Scandinavian Journal of Statistics, Vol. 36: 782–799, 2009
doi: 10.1111/j.1467-9469.2009.00648.x

Goodness-of-Fit Tests for Multiplicative Models with Dependent Data

HOLGER DETTE

Fakultät für Mathematik, Ruhr-Universität Bochum

JUAN CARLOS PARDO-FERNÁNDEZ

Departamento de Estatística e IO, Universidade de Vigo

INGRID VAN KEILEGOM

Institut de Statistique, Université catholique de Louvain

ABSTRACT. Several classical time series models can be written as a regression model between the components of a strictly stationary bivariate process. Some of those models, such as the ARCH models, share the property of proportionality of the regression function and the scale function, which is an interesting feature in econometric and financial models. In this article, we present a procedure to test for this feature in a non-parametric context. The test is based on the difference between two non-parametric 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: bootstrap, dependent data, error distribution, kernel smoothing, location-scale model, mixing sequence, multiplicative model, non-parametric regression

1. Introduction and motivation of the test

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

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

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

This general non-parametric 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, e.g. Fan & Yao, 2003, p. 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 rewrite this 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

This article loaded on 25 September 2009 replaces the version originally published and several critical corrections have been made to the references to formulas in the Appendix.

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.

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 hypothesis $H_1 : m(\cdot) \neq c\sigma(\cdot)$. For simplicity of the exposition, we restrict our attention to the case of $c > 0$; in section 6 we will explain how to consider the case $c < 0$.

The feature stated in the null hypothesis (3) is not exclusive for ARCH models, but it holds for other time series models with a multiplicative structure of the form $Z_t = \sigma_t \epsilon_t$, where $\sigma_t = g(Z_{t-1})$ for some function g. Different choices of the function g lead to different models, such as ARCH and fractionally integrated ARCH (FIARCH) models (see Fan & Yao, 2003), autoregressive conditional duration models in Engle & Russell (1998), or their corresponding non-parametric versions. Therefore, in time series analysis, this hypothesis is a preliminary step to be tested before applying other procedures, such as specific tests for ARCH models.

In econometric and financial models, some relation might be expected between the return (here represented by the regression function, m) and the risk (or the scale function, σ), so the proportionality of m and σ is a feature of interest. For a general motivation about the relationship between the mean and variance functions see, for instance, Engle *et al.* (1987) or Linton (2009).

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 before. For example, McCullagh & Nelder (1989) considered generalized linear models, Carroll & Ruppert (1988) investigated a parametric model with a constant coefficient of variation, while Eagleson & Müller (1997) considered the problem of non-parametric estimation of the regression function in a model where the SD function is proportional to the regression function.

The problem of specification testing for non-parametric 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, for example, Masry & Tjøstheim (1995), Hjellvik *et al.* (1998), Fan & Li (1999) or Dette & 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 weighted empirical processes of the standardized non-parametric residuals calculated under the hypothesis of a multiplicative structure and the alternative of a general non-parametric regression model. Some asymptotic results establishing weak convergence of the (appropriately standardized) difference of the processes and the asymptotic distributions of the test statistics are stated in section 3. The asymptotic distributions of the proposed test statistics are difficult to use in practice. Therefore, in section 4, we describe a consistent bootstrap procedure to approximate the critical values of the test, and in section 5 we present the results of a small simulation study that illustrates the finite sample properties of the bootstrap version of the test. For the sake of simplicity, in this paper we consider a bivariate time series, while extensions to more general models are briefly indicated in section 6. The proofs of the main results are complicated 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 $F_\varepsilon(y) = P(\varepsilon_t \le y)$. 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 $F_{\varepsilon 0}(y) = P(\varepsilon_{t0} \le y)$.

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 theorem 1, the proof of which 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 non-parametric estimators of the regression and variance functions:

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

where $B_t(x, h) = K((x - X_t)h^{-1}) / [\sum_{t'=1}^{T} K((x - X_{t'})h^{-1})]$ 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

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

where w is a weight function with support R_w, which arises from the weighted least squares problem

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

Note that the minimum is attained for

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

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 structurally different estimate can be obtained from the method of moments that yields

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

as an estimate of c^2, where $\bar{w}(\cdot) = w(\cdot) / \sum_{t=1}^{T} w(X_t)$, as $E[w(X_t)(Y_t/m(X_t)-1)^2] = c^{-2}E(w(X_t))$ 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).

In the general non-parametric model (1) the error distribution is estimated by the weighted empirical distribution of the estimated residuals, that is

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

where $I(\cdot)$ denotes the indicator function: $I(t \le y) = 1$ if $t \le y$, and $I(t \le y) = 0$ if $t > y$. 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}, that is,

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

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 Fig. 1, where we show the empirical distribution functions $\hat{F}_{\varepsilon}$ and $\hat{F}_{\varepsilon 0}$ corresponding to the cases: (a) $m(x) = \sigma(x) = 1 + 0.1x$, and (b) $m(x) = 1 + 0.1x, \sigma(x) = 0.5\sqrt{|x|}$. 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) :

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

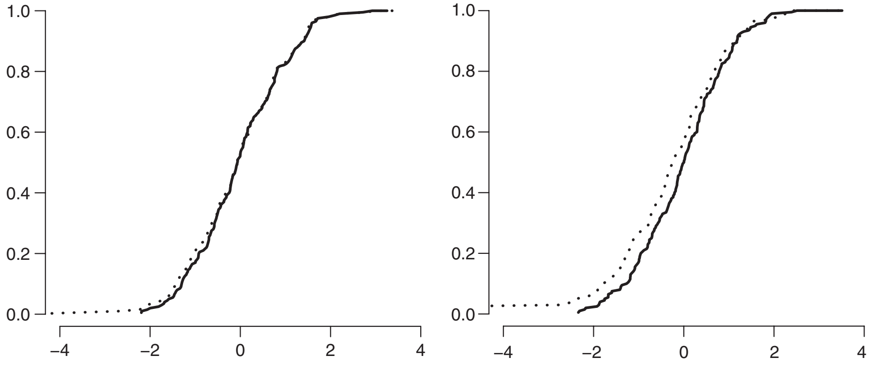

Fig. 1. The empirical processes $\hat{F}_{\varepsilon}$ (solid line) and $\hat{F}_{\varepsilon 0}$ (dotted line) correspond to the testing problem (3). The sample size is $T = 200, m(x) = 1 + 0.1x$. The left panel corresponds to the null hypothesis of a multiplicative model $H_0 : m(x) = c\sigma(x)$,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

The null hypothesis is rejected for large values of the test statistics. In section 3, 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}.

3. Asymptotic results

We assume that the process $(X_t, Y_t), t=0, \pm 1, \pm 2, \ldots,$ is strictly stationary and absolutely regular (β-mixing). This means that the sequence of mixing coefficients

$$\beta_t = E\Big\{ \sup_{A \in \mathcal{F}_t^{\infty}(X,Y)} |P(A) - P(A \mid (X_0, Y_0), (X_{-1}, Y_{-1}), \ldots)| \Big\}$$

converges to 0 as $t \to \infty$, where $\mathcal{F}_t^{\infty}(X, Y)$ is the σ-algebra generated by $\{(X_j, Y_j), j=t, \ldots, \infty\}$. This definition of absolutely regular sequences is taken from Fan & Yao (2003). For a more abstract definition of absolute regular sequences and their properties, see Doukhan (1994).

Let us now introduce some additional 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), and $F_\varepsilon(y)=P(\varepsilon_t \le y)$ the distribution function of the error. 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 mixing coefficients satisfy $\beta_t = O(t^{-b})$, for some $b > 2$.

(A2). X_t is absolutely continuous with density f_X. The functions f_X, m and σ^2 are twice continuously differentiable, $\inf_{x \in R_w} f_X(x) > 0$ and $\inf_{x \in R_w} \sigma^2(x) > 0$. The weight function w satisfies $w(x) \ge 0$ for all x, $\sup_x w(x) < \infty$ and $E(w(X_t)) > 0$.

(A3). (i) $E(|Y_0|^s) < \infty$ and $\sup_{x \in R_w} E(|Y_0|^s \mid X_0 = x) < \infty$ for some $s > 2 + 2/(b-2)$.

(ii) There exists some j' such that for all $j \ge j'$,

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

where $f_j(x_0, x_j)$ denotes the joint density of (X_0, X_j).

(iii) The errors of the regression model satisfy

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

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), and twice continuously differentiable with respect to x and y. Let $L(x,y)$ denote generically the derivatives $\frac{\partial}{\partial x}F(x,y)$, $\frac{\partial}{\partial y}F(x,y)$, $\frac{\partial^2}{\partial x^2}F(x,y)$, $\frac{\partial^2}{\partial y^2}F(x,y)$ or $\frac{\partial^2}{\partial x \partial y}F(x,y)$. Then, $L(x,y)$ is continuous in (x,y) and satisfies $\sup_y |y^2 L(x,y)| < \infty$.

(A5). (i) The bandwidth sequence h_T satisfies the following three conditions:

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

$$(\log h_T^{-1})^{-1} T h_T^{3+\delta} \to \infty \text{ for some } \delta > 0,$$

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

(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

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

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

Note that assumption (A2) implies that the support of the weight function, R_w, is compact. Introducing a weight function w in the empirical processes (6) and (7) has useful consequences in the model, as it allows us to consider covariates with unbounded support, as it usually occurs when the covariate X_t represents lagged variables of the time series Y_t. As an additional remark, note that in the case of independence these assumptions can be relaxed in the following sense: (A1) disappears as the mixing coefficients are zero; in (A3) it suffices to take $s = 2$ as in Akritas & 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 1. 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:*

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

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 centred Gaussian process* $W(y)$ *with covariance structure given by* $\mathrm{cov}(W(y), W(y')) = f_{\varepsilon}(y) f_{\varepsilon}(y') E(w(X_t))^{-2} \sum_{t=1}^{\infty} \mathrm{cov}(w(X_1) W_1, w(X_t) W_t)$.

Corollary 1

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

$$T_{KS} \xrightarrow{d} \sup_y |W(y)| \quad \textit{and} \quad T_{CM} \xrightarrow{d} \int W^2(y)\, \mathrm{d}F_{\varepsilon}(y).$$

To conclude this section, we show that the least squares estimator $\hat{c}_{ls}$ and the moment estimator $\hat{c}_{mom}$, 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,*

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

Theorem 6

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

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

Note that the representations in theorems 5 and 6 have a very similar structure, but there appear additional factors $\sigma^4(X_t)/E[\sigma^4(X_1)]$ 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 6 appears to be simpler, it 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 limit distribution in theorem 3 is degenerate in that case. In order to get the asymptotic distribution of the test statistics when using $\hat{c}_{mom}$, the standardization of the process should be changed and a deeper theoretical analysis carried out. This is, however, beyond the scope of this paper.

In practice, we use a bootstrap calibration to obtain the critical values of the test (see section 4), and we have found in simulations that the least squares estimate $\hat{c}_{ls}$ yields better results. In the simulation study presented in section 5 we only show results for that estimator.

4. Bootstrap calibration

In section 3, we have obtained the asymptotic distribution of the test statistics T_{KS} and T_{CM}. Although these asymptotic limits have explicit and relatively easy formulae, we prefer not to use them in practice, as simulations have shown that the convergence to these limiting distributions is slow, and for small samples the level is not well approximated. To circumvent this problem, in this section we propose to approximate the distribution of the test statistics by means of a smooth bootstrap procedure.

Define

$$Y_t^* = \hat{c}\hat{\sigma}(X_t) + \hat{\sigma}(X_t)\tilde{\tilde{\varepsilon}}_t^*, \quad t = 1, \ldots, T, \tag{9}$$

where

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

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

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

$(t=1,\dots,T), Z_1,\dots,Z_T$ are i.i.d. standard normal random variables, independent of the original sample, and $v=v_n$ is a sufficiently small constant; in our case, we take $v=0.1$. Note that the independence between the error and the covariate allows to consider a regular bootstrap based on the residuals, as all the dependence structure is kept by the regression model. Next, let

$$\hat{F}^*_{\varepsilon 0}(y)=\sum_{t=1}^{T}\bar{w}(X_t)I(\hat{\varepsilon}^*_t\le y),\quad \hat{F}^*_{\varepsilon}(y)=\sum_{t=1}^{T}\bar{w}(X_t)I(\hat{\varepsilon}^*_{t0}\le y),$$

where $\hat{\varepsilon}^*_t$ and $\hat{\varepsilon}^*_{t0}$ are defined similarly as $\hat{\varepsilon}_t$ and $\hat{\varepsilon}_{t0}=(Y_t-\hat{c}\hat{\sigma}(X_t))/\hat{\sigma}(X_t)$, but by using the bootstrap data $\mathcal{Y}_T=\{(X_1,Y^*_1),\dots,(X_T,Y^*_T)\}$ instead of the original data. Now, use these bootstrap empirical distributions to compute the bootstrap version of the test statistics T_{KS} and T_{CM}. Repeat this procedure B times and denote the ordered outcomes by $T^{(1)*}_{CM}\le\cdots\le T^{(B)*}_{CM}$ (and similarly for the statistic T_{KS}). Then, the null hypothesis of a multiplicative model is rejected if

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

where $\lfloor u\rfloor$ denotes the integer part of u.

Note that we use a smooth bootstrap procedure. This is because the asymptotic representation of $\hat{F}_{\varepsilon 0}(y)-\hat{F}_{\varepsilon}(y)$ given in theorem 2 contains the density function $f_{\varepsilon}(y)$, and without using smoothing this would lead to an inconsistent bootstrap procedure; see, for example, Silverman & Young (1987) and Hall *et al.* (1989).

This bootstrap procedure is consistent if, conditionally on the sample $\mathcal{Y}_T$, and under both the null hypothesis H_0 and the fixed alternatives, the process $T^{1/2}(\hat{F}^*_{\varepsilon 0}(y)-\hat{F}^*_{\varepsilon}(y))(-\infty<y<\infty)$ converges weakly to the process $W(y)$, in probability, where $W(y)$ is defined in theorem 3. To prove this, a similar method of proof can be followed as in Neumeyer (2006, 2009) and Dette *et al.* (2007). In the former two manuscripts, it is shown that the smooth bootstrap procedure described before is consistent for approximating the distribution of $\hat{F}_{\varepsilon}(y)$ when the data are i.i.d. In the latter paper, the consistency of the bootstrap is proved for a test for the form of the variance function in regression, that is similar in structure to the test considered in this paper, but that is restricted to i.i.d. data.

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 two multiplicative models including the ARCH(1). Note that by corollary 1, 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 the smooth bootstrap test, outlined in section 4.

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 is used for the estimation of the regression function in order to address 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 functions have been chosen separately by least squares cross-validation, and the same bandwidths have been used in the bootstrap procedure. 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. In a next step, we have generated bootstrap data according to the procedure described in the previous section. In each scenario, 1000 simulation runs with $B=100$ bootstrap replications have been performed to estimate the empirical level of the bootstrap test.

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

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

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

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

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 power. The empirical distribution functions $\hat{F}_\varepsilon$ and $\hat{F}_{\varepsilon 0}$ corresponding to the null hypothesis and alternative have been depicted in Fig. 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

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

with alternative

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

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.

Example 2. We will conclude this section discussing the application of the methodology for testing multiplicative, in particular ARCH structures. For this purpose, we consider two

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 a non-multiplicative model* [*model* (13)]

			$T=50$			$T=100$			$T=200$		
Model	c	α:	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 a non-multiplicative model* [*model* (15)]

			$T=50$			$T=100$			$T=200$		
Model	c	α:	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
(14)	0.5		0.025	0.038	0.079	0.026	0.047	0.086	0.033	0.052	0.097
	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
(15)	0.5		0.232	0.312	0.428	0.356	0.445	0.548	0.593	0.641	0.713
	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

Table 3. *Simulated rejection probabilities of the bootstrap test* (11) *for an ARCH* (1) *structure. Equations* (16) *and* (17) *correspond to the 'null hypothesis' of a multiplicative model while* (18) *and* (19) *correspond to two alternatives*

		$T=50$			$T=100$			$T=200$		
Model	α:	0.025	0.05	0.10	0.025	0.05	0.10	0.025	0.05	0.10
(16)		0.040	0.064	0.132	0.021	0.042	0.091	0.022	0.043	0.085
(17)		0.050	0.071	0.122	0.021	0.042	0.082	0.020	0.043	0.087
(18)		0.367	0.463	0.570	0.453	0.581	0.721	0.678	0.734	0.845
(19)		0.366	0.431	0.604	0.481	0.568	0.701	0.634	0.691	0.783

multiplicative models. 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. Next, we study the level of the bootstrap test if the data is generated by the multiplicative model

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

where the random variables ϵ_t are i.i.d. and standard normally distributed. We also observe a rather precise estimation of the nominal level. In order to address the question if the test is able to detect non-multiplicative structures, we consider two alternatives. First, we consider the classical AR(1) model

$$Z_t = 0.25 + 0.75 Z_{t-1} + 0.5\epsilon_t, \quad t \in \mathbb{Z}, \tag{18}$$

where the random variables ε_t are standard normally distributed. The corresponding results are depicted in the second part of Table 3 and show that the test clearly detects the alternative of a non-multiplicative model. As a further alternative, we have considered the model

$$Z_t = \begin{cases} -0.7 Z_{t-1} + 0.5\varepsilon_t & \text{if } Z_{t-1} > 0.5, \\ 0.7 Z_{t-1} + 0.5\varepsilon_t & \text{if } Z_{t-1} \le 0.5 \end{cases} \quad t \in \mathbb{Z}, \tag{19}$$

(see Fan & Yao, 2003, p. 127) where the random variables ϵ_t are again standard normally distributed. The corresponding results are depicted in the fourth row of Table 3 and this alternative is also detected with reasonable power.

6. Discussion and extensions

In this article, we have presented a method to test for the proportionality of the regression function and the scale function in a non-parametric regression set-up with dependent data, which, to the best of our knowledge, has not been treated in the literature. The proportionality of those two functions is motivated in time series analysis (e.g. ARCH models) and in financial and econometric models. The implementation of a smoothed bootstrap procedure yields a good approximation of the nominal level and reasonable power for finite sample sizes.

We have restricted our attention to the case $m(\cdot)=c\sigma(\cdot)$, with $c>0$. The case $c<0$ is also interesting for practical applications, and it can be treated in a similar way. Note that c is defined in terms of c^2, and the same happens for its estimators. So, to test the multiplicative model with negative c, it suffices to consider the negative square root of c^2, and similarly in the expressions of the estimators. It is important to note that the use of our test requires that the sign of c is specified in advance. We believe that in practice, one will usually have an idea of the sign to be employed. A totally different problem, which we do not treat in this paper, is developing a formal test for the sign of c.

More general null hypotheses could be considered in the future, such as testing for a general relationship of the form

$$H_0: m(\cdot)=g(\sigma(\cdot), c),$$

where g is a specified function and c is a parameter. Note that the null hypothesis (3) corresponds to the choice $g(t,c)=ct$. The estimation of the parameter c under those general null hypotheses would deserve more attention.

Extensions to models with more than one covariate are also 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 non-parametric 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 non-parametrically: backfitting, marginal integration, etc.; see, for example, Hastie & Tibshirani (1990), Linton & Nielsen (1995), Nielsen & Sperlich (2005) among many others. A more delicate issue, which has not been sufficiently addressed in the literature yet, is the appropriate modelling 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 $m(\mathbf{x})=c\sigma(\mathbf{x})=a_0+a_1x_1+\cdots+a_qx_q$, we can consider the ARCH(q) model as a special case of a non-parametric regression model where the regression and SD 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 1 in the Appendix.

Acknowledgements

The authors thank Qiwei Yao (London School of Economics) and Christian Hafner (Institut de Statistique, Université catholique de Louvain) for helpful discussions and for pointing out the book by Dehling, Mikosch and Sørensen (2002)–see Dedecker & Louhichi (2002). This article has also benefitted from the comments of two referees and an associate editor. The authors are also grateful to M. Stein, who typed parts of this paper with considerable technical expertise. H. Dette's research is supported by the Deutsche Forschungsgemeinschaft (SFB 475; Komplexitätsreduktion in multivariaten Datenstrukturen) and the NIH (IR01GM072876). J.C. Pardo-Fernández's research is supported by Spanish Ministerio de Ciencia e Innovación (project MTM2008-03129), Xunta de Galicia (project PGIDIT07PXIB300191PR) and Vicerreitorado de Investigación of the Universidade de Vigo. I. Van Keilegom's research is supported by IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 203650.

References

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

Carroll, R. J. & Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman & Hall, London.

Dedecker, J. & Louhichi, L. (2002). Maximal inequalities and empirical central limit theorems. In *Empirical process techniques for dependent data* (eds H. Dehling, T. Mikosch & M. Sørensen), 137–159. Birkhäuser, Boston.

Dette, H., Neumeyer, N. & Van Keilegom, I. (2007). A new test for the parametric form of the variance function in non-parametric regression. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **69**, 903–917.

Dette, H. & 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. & Müller, H. G. (1997). Transformations for smooth regression models with multiplicative errors. *J. Roy. Statist. Soc. Ser. B* **59**, 173–189.

Engle, R. F., Lilien, D. M. & Robins, R. P. (1987). Estimating time varying risk premia in the term structure: the Arch-M model. *Econometrica* **55**, 391–407.

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

Fan, Y. & 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.

Fan, J. & Yao, Q. (2003). *Nonlinear time series. Nonparametric and parametric methods*. Springer, New York.

Hall, P., DiCiccio, T. J. & Romano, J. P. (1989). On smoothing and the bootstrap. *Ann. Statist.* **17**, 692–704.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized additive models*. Chapman & Hall, London.

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

Linton, O. (2009). Semiparametric and nonparametric ARCH modelling. To appear in *Handbook of financial time series* (eds T. G. Andersen, R. A. Davis, J.-P. Kreiss & Th. Mikosch). Springer.

Linton, O. & Nielsen, J. P. (1995). A kernel method of estimating nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.
Masry, E. & Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* **11**, 258–289.
McCullagh, P. & Nelder, J. (1989). *Generalized linear models*, 2nd edn. Chapman & Hall, London.
Neumeyer, N. (2006). *Bootstrap procedures for empirical processes of nonparametric residuals*. Habilitationsschrift, Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum.
Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. *Scand. J. Statist*. **36**, 204–208.
Nielsen, J. P. & Sperlich, S. (2005). Smooth backfitting in practice. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **67**, 43–61.
Pardo-Fernández, J. C., Van Keilegom, I. & 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.
Silverman, B. W. & Young, G. A. (1987). The bootstrap: to smooth or not to smooth? *Biometrika* **74**, 469–479.
Van der Vaart, A. W. & 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* (in German), Ruhr-Universität Bochum, Bochum.

Received September 2007, in final form February 2009

Juan Carlos Pardo-Fernández, Departamento de Estatística e Investigación Operativa, Universidade de Vigo. E.U. Enxeñaría Técnica Industrial, Rúa Torrecedeira 86, Vigo (36208), Spain.
E-mail: juancp@uvigo.es

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 $\mathrm{var}(\varepsilon_{t0})=\mathrm{var}(\varepsilon_t)=1$. Consider the representation $\varepsilon_{t0}=(Y_t-c\sigma(X_t))/\sigma(X_t)=\varepsilon_t+(m(X_t)/\sigma(X_t)-c)$. By applying expectations on both sides of the previous expression, we obtain $E[m(X_t)/\sigma(X_t)-c]=0$. On the other hand, by calculating variances (taking into account that X_t and ε_t are independent), we obtain $\mathrm{var}(\varepsilon_{t0})=\mathrm{var}(\varepsilon_t)+\mathrm{var}[m(X_t)/\sigma(X_t)-c]$. It follows that $E[m(X_t)/\sigma(X_t)-c]=0$ and $\mathrm{var}[m(X_t)/\sigma(X_t)-c]=0$. 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.

Before writing the proofs of the asymptotic results, we introduce lemma 1.

Lemma 1
Assume that conditions (A1–A6) *are satisfied. Then, the following representation holds:*

$$\hat{F}_\varepsilon(y)-F_\varepsilon(y)=\frac{1}{E(w(X_t))}\frac{1}{T}\sum_{t=1}^{T}w(X_t)\,[I(\varepsilon_t\le y)-F_\varepsilon(y)]$$
$$+\frac{f_\varepsilon(y)}{E(w(X_t))}\int\frac{w(x)}{\sigma(x)}\,[y(\hat{\sigma}(x)-\sigma(x))+\hat{m}(x)-m(x)]f_X(x)\,\mathrm{d}x$$
$$+o_P(T^{-1/2}), \tag{20}$$

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

Proof. The proof is based on theorem 1 in Akritas & 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 covariate X_t has a compact support, $w(X_t) \equiv 1$, 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 this theorem need to be modified. All of the notations used next 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

$$\sup_{x \in R_w} |\hat{m}(x) - m(x)| = O_P((\log T)^{1/2}(Th_T)^{-1/2}), \tag{21}$$

$$\sup_{x \in R_w} |\hat{\sigma}(x) - \sigma(x)| = O_P((\log T)^{1/2}(Th_T)^{-1/2}), \tag{22}$$

and that

$$\begin{aligned}
&\sup_{x \in R_w} |\hat{m}'(x) - m'(x)| = O_P((\log T)^{1/2}(Th_T^3)^{-1/2}),\\
&\sup_{x \in R_w} |\hat{\sigma}'(x) - \sigma'(x)| = O_P((\log T)^{1/2}(Th_T^3)^{-1/2}),\\
&\sup_{x, x' \in R_w} \frac{|\hat{m}'(x) - m'(x) - \hat{m}'(x') + m'(x')|}{|x - x'|^\delta} = O_P((\log T)^{1/2}(Th_T^{3+2\delta})^{-1/2}),\\
&\sup_{x, x' \in R_w} \frac{|\hat{\sigma}'(x) - \sigma'(x) - \hat{\sigma}'(x') + \sigma'(x')|}{|x - x'|^\delta} = O_P((\log T)^{1/2}(Th_T^{3+2\delta})^{-1/2}),
\end{aligned} \tag{23}$$

for some $\delta > 0$. Regarding the validity of (21), this follows from theorem 8 in Hansen (2008). 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 (22) and (23) 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 (2008).

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

$$\int_0^\infty \sqrt{\log N_{[]}(\lambda, \mathcal{F}, \|\cdot\|_{2,\beta})}\, \mathrm{d}\lambda < \infty, \tag{24}$$

where the class $\mathcal{F}$ is now defined as

$$\begin{aligned}
\mathcal{F} = \Big\{ (x, e) \rightarrow &\frac{w(x)}{E(w(X_t))} [I(e \leq y d_2(x) + d_1(x)) - I(e \leq y)] \\
&- \frac{1}{E(w(X_t))} E[w(X_t) F_\varepsilon(y d_2(X_t) + d_1(X_t))] + F_\varepsilon(y) : \\
&- \infty < y < \infty, d_1 \in C_1^{1+\delta}(R_w), d_2 \in \tilde{C}_2^{1+\delta}(R_w) \Big\}.
\end{aligned}$$

Here, $C_1^{1+\delta}(R_w)$ and $\tilde{C}_2^{1+\delta}(R_w)$ are as in AVK, and for any function g,

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

where β^{-1} is the inverse cadlag of the decreasing function $u \rightarrow \beta_{\lfloor u \rfloor}$ ($\lfloor u \rfloor$ being the integer part of u, and β_t being the mixing coefficient) and Q_g is the inverse cadlag of the tail function $u \rightarrow P(|g| > u)$ (see section 4.3 in Dedecker & Louhichi, 2002). Note that we can restrict the functions d_1 and d_2 to R_w, as w lives on R_w.

For verifying (24), let us fix $\lambda > 0$ and let $[d^L, d^U]$, $[\tilde{d}^L, \tilde{d}^U]$, and $[y^L, y^U]$ be λ^2-brackets for a given d_1, d_2, and y, defined in the same way as in AVK, except that we omit the indices i, j, k for notational simplicity. Take $y \geq 0$ (the case $y \leq 0$ can be dealt with in a similar way). For any $0 \leq z \leq M$, where $M = \sup_x w(x)/E(w(X_t)) < \infty$, let

$$p(z) = P\left\{ \frac{w(X_t)}{E(w(X_t))}[I(\varepsilon_t \leq y^U \tilde{d}^U(X_t) + d^U(X_t)) - I(\varepsilon_t \leq y^L \tilde{d}^L(X_t) + d^L(X_t))] > z \right\}.$$

Then,

$$\begin{aligned}
p(z) &\leq p(0) \\
&\leq P\{I(\varepsilon_t \leq y^U \tilde{d}^U(X_t) + d^U(X_t)) - I(\varepsilon_t \leq y^L \tilde{d}^L(X_t) + d^L(X_t)) > 0\} \\
&\leq P(y^L \tilde{d}^L(X_t) + d^L(X_t) \leq \varepsilon_t \leq y^U \tilde{d}^U(X_t) + d^U(X_t)) \\
&= \int P(y^L \tilde{d}^L(x) + d^L(x) \leq \varepsilon_t \leq y^U \tilde{d}^U(x) + d^U(x) \,|\, x) \, \mathrm{d}F_X(x) \\
&\leq \int \{F_\varepsilon(y\tilde{d}^U(x) + d^U(x) \,|\, x) - F_\varepsilon(y\tilde{d}^L(x) + d^L(x) \,|\, x)\} \, \mathrm{d}F_X(x) + K_1\lambda^2 \\
&= \int f_\varepsilon(y\tilde{\xi}(x) + \xi(x) \,|\, x)\{y[\tilde{d}^U(x) - \tilde{d}^L(x)] + [d^U(x) - d^L(x)]\} \, \mathrm{d}F_X(x) + K_1\lambda^2 \\
&\leq K_2\|\tilde{d}^U - \tilde{d}^L\|_1 + K_3\|d^U - d^L\|_1 + K_1\lambda^2 \\
&\leq K_2\|\tilde{d}^U - \tilde{d}^L\|_2 + K_3\|d^U - d^L\|_2 + K_1\lambda^2 \leq (K_1 + K_2 + K_3)\lambda^2,
\end{aligned}$$

for some $\xi(x)$ between $d^L(x)$ and $d^U(x)$, some $\tilde{\xi}(x)$ between $\tilde{d}^L(x)$ and $\tilde{d}^U(x)$, and some $K_1, K_2, K_3 > 0$. It follows that the quantile function $Q(u)$ corresponding to the previous survival function is bounded by

$$Q(u) \leq \begin{cases} M & \text{if } 0 \leq u < p(0), \\ 0 & \text{if } p(0) \leq u \leq 1. \end{cases} \tag{25}$$

Hence,

$$\begin{aligned}
&\left\| \frac{w(X_t)}{E(w(X_t))}[I(\varepsilon_t \leq y^U \tilde{d}^U(X_t) + d^U(X_t)) - I(\varepsilon_t \leq y^L \tilde{d}^L(X_t) + d^L(X_t))] \right\|_{2,\beta}^2 \\
&\leq M^2 \int_0^{p(0)} \beta^{-1}(u) \, \mathrm{d}u \leq M^2 \beta^{-1}(0) p(0) \leq K\lambda^2,
\end{aligned}$$

for some constant $0 < K < \infty$. This shows that

$$\begin{aligned}
&\sup_y \left| T^{-1} \sum_{t=1}^T \frac{w(X_t)}{E(w(X_t))} [I(\hat{\varepsilon}_t \leq y) - I(\varepsilon_t \leq y)] \right. \\
&\left. - \frac{1}{E(w(X_t))} E\left\{ w(X_t) \left[F_\varepsilon\left(\frac{y\hat{\sigma}(X_t) + \hat{m}(X_t) - m(X_t)}{\sigma(X_t)} \right) - F_\varepsilon(y) \right] \right\} \right| = o_P(T^{-1/2}),
\end{aligned} \tag{26}$$

and it is easy to show that (26) remains valid when the factor $E(w(X_t))$ in front of the first two terms is replaced by $T^{-1}\sum_{t=1}^T w(X_t)$, by using the fact that $T^{-1}\sum_{t=1}^T w(X_t) - E(w(X_t)) = O_P(T^{-1/2})$. Now, write

$$\hat{F}_\varepsilon(y) - F_\varepsilon(y) = \sum_{t=1}^{T} \bar{w}(X_t)\{I(\varepsilon_t \le y) - F_\varepsilon(y)\}$$
$$+ \frac{1}{E(w(X_t))} E\left\{ w(X_t)\left[F_\varepsilon\left(\frac{y\hat{\sigma}(X_t) + \hat{m}(X_t) - m(X_t)}{\sigma(X_t)} \right) - F_\varepsilon(y) \right] \right\} + o_P(T^{-1/2})$$
$$= \frac{1}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t)\{I(\varepsilon_t \le y) - F_\varepsilon(y)\}$$
$$+ \frac{f_\varepsilon(y)}{E(w(X_t))} \int w(x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)} f_X(x)\,\mathrm{d}x + o_P(T^{-1/2}),$$

where the last equality follows in a similar way as in the proof of theorem 1 in AVK. This completes the proof.

Proof of theorem 2. Lemma 1 states that

$$\hat{F}_\varepsilon(y) - F_\varepsilon(y) = \frac{1}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t)\left\{ I\left(\frac{Y_t - m(X_t)}{\sigma(X_t)} \le y \right) - F_\varepsilon(y) \right\}$$
$$+ \frac{f_\varepsilon(y)}{E(w(X_t))} \int w(x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{m}(x) - m(x)}{\sigma(x)} f_X(x)\,\mathrm{d}x + o_P(T^{-1/2}) \quad (27)$$

and similarly it can be shown that

$$\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y) = \frac{1}{E(w(X_t))} \frac{1}{T} \sum_{t=1}^{T} w(X_t)\left\{ I\left(\frac{Y_t - c\sigma(X_t)}{\sigma(X_t)} \le y \right) - F_\varepsilon(y) \right\}$$
$$+ \frac{f_\varepsilon(y)}{E(w(X_t))} \int w(x) \frac{y(\hat{\sigma}(x) - \sigma(x)) + \hat{c}\hat{\sigma}(x) - c\sigma(x)}{\sigma(x)} f_X(x)\,\mathrm{d}x + o_P(T^{-1/2}),$$

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, Fan & Yao (2003, theorem 2.20). Now, taking into account that under the null hypothesis $m(x) = c\sigma(x)$, we obtain

$$\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y) = \frac{f_\varepsilon(y)}{E(w(X_t))} \int w(x) \frac{\hat{c}\hat{\sigma}(x) - \hat{m}(x)}{\sigma(x)} f_X(x)\,\mathrm{d}x + o_P(T^{-1/2}).$$

The uniform rates given in (21) ensure that lemmas 8 and 9 in Pardo-Fernández *et al.* (2007) can be applied here:

$$\int w(x) \frac{\hat{m}(x) - m(x)}{\sigma(x)} f_X(x)\,\mathrm{d}x = \frac{1}{T} \sum_{t=1}^{T} w(X_t) \frac{Y_t - m(X_t)}{\sigma(X_t)} + o_P(T^{-1/2})$$

and

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

Hence,

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

which equals the representation given in the statement of the theorem.

Proof of theorem 3. The leading term of the representation given in theorem 2 factorizes in a deterministic function, $f_\varepsilon(y)/E(w(X_t))$, and a sum of random variables, $T^{-1}\sum_{t=1}^{T} w(X_t)W_t$, where $W_t = 0.5c\varepsilon_t^2 - \varepsilon_t - 0.5c + s(X_t, \varepsilon_t)$, not depending on y. The weak convergence of the process $T^{1/2}(\hat{F}_{\varepsilon 0}(y) - \hat{F}_\varepsilon(y))$ follows from the central limit theorem for the α-mixing processes. See, for instance, theorem 2.21 in Fan & Yao (2003). Indeed, the process $w(X_t)W_t$ has expectation 0 and verifies condition (i) of the aforementioned theorem owing to conditions (A1), (A3), and (A6). Note that, as $w(X_t)W_t$ is a transformation of the bidimensional process (X_t, Y_t), it inherits its mixing property [see remark (ii) on p. 69 in Fan & Yao (2003)].

The variance of the limit distribution is: $\text{var}(W_1) + 2\sum_{j=1}^{\infty}\text{cov}(W_{1+j}, W_1)$. Taking into account (A1(ii)), it is easy to check that $\text{cov}(W_{1+j}, W_1) = \text{cov}(s(X_{1+j}, Y_{1+j}), s(X_1, Y_1))$.

Proof of corollary 1. 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, p. 23) implies

$$\sup_y |\hat{W}(y) - W(y)| \to_{\text{a.s.}} 0 \quad \text{and} \quad \sup_y |\hat{F}_\varepsilon(y) - F_\varepsilon(y)| \to_{\text{a.s.}} 0. \tag{28}$$

Now write

$$\begin{aligned}&\left|\int \hat{W}^2(y)\,\mathrm{d}\hat{F}_\varepsilon(y) - \int W^2(y)\,\mathrm{d}F_\varepsilon(y)\right|\\ &\le \left|\int (\hat{W}^2(y) - W^2(y))\mathrm{d}\hat{F}_\varepsilon(y)\right| + \left|\int W^2(y)\mathrm{d}(\hat{F}_\varepsilon(y) - F_\varepsilon(y))\right|.\end{aligned}$$

Both terms on the right-hand side of this inequality are negligible a.s. The first one is $o(1)$ a.s. owing to the first expression in (28). The second one is also $o(1)$ a.s. because of the second expression in (28) and the application of the Helly–Bray theorem (see p. 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 theorems 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

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

Hence, it is sufficient to consider

$$\begin{aligned}\hat{c}_{mom}^{-2} - c^{-2} = &\frac{1}{E(w(X_t))}\frac{1}{T}\sum_{t=1}^{T} w(X_t)[\hat{\eta}_t^2 - \eta_t^2] + \left[\frac{1}{E(w(X_t))}\frac{1}{T}\sum_{t=1}^{T} w(X_t)(\eta_t^2 - c^{-2})\right]\\ &+ \sum_{t=1}^{T}\left[\bar{w}(X_t) - \frac{w(X_t)}{TE(w(X_t))}\right][\hat{\eta}_t^2 - c^{-2}],\end{aligned} \tag{29}$$

where $\eta_t = Y_t/m(X_t) - 1 = c^{-1}\varepsilon_t$ and $\hat{\eta}_t = Y_t/\hat{m}(X_t) - 1$. For the first term here, 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 on R_w, which follows from (21). Let $v(x,y)=-2y(y/m(x)-1)/m^2(x)$. Then,

$$\begin{aligned}&\frac{1}{E(w(X_t))}\frac{1}{T}\sum_{t=1}^{T}w(X_t)\left[\hat{\eta}_t^2-\eta_t^2\right]\\&=\frac{1}{E(w(X_t))}\int w(x)v(x,y)(\hat{m}(x)-m(x))\,\mathrm{d}(\hat{F}(x,y)-F(x,y))\\&\quad+\frac{1}{E(w(X_t))}\int w(x)v(x,y)(\hat{m}(x)-m(x))\,\mathrm{d}F(x,y)+o_P(T^{-1/2}),\end{aligned}\tag{30}$$

where $\hat{F}(x,y)=T^{-1}\sum_{t=1}^{T}I(X_t\le x,\,Y_t\le y)$. The second term of (30) equals

$$\begin{aligned}&\frac{1}{E(w(X_t))}\int w(x)v(x,y)f_X^{-1}(x)\frac{1}{T}\sum_{t=1}^{T}K_h(x-X_t)\sigma(X_t)\varepsilon_t\,\mathrm{d}F(x,y)+o_P(T^{-1/2})\\&=-\frac{1}{E(w(X_t))}\frac{2}{Tc^3}\sum_{t=1}^{T}w(X_t)\varepsilon_t+o_P(T^{-1/2}),\end{aligned}$$

as $E(v(X_t,\,Y_t)\,|\,X_t)=-2/[c^2m(X_t)]$ under H_0. The first term of (30) can be written as:

$$\frac{c_T}{E(w(X_t))}\int w(x)v(x,y)\,d_T(x)d(\hat{F}(x,y)-F(x,y)),\tag{31}$$

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 the techniques from empirical processes. Let $C_1^{1+\alpha}(R_w)$, $\alpha>0$, be the class of all differentiable functions d defined on R_w 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 (21) and (23), we have that $P(d_T\in C_1^{1+\alpha}(R_w))\to 1$ as $T\to\infty$, if c_T and $\alpha>0$ are chosen such that $c_T^{-1}h^{-\alpha}=O(h^{-\delta})$ (we can restrict the function d_T to R_w as $w\equiv 0$ outside R_w). Next, note that the class $\mathcal{F}=\{(x,y)\to w(x)v(x,y)d(x):d\in C_1^{1+\alpha}(R_w)\}$ is P-Donsker, where P is the joint probability measure of $(X_t,\,Y_t)$. This is because the bracketing number $N_{[]}(\lambda,C_1^{1+\alpha}(R_w),L_r)$ of the class $C_1^{1+\alpha}(R_w)$ satisfies ($\lambda>0$)

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

(see corollary 2.7.2 in Van der Vaart & Wellner, 1996), and hence

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

for $r>2b/(b-1)$. See p. 146 in Dedecker & Louhichi (2002) and the proof of theorem 3 for more details. It now follows that

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

and hence (31) is $O_P(c_T T^{-1/2})=o_P(T^{-1/2})$. It now also follows that the third term of (29) is $o_P(T^{-1/2})$.