \, \epsilon_t \, ; \quad t \in \mathbb{Z} \, , \tag{18}$$

where the random variables ϵ_t are again standard normally distributed. The corresponding results are depicted in the third row of Table 3 and this alternative is also detected with reasonable power.

Appendix: Proofs

In this Appendix, we include the proofs of the theoretical results.

Proof of Theorem 1. Assume that the random variables ε_t and ε_{t0} have the same distribution. In particular, $E(\varepsilon_{t0}) = E(\varepsilon_t)$ and $\text{Var}(\varepsilon_{t0}) = \text{Var}(\varepsilon_t) = 1$. Consider the representation

$$\varepsilon_{t0} = \frac{Y_t - c\sigma(X_t)}{\sigma(X_t)} = \varepsilon_t + \left(\frac{m(X_t)}{\sigma(X_t)} - c \right) .$$

By applying expectations on both sides of the above expression, we obtain $E[m(X_t)/\sigma(X_t) - c] = 0$. On the other hand, by calculating variances, we get

$$\text{Var}(\varepsilon_{t0}) = \text{Var}(\varepsilon_t) + \text{Var}\left(\frac{m(X_t)}{\sigma(X_t)} - c\right) + 2\text{Cov}\left(\varepsilon_t, \frac{m(X_t)}{\sigma(X_t)}\right).$$

It is easy to check that $\text{Cov}(\varepsilon_t, m(X_t)/\sigma(X_t)) = 0$ since $E(\varepsilon_t) = E[E(\varepsilon_t|X_t)] = 0$ and $E(\varepsilon_t m(X_t)/\sigma(X_t)) = E[E(\varepsilon_t m(X_t)/\sigma(X_t)|X_t)] = E[(m(X_t)/\sigma(X_t))E(\varepsilon_t|X_t)] = 0$. It follows that $E[m(X_t)/\sigma(X_t) - c] = 0$ and $\text{Var}[m(X_t)/\sigma(X_t) - c] = 0$. This means that $m(x) = c\sigma(x)$ with probability 1. The continuity of the functions m and σ allows us to extend the result to the whole support of X_t. The converse implication is obvious. $\qquad\square$

Before writing the proofs of the asymptotic results, we introduce a technical lemma.

Lemma 7 *Assume that conditions (A1)-(A6) are satisfied. Then, the following representation holds :*

$$
\begin{aligned}
\hat{F}_\varepsilon(y) - F_\varepsilon(y) \;=\; & \frac{1}{T}\sum_{t=1}^{T} I\left(\varepsilon_t \le y\right) - F_\varepsilon(y) & (19) \\
& + \int f_\varepsilon(y|x)\frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)} f_X(x)dx + o_P(T^{-1/2}),
\end{aligned}
$$

uniformly in $-\infty < y < \infty$.

Proof. The proof is based on Theorem 1 in Akritas and Van Keilegom (2001) (AVK in the sequel). In that theorem an i.i.d. representation for the empirical process $\hat{F}_\varepsilon(y) - F_\varepsilon(y)$ is established when the error variable ε_t is independent of the covariate X_t, and when it is assumed that the data (X_t, Y_t), $t = 1, \ldots, T$, are i.i.d.

We will restrict attention to indicating which steps in the proof of the above theorem need to be modified. All of the notations used below are taken over from that proof. We start by proving Propositions 3-5 in AVK, which are required in the main proof of the theorem. These propositions state that

$$\sup_x |\hat{m}(x) - m(x)| = O_P((\log T)^{1/2}(Th_T)^{-1/2}), \tag{20}$$

$$\sup_x |\hat{\sigma}(x) - \sigma(x)| = O_P((\log T)^{1/2}(Th_T)^{-1/2}), \tag{21}$$

and that

$$\sup_x |\hat{m}'(x) - m'(x)| = O_P((\log T)^{1/2}(Th_T^3)^{-1/2}), \tag{22}$$

$$\sup_x |\hat{\sigma}'(x) - \sigma'(x)| = O_P((\log T)^{1/2}(Th_T^3)^{-1/2}),$$

$$\sup_{x,x'} \frac{|\hat{m}'(x) - m'(x) - \hat{m}'(x') - m'(x')|}{|x - x'|^\delta} = O_P((\log T)^{1/2}(Th_T^{3+2\delta})^{-1/2}),$$

$$\sup_{x,x'} \frac{|\hat{\sigma}'(x) - \sigma'(x) - \hat{\sigma}'(x') - \sigma'(x')|}{|x - x'|^\delta} = O_P((\log T)^{1/2}(Th_T^{3+2\delta})^{-1/2}),$$

for some $\delta > 0$. Regarding the validity of (20), this follows from Theorem 8 in Hansen (2006). In that paper, the uniform consistency of kernel estimators in regression is proved when the data (X_t, Y_t) are assumed to come from a stationary β-mixing process. The rates in (21) and (22) can be obtained in a similar way, taking into account that the regularity conditions imposed in assumption (A2) are stronger than the corresponding ones in Hansen (2006).

We now verify how the proof of Lemma 1 in AVK can be adapted to the present setup. One major change is required in this proof: the condition on the boundedness of the bracketing integral (see equation (20) in AVK) should be replaced by

$$\int_0^\infty \sqrt{\log N_{[]}(\lambda, \mathcal{F}, \|\cdot\|_{2,\beta})} \, d\lambda < \infty, \tag{23}$$

where the class $\mathcal{F}$ is defined as in the proof of AVK, and for any function g,

$$\|g\|_{2,\beta}^2 = \int_0^1 \beta^{-1}(u)Q_g^2(u)du,$$

where β^{-1} is the inverse cadlag of the decreasing function $u \to \beta_{\lfloor u \rfloor}$ ($\lfloor u \rfloor$ being the integer part of u, and β_t being the mixing coefficient) and Q_g is the inverse cadlag of the tail function $u \to P(|g| > u)$ (see Section 4.3 in Dedecker and Louhichi, 2002).

For verifying (23), consider for simplicity the case where $\sigma \equiv 1$, i.e. no brackets for σ

need to be constructed. First consider for $0 \le y < 1$,

$$P\{I(\varepsilon_t \le d^U(X_t)) - I(\varepsilon_t \le d^L(X_t)) > y\}$$

$$= P(d^L(X_t) \le \varepsilon_t \le d^U(X_t))$$

$$= \int P(d^L(x) \le \varepsilon_t \le d^U(x)|x)dF_X(x)$$

$$= \int [F_\varepsilon(d^U(x)|x) - F_\varepsilon(d^L(x)|x)] \, dF_X(x)$$

$$= \int f_\varepsilon(\xi(x)|x)[d^U(x) - d^L(x)] \, dF_X(x)$$

$$\le K\|d^U - d^L\|_1 \le K\|d^U - d^L\|_2 < K\lambda^2,$$

for some $\xi(x)$ between $d^L(x)$ and $d^U(x)$ (see the proof in AVK for the definition of the functions d_L and d_U), whereas for $y = 1$, the above probability equals 0. It follows that the quantile function $Q(u)$ corresponding to the above distribution function equals

$$Q(u) = \begin{cases} 1 & \text{if } 0 \le u < p \\ 0 & \text{if } p \le u \le 1, \end{cases} \tag{24}$$

where $p = P(d^L(X_t) \le \varepsilon_t \le d^U(X_t))$. Hence,

$$\|I(\varepsilon_t \le d^U(X_t)) - I(\varepsilon_t \le d^L(X_t))\|_{2,\beta}^2 = \int_0^p \beta^{-1}(u) \, du \le \beta^{-1}(0)p \le K\lambda^2, \tag{25}$$

for some constant $0 < K < \infty$. This shows that Lemma 1 in AVK continues to hold true in the context of this paper.

Finally, we verify the main proof of Theorem 1 in AVK. Careful verification of the different steps in that proof reveals that the only change is the replacement of $f_\varepsilon(y)$ by $f_\varepsilon(y|x)$ in the main term of the representation. This finishes the proof. $\qquad\square$

Proof of Theorem 2. Lemma 7 states that

$$\hat{F}_\varepsilon(y) - F_\varepsilon(y) = \frac{1}{T} \sum_{t=1}^T I\left(\frac{Y_t - m(X_t)}{\sigma(X_t)} \le y\right) - F_\varepsilon(y) \tag{26}$$

$$+ \int f_\varepsilon(y|x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)} f_X(x) dx + o_P(T^{-1/2})$$

and similarly it can be shown that

$$\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y) = \frac{1}{T} \sum_{t=1}^T I\left(\frac{Y_t - c\sigma(X_t)}{\sigma(X_t)} \le y\right) - F_\varepsilon(y)$$

$$+ \int f_\varepsilon(y|x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{c}\hat{\sigma}(x) - c\sigma(x)}{\sigma(x)} f_X(x) dx + o_P(T^{-1/2}),$$

uniformly in y, provided $\hat{c} - c = O_P(T^{-1/2})$, which follows from (A6) and the central limit theorem for mixing sequences [see, for instance, Theorem 2.20 in Fan and Yao, (2003)]. Now, taking into account that under the null hypothesis $m(x) = c\sigma(x)$, we obtain

$$\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y) = \int f_{\varepsilon}(y|x)\frac{\hat{c}\hat{\sigma}(x) - \hat{m}(x)}{\sigma(x)}f_X(x)dx + o_P(T^{-1/2}).$$

The uniform rates given in (20) ensure that Lemmas 8 and 9 in Pardo-Fernández, Van Keilegom and González-Manteiga (2007) can be applied here (note that in the cited results in that reference the factor $f_{\varepsilon}(y|x)$ does not appear inside the integral, but it does not represent any additional difficulty in the proof):

$$\int f_{\varepsilon}(y|x)\frac{\hat{m}(x) - m(x)}{\sigma(x)}f_X(x)dx = \frac{1}{T}\sum_{t=1}^{T} f_{\varepsilon}(y|X_t)\frac{Y_t - m(X_t)}{\sigma(X_t)} + o_P(T^{-1/2})$$

and

$$\int f_{\varepsilon}(y|x)\frac{\hat{c}\hat{\sigma}(x) - c\sigma(x)}{\sigma(x)}f_X(x)dx$$
$$= \frac{1}{T}\sum_{t=1}^{T} f_{\varepsilon}(y|X_t)\left\{\frac{c(Y_t - m(X_t))^2 - c\sigma^2(X_t)}{2\sigma^2(X_t)} + s(X_t, \varepsilon_t)\right\} + o_P(T^{-1/2}).$$

Hence,

$$\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y)$$
$$= \frac{1}{T}\sum_{t=1}^{T} f_{\varepsilon}(y|X_t)\left\{\frac{c}{2}\left(\frac{Y_t - m(X_t)}{\sigma(X_t)}\right)^2 - \frac{Y_t - m(X_t)}{\sigma(X_t)} - \frac{c}{2} + s(X_t, \varepsilon_t)\right\} + o_P(T^{-1/2}),$$

which equals the representation given in the statement of the theorem. $\qquad\square$

Proof of Theorem 3. Let us consider the following class of functions

$$\mathcal{F} = \left\{(u, e) \to f_{\varepsilon}(y|u)(0.5ce^2 - e - 0.5c + s(u, e)), -\infty < y < +\infty\right\}.$$

Then, $T^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_{\varepsilon}(y))$, $-\infty < y < +\infty$, is asymptotically equivalent to the empirical process indexed by $\mathcal{F}$. General results concerning weak convergence of empirical processes with dependent sequences are provided in Dedecker and Louhichi (2002), where some special cases are treated in detail, such as absolutely regular sequences (β-mixing).

In our case, the mixing coefficients of the sequence (X_t, Y_t) satisfy $\sum_t t^{b-1}\beta_t < \infty$ for $b \in (1, \beta)$ (assumption A1), and, due to property number iv on page 170 in Bradley

19

(1986), so do the corresponding coefficients of the sequences (X_t, ε_t) and $f_\varepsilon(y|X_t)(0.5c\varepsilon_t^2 - \varepsilon_t - 0.5c + s(X_t, \varepsilon_t))$. In this situation, as explained on page 146 in Dedecker and Louhichi (2002), the process $T^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y))$ is weakly convergent as long as the entropy integral $\int \log N_{[\ ]}(\lambda, \mathcal{F}, L_r(P)) d\lambda$ is finite, where $N_{[\ ]}(\lambda, \mathcal{F}, L_r(P))$ is the bracketing number, P is the probability measure induced by the pair (X_t, ε_t) and $L_r(P)$ is the L_r-norm, with $r > 2\beta/(\beta - 1)$.

We will use Corollary 2.7.4 in Van der Vaart and Wellner (1996) to show the finiteness of the entropy integral for the class $\mathcal{F}$ with the L_r-norm. Given that $\mathbb{Z}^2$ is countable, there exists a bijective application ν between $\mathbb{N}$ and $\mathbb{Z}^2$, such that for each $(j_1, j_2) \in \mathbb{Z}^2$ there is one and only one $j \in \mathbb{N}$ such that $\nu(j) = (j_1, j_2)$. This identification satisfies $j_1 = O(j^{1/2})$ and $j_2 = O(j^{1/2})$. Let $I_j = \{(x_1, x_2) \in \mathbb{R}^2 : \max\{|x_1 - j_1|, |x_2 - j_2|\} \le 0.5, \text{ where } (j_1, j_2) = \nu(j)\}$ be the unit square centered at $\nu(j) = (j_1, j_2)$. Obviously, $\mathbb{R}^2 = \bigcup_{j=1}^\infty I_j$.

If $g \in \mathcal{F}$, assumptions (A4) and (A6) ensure that the restriction of g to any I_j is a function of the class $C_M^{1+\delta}(I_j)$, as defined on page 154 of Van der Vaart and Wellner (1996), with $\delta \in (0, 1)$ and where M is a global bound for $\|g\|_{1+\delta}$ on the whole real plane. Under these circumstances, Corollary 2.7.4 in Van der Vaart and Wellner (1996) provides the following bound for the logarithm of the bracketing number (k will denote a generic constant which can vary from line to line along this proof):

$$\log N_{[\ ]}(\lambda, \mathcal{F}, L_r(P)) \le k\lambda^{-V} \left(\sum_{j=1}^\infty p_j^{\frac{V}{V+r}} \right)^{\frac{V+r}{r}}, \tag{27}$$

for any $V > 2/(1 + \delta)$, where $p_j = P((X_t, \varepsilon_t) \in I_j)$. Note that the constant k involves the Lebesgue measure of the set $\{(x_1, x_2) \in \mathbb{R}^2, \|(x_1, x_2) - I_j\| < 1\}$ (which is a constant), and the global bound M. If we put $V = r$ (this can always be done since $r > 2$), then (27) becomes

$$\log N_{[\ ]}(\lambda, \mathcal{F}, L_r(P)) \le k\lambda^{-r} \left(\sum_{j=1}^\infty p_j^{1/2} \right)^2. \tag{28}$$

Assumption (A3-i) concerning moment conditions and the boundedness of X_t ensure that $E(|X_t|^{a_1}|\varepsilon_t|^{a_2}) < \infty$, where $a_1 > 2$ and $a_2 > 2$. This implies that, for j sufficiently large,

$$p_j = \int_{I_j} f_{(X_t, \varepsilon_t)}(x, e) dx de \le \int |x|^{-a_1}|e|^{-a_2} dx de = O(j_1^{-a_1} j_2^{-a_2}) = O(j^{-\frac{a_1 + a_2}{2}}) = o(j^{-2}),$$

where $f_{(X_t, \varepsilon_t)}$ is the density of the pair (X_t, ε_t). This shows that the series in (28) is finite, and therefore $\log N_{[\,]}(\lambda, \mathcal{F}, L_r(P)) \leq k\lambda^{-r}$ and $\int \log N_{[\,]}(\lambda, \mathcal{F}, L_r(P))d\lambda < \infty$. As explained above, this implies the weak convergence of the process. The covariance structure of the limit process follows directly from Theorem 5.2 (central limit theorem for stationary sequences) in Dedecker and Louhichi (2002). □

Proof of Corollary 4. The Continuous Mapping Theorem ensures the convergence of the statistic T_{KS}. For T_{CM}, we will show that $d\hat{F}_\varepsilon(y)$ can be replaced by $dF_\varepsilon(y)$ in the integral. Given that the processes $\hat{W}(y)$ and $T^{1/2}(\hat{F}_\varepsilon(y) - F_\varepsilon(y))$ are weakly convergent (the weak convergence of the second process can be obtained in a similar way as the weak convergence of $\hat{W}(y)$ in Theorem 3), the Skorohod construction (see Serfling, 1980, page 23) implies

$$\sup_y |\hat{W}(y) - W(y)| \to_{a.s.} 0 \tag{29}$$

and

$$\sup_y |\hat{F}_\varepsilon(y) - F_\varepsilon(y)| \to_{a.s.} 0. \tag{30}$$

Now write

$$\left| \int \hat{W}^2(y)d\hat{F}_\varepsilon(y) - \int W^2(y)dF_\varepsilon(y) \right|$$
$$\leq \left| \int (\hat{W}^2(y) - W^2(y))d\hat{F}_\varepsilon(y) \right| + \left| \int W^2(y)d(\hat{F}_\varepsilon(y) - F_\varepsilon(y)) \right|.$$

Both terms on the right hand side of the above inequality are negligible a.s. The first one is $o(1)$ a.s. due to (29). The second one is also $o(1)$ a.s. because of (30) and the application of the Helly-Bray Theorem (see page page 97 in Rao, 1965) to each of the trajectories of the corresponding limit process, which are bounded and continuous almost surely. This concludes the proof. □

Proof of Theorem 5 and 6. For the sake of brevity we restrict ourselves to a derivation of the stochastic expansion for the moment estimate $\hat{c}_{mom}$. The corresponding result for the least squares estimate can be obtained by similar arguments [see Wieczorek (2007)]. Write

$$\hat{c}_{mom}^{-2} - c^{-2} = \frac{c^2 - \hat{c}_{mom}^2}{c^2\hat{c}_{mom}^2} = -\frac{c + \hat{c}_{mom}}{c^2\hat{c}_{mom}^2}(\hat{c}_{mom} - c) = -\frac{2}{c^3}(\hat{c}_{mom} - c) + O_P(|\hat{c}_{mom} - c|^2).$$

21

Hence, it is sufficient to consider

$$\hat{c}_{mom}^{-2} - c^{-2} = \frac{1}{T}\sum_{t=1}^{T}[\hat{\eta}_t^2 - \eta_t^2] + \left[\frac{1}{T}\sum_{t=1}^{T}\eta_t^2 - c^{-2}\right],$$

where $\eta_t = \frac{Y_t}{m(X_t)} - 1 = c^{-1}\varepsilon_t$ and $\hat{\eta}_t = \frac{Y_t}{\hat{m}(X_t)} - 1$. For the first term above, consider

$$\begin{aligned}
\hat{\eta}_t^2 - \eta_t^2 &= Y_t^2\frac{m^2(X_t) - \hat{m}^2(X_t)}{\hat{m}^2(X_t)m^2(X_t)} + 2Y_t\frac{\hat{m}(X_t) - m(X_t)}{\hat{m}(X_t)m(X_t)} \\
&= -\frac{2Y_t\eta_t}{m^2(X_t)}(\hat{m}(X_t) - m(X_t)) + o_P(T^{-1/2}),
\end{aligned}$$

uniformly in t, which follows from (20). Let $v(x,y) = -\frac{2y}{m^2(x)}\left(\frac{y}{m(x)} - 1\right)$. Then,

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}[\hat{\eta}_t^2 - \eta_t^2] &= \int v(x,y)(\hat{m}(x) - m(x))\,d(\hat{F}(x,y) - F(x,y)) \\
&\quad + \int v(x,y)(\hat{m}(x) - m(x))\,dF(x,y) + o_P(T^{-1/2}), \qquad (31)
\end{aligned}$$

where $\hat{F}(x,y) = T^{-1}\sum_{t=1}^{T}I(X_t \le x, Y_t \le y)$. The second term of (31) equals

$$\int v(x,y)f_X^{-1}(x)\frac{1}{T}\sum_{t=1}^{T}K_h(x - X_t)\sigma(X_t)\varepsilon_t\,dF(x,y) + o_P(T^{-1/2})$$

$$= -\frac{2}{Tc^3}\sum_{t=1}^{T}\varepsilon_t + o_P(T^{-1/2}),$$

since $E(v(X_t, Y_t)|X_t) = -2/[c^2 m(X_t)]$ under H_0. The first term of (31) can be written as

$$c_T\int v(x,y)d_T(x)\,d(\hat{F}(x,y) - F(x,y)), \qquad (32)$$

where $c_T \to 0$, and $d_T(x) = c_T^{-1}(\hat{m}(x) - m(x))$. We will show that this term is $o_P(T^{-1/2})$ by making use of techniques from empirical processes. Let $C_1^{1+\alpha}(R_X)$, $\alpha > 0$, be the class of all differentiable functions d defined on the domain R_X of X_t such that $\|d\|_{1+\alpha} \le 1$, where

$$\|d\|_{1+\alpha} = \max\{\sup_x |d(x)|, \sup_x |d'(x)|\} + \sup_{x,x'}\frac{|d'(x) - d'(x')|}{|x - x'|^\alpha}.$$

Note that by (20) and (22), we have that $P(d_T \in C_1^{1+\alpha}(R_X)) \to 1$ as $T \to \infty$, if c_T and $\alpha > 0$ are chosen such that $c_T^{-1}h^{-\alpha} = O(h^{-\delta})$. Next, note that the class

$$\mathcal{F} = \left\{(x,y) \to v(x,y)d(x) : d \in C_1^{1+\alpha}(R_X)\right\}$$

22

is P-Donsker, where P is the joint probability measure of (X_t, Y_t). This is because the bracketing number $N_{[]}(\lambda, C_1^{1+\alpha}(R_X), L_r)$ of the class $C_1^{1+\alpha}(R_X)$ satisfies $(\lambda > 0)$

$$\log N_{[]}(\lambda, C_1^{1+\alpha}(R_X), L_r) \le K\lambda^{-1/(1+\alpha)}$$

[see Corollary 2.7.2 in Van der Vaart and Wellner (1996)], and hence

$$\int_0^\infty \log N_{[]}(\lambda, \mathcal{F}, L_r)\, d\lambda < \infty$$

for $r > 2\beta/(\beta - 1)$. See page 146 in Dedecker and Louhichi (2002) and the proof of Theorem 3 for more details. It now follows that

$$\sup_{d \in C_1^{1+\alpha}(R_X)} \left| \frac{1}{T} \sum_{t=1}^T v(X_t, Y_t)d(X_t) - E\left\{ v(X_t, Y_t)d(X_t) \right\} \right|$$
$$= \sup_{d \in C_1^{1+\alpha}(R_X)} \left| \int v(x,y)d(x)\, d(\hat{F}(x,y) - F(x,y)) \right|$$
$$= O_P(T^{-1/2}),$$

and hence (32) is $O_P(c_T T^{-1/2}) = o_P(T^{-1/2})$. $\qquad\qquad\square$

Acknowledgements. The authors would like to thank Qiwei Yao (London School of Economics) for helpful discussions and for pointing out the book by Dehling, Mikosch and Sørensen (2002). The authors are also grateful to M. Stein, who typed parts of this paper with considerable technical expertise.

References

Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, 28, 549-568.

Bradley, R.C. (1986). Basic properties of strong mixing conditions, In *Dependence in Probability and Statistics* (E. Eberlein and M.S. Taqqu, editors). Birkhäuser, Boston.

Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, London.

Carroll, R.J., Härdle, W. and Mammen, E. (2002). Estimation in an additive model when the components are linked parametrically. *Econometric Theory*, 18, 886-912.

Dedecker, J. and Louhichi, L. (2002). Maximal inequalities and empirical central limit theorems, in *Empirical Process Techniques for Dependent Data* (H. Dehling, T. Mikosch and M. Sørensen, editors). Birkhäuser, Boston.

Dette, H. and Spreckelsen, I. (2004). Some comments on specification tests in nonparametric absolutely regular processes. *J. Time. Ser. Anal.*, 25, 159-172.

Doukhan, P. (1994). *Mixing. Properties and Examples.* Springer, New York.

Eagleson, G.K. and Müller, H.G. (1997). Transformations for smooth regression models with multiplicative errors. *J. Roy. Statist. Soc.-Series B*, 59, 173-189.

Engle, R.F. and Russell, J.R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, 66, 1127-1162.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series. Nonparametric and Parametric Methods.* Springer, New York.

Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Statist.*, 10, 245-271.

Hansen, B.E. (2006). Uniform convergence rates for kernel estimation with dependent data. To appear in *Econometric Theory*.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Infer.*, 68, 295-321.

Linton, O. and Nielsen, J.P. (1995). A kernel method of estimating nonparametric regression based on marginal integration. *Biometrika*, 82, 93-100.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory*, 11, 258-289.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

Nielsen, J.P. and Sperlich, S. (2005). Smooth backfitting in practice. *J. Roy. Statist. Soc.-Series B*, 67, 43-61.

Pardo-Fernández, J.C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statist. Sinica*, 17, 1115-1137.

Rao, C.R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

Wieczorek, G. (2007). *Tests auf multiplikative Struktur in nicht parametrischen Zeitreihenmodellen.* PhD Thesis, Ruhr-Universität Bochum (in German).