REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 137–166

COMPUTATIONALLY EFFICIENT GOODNESS-OF-FIT TESTS FOR THE ERROR DISTRIBUTION IN NONPARAMETRIC REGRESSION

Authors: G.I. Rivas-Martínez
– Laboratorio de Sistemas de Potencia y Control,
Universidad Nacional de Asunción, Paraguay
gusyri@hotmail.com

M.D. Jiménez-Gamero
– Dpto. de Estadística e Investigación Operativa,
Universidad de Sevilla, España
dolores@us.es

Received: April 2016 Revised: August 2016 Accepted: August 2016

Abstract:

- Several procedures have been proposed for testing goodness-of-fit to the error distribution in nonparametric regression models. The null distribution of the associated test statistics is usually approximated by means of a parametric bootstrap which, under certain conditions, provides a consistent estimator. This paper considers a goodness-of-fit test whose test statistic is an L_2 norm of the difference between the empirical characteristic function of the residuals and a parametric estimate of the characteristic function in the null hypothesis. It is proposed to approximate the null distribution through a weighted bootstrap which also produces a consistent estimator of the null distribution but, from a computational point of view, is more efficient than the parametric bootstrap.

Key-Words:

- *goodness-of-fit; empirical characteristic function; regression residuals; weighted bootstrap; consistency.*

AMS Subject Classification:

- 62G08, 62G09, 62G10.

1. INTRODUCTION

Let (X,Y) be a bivariate random vector satisfying the general nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where $m(x) = E(Y \mid X = x)$ is the regression function, $\sigma^2(x) = \text{Var}(Y \mid X = x)$ is the conditional variance function and ε is the regression error, which is assumed to be independent of X. Note that, by construction, $E(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = 1$. The covariate X is continuous with density function f_X. The regression function, the variance function, the error distribution and that of the covariate are unknown and no parametric models are assumed for them.

Because the knowledge of the error distribution will improve the statistical analysis of model (1.1), several authors have proposed tests for such distribution, that is, tests of the null hypothesis

$$H_0\colon F \in \mathcal{F},$$

versus the alternative

$$H_1\colon F \notin \mathcal{F},$$

where F stands for the cumulative distribution function (CDF) of ε and $\mathcal{F}$ is a parametric family,

$$\mathcal{F} = \left\{F(\cdot;\theta),\ \ \theta \in \Theta\right\}, \qquad \Theta \subseteq \mathbb{R}^p.$$

Examples are the tests in Neumeyer *et al.* [17] and Heuchenne and Van Keilegom [6], which are based on comparing the empirical CDF of the residuals to a parametric estimator of the CDF under the null hypothesis. Since the equality of the CFDs can be also interpreted in terms of the associated characteristic functions (CFs), Hušková and Meintanis [11] have proposed a test for H_0 that is based on comparing the empirical CF of the residuals to a parametric estimator of the CF under the null hypothesis. As commented in Jiménez-Gamero [13], it is interesting to observe that the last paper requires weaker conditions for the validity of the procedures than the ones based on the CDF. Nevertheless, in all cases the limit distribution of the proposed test statistics is unknown, even under the null distribution, because it depends on the unknown value of the parameter θ. To overcome this difficulty, these papers propose to use a parametric bootstrap (PB) for approximating the null distribution of the test statistic. Although very easy to implement, the PB can become very computationally expensive as the sample size and/or the number of unknown parameters increase.

This paper studies another method for estimating the null distribution of the test statistic $T_{n,w}(\hat{\theta})$ in [11]. Specifically, a weighted bootstrap (WB) approximation in the sense of Burke [2] is considered (see also Zhu [23]). This method has been previously suggested in Kojadinovic and Yan [15], to approximate the null distribution of goodness-of-fit (GOF) tests based on the empirical CDF, and in Jiménez-Gamero and Kim [14], to approximate the null distribution of GOF tests based on the empirical CF (ECF), among others. Both papers assume observable independent and identically distributed (IID) data. They show that the properties of the WB are quite similar to those of the PB (it provides a consistent estimator of the null distribution and the resulting test is able to detect any alternative) but, from a computational point of view, it is more efficient. In view of the good properties of the WB in these and other papers, it is also expected to work satisfactorily for estimating the null distribution of the test statistic considered in this paper. The purpose of the current study is to investigate, both theoretically and empirically, the use of the WB for approximating the null distribution of $T_{n,w}(\hat{\theta})$. A main difference between the setting in this paper and the one in [14, 15] is that in our case the errors are not observable. So we replace the errors by the residuals, but the residuals are not independent.

The paper is organized as follows. Section 2 describes the test statistic and explains some problems with the WB approximation. Section 3 gives a solution to the problems described in the previous section and proves the consistency of the proposed WB approximation. It also shows that the resulting test is consistent, in the sense of being able to detect any alternative. The application of the proposed WB approximation requires the estimation of certain functions appearing in the linear expansion of the parameter estimators. The estimation of such functions is dealt with in Section 4. Section 5 reports the results of some simulation experiments designed to study the finite sample performance of the proposed approximation and to compare it to the PB. From this numerical study it is concluded that both approximations behave quite closely but, from a computational point of view, the WB outperforms the PB. Section 6 concludes and outlines possible extensions of the results presented in this paper. All proofs and technical details are deferred to the last section.

The following notation will be used along the paper: all vectors are column vectors; for any vector a, a_k denotes its k-th coordinate and $\|a\|$ its Euclidean norm; the superscript ^{T} denotes transpose; E_θ and P_θ denote expectation and probability, respectively, assuming that the data has CDF $F(\cdot;\theta)$; P_* denotes the conditional probability law, given the data; all limits in this paper are taken when $n \to \infty$; $\xrightarrow{\mathcal{L}}$ denotes convergence in distribution; $\xrightarrow{P}$ denotes convergence in probability; $\xrightarrow{a.s.}$ denotes the almost sure convergence; for any complex number $z = a + \mathrm{i}b$, $|z|$ is its modulus; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; for a given non-negative real-valued function w we denote $\|\cdot\|_w$ to the norm and $\langle\cdot,\cdot\rangle_w$ to the scalar product in the Hilbert space $L^2(w) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 w(t)dt < \infty\}$; if F is a CDF, then $L^2(F) = \{g : \mathbb{R} \to \mathbb{C}, \int |g(t)|^2 dF(t) < \infty\}$; for any real function $f(t;\theta)$ differentiable at $t \in \mathbb{R}$ and at $\theta = (\theta_1, \theta_2, ..., \theta_p)^T \in \mathbb{R}^p$ the following notations will be used:

$$f'(t;\theta) = \frac{\partial}{\partial t} f(t;\theta)\,, \qquad f_{(r)}(t;\theta) = \frac{\partial}{\partial \theta_r} f(t;\theta), \quad 1 \le r \le p,$$

$$\nabla f(t;\theta) = \Big(f_{(1)}(t;\theta), f_{(2)}(t;\theta), ..., f_{(p)}(t;\theta) \Big)^T.$$

2. THE TEST STATISTIC

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be IID from model (1.1), that is, $Y_j = m(X_j) + \sigma(X_j)\varepsilon_j$, $1 \le j \le n$. Since the hypothesis H_0 is on the common error distribution, $\varepsilon_1, ..., \varepsilon_n$, and the errors are not observable, the inference must be based on the residuals,

$$\hat{\varepsilon}_j = \frac{Y_j - \hat{m}(X_j)}{\hat{\sigma}(X_j)}, \quad 1 \le j \le n,$$

where $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ are estimators of $m(\cdot)$ and $\sigma(\cdot)$, respectively. Several choices are possible for $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$. Here, as in [11], we use the following kernel estimators for the density function f_X of X, the regression function $m(\cdot)$ and the variance function $\sigma^2(\cdot)$,

$$\begin{aligned}
\hat{f}_X(x) &= \frac{1}{n} \sum_{j=1}^{n} K_{h_n}(X_j - x)\,, \\
\hat{m}(x) &= \frac{1}{n \hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x)\, Y_j, \\
\hat{\sigma}^2(x) &= \frac{1}{n \hat{f}_X(x)} \sum_{j=1}^{n} K_{h_n}(X_j - x) \left\{ Y_j - \hat{m}(x) \right\}^2,
\end{aligned}$$

where $K_{h_n}(\cdot) = \frac{1}{h_n} K(\frac{\cdot}{h_n})$, $K(\cdot)$ is a kernel and h_n is the bandwidth, satisfying certain conditions that will be specified later.

Hušková and Meintanis [11] proposed the following test for testing H_0,

$$\Psi = \begin{cases} 1, \text{ if } \quad T_{n,w}(\hat{\theta}) \ge t_{n,\omega,\alpha}, \\ 0, \text{ otherwise,} \end{cases}$$

where $t_{n,\omega,\alpha}$ is the $1 - \alpha$ percentile of the null distribution of $T_{n,\omega}(\hat{\theta})$,

$$T_{n,\omega}(\hat{\theta}) = n \int |c_n(t) - c(t, \hat{\theta})|^2 \omega(t) dt = n \| c_n(t) - c(t, \hat{\theta}) \|_w^2, \tag{2.1}$$

$c_n(t)$ is the ECF of the residuals,

$$c_n(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i} t \hat{\varepsilon}_j) = \frac{1}{n} \sum_{j=1}^{n} \cos(t \hat{\varepsilon}_j) + \mathrm{i} \frac{1}{n} \sum_{j=1}^{n} \sin(t \hat{\varepsilon}_j),$$

$c(t;\theta)$ is the CF associated to $F(\varepsilon;\theta)$, that is, $c(t;\theta) = E_\theta\{\exp(\mathrm{i}t\varepsilon)\} = R(t;\theta) + \mathrm{i}I(t;\theta)$, $\omega(t)$ is a nonnegative function such that $\int \omega(t)dt < \infty$, which may depend on θ, and $\hat{\theta}$ is a consistent estimator of θ satisfying the following assumption.

(**A.1**) Under H_0, $\sqrt{n}(\hat{\theta}-\theta_0) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi(\varepsilon_j;\theta_0) + o_p(1)$, where θ_0 is the true parameter value, $E_{\theta_0}\{\psi(\varepsilon_j;\theta_0)\} = 0$ and $E_{\theta_0}\{\|\psi(\varepsilon_j;\theta_0)\|^2\} < \infty$.

Assumption (A.1) implies that, when the null hypothesis is true and θ_0 denotes the true parameter value, $\sqrt{n}(\hat{\theta}-\theta_0)$ is asymptotically normally distributed. This assumption is satisfied by commonly used estimators such as maximum likelihood estimators and method of moment estimators when $\varepsilon_1,\ldots,\varepsilon_n$ are observable and, in such a case, the expression of the function ψ is well-known (see, for example, [1, Ch. 5]). In our setting, the errors are not observable and the expression of the function ψ differs from the observable case. This topic will be discussed in detail in Section 4.

Theorem 1 in [11] states that if $\hat{\theta}$ satisfies (A.1), H_0 is true and θ_0 is the true parameter value, under certain additional conditions (assumptions (A.2)–(A.7) in Section 7),

$$T_{n,\omega}(\hat{\theta}) \xrightarrow{\mathcal{L}} \|Z(t;\theta_0)\|_\omega^2, \tag{2.2}$$

where $\{Z(t;\theta_0),\, t\in\mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov_{\theta_0}\{Z_1(\varepsilon;t,\theta_0,\psi), Z_1(\varepsilon;s,\theta_0,\psi)\}$,

$$\begin{aligned} Z_1(\varepsilon;t,\theta,\psi) \;=\; & \cos(t\varepsilon)+\sin(t\varepsilon)-R(t;\theta)-I(t;\theta)-t\varepsilon\{R(t;\theta)-I(t;\theta)\} \\ & -t\tfrac{\varepsilon^2-1}{2}\{R'(t;\theta)+I'(t;\theta)\}-\psi^T(\varepsilon;\theta)\{\nabla R(t;\theta)+\nabla I(t;\theta)\}. \end{aligned} \tag{2.3}$$

Clearly, the asymptotic null distribution of $T_{n,\omega}(\hat{\theta})$ is unknown. It depends on the hypothetical the error distribution, on the chosen estimator and the true unknown value of the parameter.

In order to try to approximate the null distribution of $T_{n,\omega}(\hat{\theta})$ we first observe that it resembles a degree-2 V-statistic, because

$$T_{n,\omega}(\hat{\theta}) = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\rho(\hat{\varepsilon}_j,\hat{\varepsilon}_k;\hat{\theta}),$$

with $\rho(\varepsilon,z;\theta) = u(\varepsilon-z) - u_0(\varepsilon;\theta) - u_0(z;\theta) + u_{00}(\theta)$, $u_0(\varepsilon;\theta) = \int u(\varepsilon-z)dF(z;\theta)$, $u_{00}(\theta) = \int u(\varepsilon-z)dF(\varepsilon;\theta)dF(z;\theta)$, and $u(t) = \int \cos(t\varepsilon)\omega(\varepsilon)d\varepsilon$.

Dehling and Mikosch [4] (see also Hušková and Janssen [10]) showed that if $\varepsilon_1,\ldots,\varepsilon_n$ are IID, $\xi_1,\ldots,\xi_n$ are IID with $E(\xi_1) = 0$ and $\mathrm{Var}(\xi_1) = 1$, independent of $\varepsilon_1,\ldots,\varepsilon_n$ and $V_n = \frac{1}{n^2}\sum_{1\le j,k\le n} g(\varepsilon_j,\varepsilon_k)$ is a degenerate degree-2 V-statistic, then the conditional distribution, given $\varepsilon_1, ..., \varepsilon_n$, of

$$\frac{1}{n} \sum_{1 \le j,k \le n} g(\varepsilon_j, \varepsilon_k) \xi_j \xi_k$$

consistently estimates that of nV_n. In the light of this result, since $\hat{\varepsilon}_j$ and $\hat{\theta}$ are approximations to ε_j and θ, respectively, one may be tempted to estimate the null distribution of $T_{n,\omega}(\hat{\theta})$ by means of the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$W^* = \frac{1}{n} \sum_{1 \le j,k \le n} \rho(\hat{\varepsilon}_j, \hat{\varepsilon}_k; \hat{\theta}) \xi_j \xi_k. \tag{2.4}$$

We will see that this approach is wrong. The next result gives the limit distribution of W^*. The required assumptions are listed in Section 7.

Theorem 2.1. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, that assumptions (A.2)–(A.6) hold, that the first partial derivatives $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $1 \le r \le p$, exist and are continuous functions $\forall \theta \in U(\theta_1) \subseteq \Theta$, an open neighborhood of θ_1, and they are bounded by functions in $L_2(\omega)$, $\forall \theta \in U(\theta_1)$, then*

$$\sup_x |P_* \{W^* \le x\} - P\{W_0 \le x\}| \xrightarrow{P} 0,$$

where $W_0 = \|Z_0(t;\theta_1)\|_\omega^2$, $\{Z_0(t;\theta_1),\, t \in \mathbb{R}\}$ is a centered Gaussian process on $L_2(\omega)$ with covariance structure of the form $Cov\{Z_0(\varepsilon; t, \theta_1), Z_0(\varepsilon; s, \theta_1)\}$, $Z_0(\varepsilon; t, \theta) = \cos(t\varepsilon) + \sin(t\varepsilon) - R(t;\theta) - I(t;\theta)$.

From the result in Theorem 2.1 and (2.2), it is clear that the conditional distribution of W^* does not provide a consistent estimator of the null distribution of $T_{n,\omega}(\hat{\theta})$ because replacing $m(\cdot)$, $\sigma(\cdot)$ and θ by $\hat{m}(\cdot)$, $\hat{\sigma}(\cdot)$ and $\hat{\theta}$, respectively, has an impact on the asymptotic null distribution of the test statistic that is not captured by the conditional distribution of W^*. The next Section shows how to deal with this problem.

Before ending this section we do some comments on the behaviour of $\hat{\theta}$ under the alternative. Theorem 2.1 assumes that $\hat{\theta}$ has a limit (in probability), θ_1. In practice, to estimate θ one proceeds as if H_0 were true. For example, θ is usually estimated by its quasi maximum likelihood estimator, which maximizes the likelihood under the null hypothesis (with the errors replaced by the residuals). If H_0 is true, under certain assumptions, the resulting estimator converges to the true parameter value (see Section 4); if H_0 is not true, then proceeding as in White [22] for observable data, it can shown that, under certain conditions, the estimator also converges to a well-defined limit. Similar comments could be done for other estimators.

3. CONSISTENCY OF THE WB APPROXIMATION

If assumptions (A.1)–(A.7) hold and H_0 is true, from the proof of Theorem 1 in [11], it follows that

$$T_{n,\omega}(\hat\theta) = T_{1,n,\omega}(\theta_0) + o_p(1), \tag{3.1}$$

where

$$T_{1,n,\omega}(\theta) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta, \psi)\|_\omega^2,$$

with $Z_1(\varepsilon; t, \theta, \psi)$ as defined in (2.3). Now, from (3.1) and applying the results in [4], we get that the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$T^*_{1,n,\omega}(\theta_0) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\varepsilon_j; t, \theta_0, \psi)\xi_j\|_\omega^2,$$

provides a consistent estimator of the distribution of $T_{n,\omega}(\hat\theta)$, when H_0 is true. From a practical point of view, this result is useless because $Z_1(\varepsilon_j; t, \theta_0, \psi)$ depends on the non-observable error ε_j, on the unknown value of θ_0 and on the function $\psi(\varepsilon_j; \theta_0)$, whose explicit expression is usually unknown. Suppose that $\|\hat\theta - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true. To overcome these difficulties we replace ε_j by $\hat\varepsilon_j$, θ_0 by $\hat\theta$ and $\psi(\varepsilon_j; \theta_0)$ by $\psi_n(\hat\varepsilon_j; \hat\theta)$, where $\psi_n(\cdot; \hat\theta)$ is a function of the data which approximates ψ in such a way that

$$\begin{aligned} &\frac{1}{n}\sum_{j=1}^{n} \|\psi_n(\hat\varepsilon_j; \hat\theta) - \psi_1(\varepsilon_j; \theta_1)\|^2 \xrightarrow{P} 0, \\ &\text{with } \; E\{\|\psi_1(\varepsilon; \theta_1)\|^2\} < \infty \;\text{ and }\; \psi_1(\varepsilon; \theta_1) = \psi(\varepsilon; \theta_1) \;\text{ if } H_0 \text{ is true.} \end{aligned} \tag{3.2}$$

The choice of ψ_n will depend on ψ, that is, on the estimator of θ considered. Section 4 studies some proposals for ψ_n satisfying (3.2) for two common choices for $\hat\theta$: the maximum likelihood estimator and the method of moments estimator, both based on the residuals. So, the null distribution of $T_{n,\omega}(\hat\theta)$ is now estimated by means of the conditional distribution, given $(X_1, Y_1), ..., (X_n, Y_n)$, of

$$T^*_{2,n,\omega}(\hat\theta) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\hat\varepsilon_j; t, \hat\theta, \psi_n)\xi_j\|_\omega^2.$$

The next theorem gives the limit of the conditional distribution of $T^*_{2,n,\omega}(\hat\theta)$, given $(X_1, Y_1), ...(X_n, Y_n)$.

Theorem 3.1. *Suppose that $\|\hat\theta - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, and that assumptions (A.1)–(A.7) and (3.2) hold, then*

$$\sup_x \left| P_*\left\{T^*_{2,n,\omega}(\hat\theta) \le x\right\} - P\left\{T_2 \le x\right\}\right| \xrightarrow{P} 0,$$

where $T_2 = \|Z_2(t;\theta_1)\|_\omega^2$, $\{Z_2(t;\theta_1),\, t \in \mathbb{R}\}$ *is a centered Gaussian process on* $L_2(\omega)$ *with covariance structure of the form* $Cov\{Z_1(\varepsilon;t,\theta_1,\psi_1), Z_1(\varepsilon;s,\theta_1,\psi_1)\}$.

The result in Theorem 3.1 is valid whether the null hypothesis H_0 is true or not. An immediate consequence of this fact and (2.2) is the following.

Corollary 3.1. *If* H_0 *is true and the assumptions in Theorem 3.1 hold, then*

$$\sup_x \left| P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \le x \right\} - P_{\theta_1} \left\{ T_{n,\omega}(\hat{\theta}) \le x \right\} \right| \xrightarrow{P} 0.$$

Let $\alpha \in (0,1)$ and

$$\Psi_* = \begin{cases} 1, \text{ if } \quad T_{n,\omega}(\hat{\theta}) \ge t^*_{2,n,\omega,\alpha}, \\ 0, \text{ otherwise}, \end{cases}$$

where $t^*_{2,n,\omega,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of $T^*_{2,n,\omega}(\hat{\theta})$, or equivalently, $\Psi_* = 1$ if $p^* \le \alpha$, where $p^* = P_* \left\{ T^*_{2,n,\omega}(\hat{\theta}) \ge T_{n,\omega}(\hat{\theta})_{obs} \right\}$ and $T_{n,\omega}(\hat{\theta})_{obs}$ is the observed value of the test statistic. The result in Corollary 3.1 states that Ψ_* is asymptotically correct, in the sense that its type I error is asymptotically equal to the nominal value α.

Corollary 3.2. *Suppose that* H_0 *is not true and let* $c(t)$ *denote the true CF of the errors. If the assumptions in Theorem 3.1 hold and* ω *is such that*

$$\kappa = \|c(t) - c(t;\theta_1)\|_\omega^2 > 0, \tag{3.3}$$

then $P(\Psi_* = 1) \to 1$.

Corollary 3.2 shows that, if ω is such that (3.3) holds, then the test Ψ_* is consistent in the sense of being able to asymptotically detect any (fixed) alternative. Since two distinct characteristic functions can be equal in a finite interval (Feller [5, p.506]), a general way to ensure (3.3) is to take ω positive for almost all (with respect to the Lebesgue measure) points in $\mathbb{R}$.

Remark 3.1. If model (1.1) is homoscedastic, that is, if $\sigma(x) = \sigma$, $\forall x$, for some unknown $\sigma > 0$, we can use the residuals $\tilde{\varepsilon}_j = Y_j - \hat{m}(X_j)$, $1 \le j \le n$, and consider σ as a parameter of the family $\mathcal{F}$. In this framework, the result in Theorem 3.1 (with weaker assumptions) keeps on being true with the following simpler expression for $Z_1(\varepsilon;t,\theta,\psi)$,

$$\begin{aligned} Z_1(\varepsilon;t,\theta,\psi) \;=\; & \cos(t\varepsilon) - R(t;\theta) + \sin(t\varepsilon) - I(t;\theta) - t\varepsilon R(t;\theta) + t\varepsilon I(t;\theta) \\ & - \psi^T(\varepsilon;\theta)\{\nabla R(t;\theta) + \nabla I(t;\theta)\}. \end{aligned}$$

Remark 3.2. If the null hypothesis is simple, then the result in Theorem 3.1 (with weaker assumptions) is also true with the following simpler expression for $Z_1(\varepsilon; t, \theta, \psi) = Z_1(\varepsilon; t)$,

$$\begin{aligned} Z_1(\varepsilon; t) &= \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t) \\ &\quad - t\frac{\varepsilon^2 - 1}{2}\{R'(t) + I'(t)\}, \end{aligned}$$

where $R(t)$ and $I(t)$ denote the real and the imaginary parts of the CF of the law in the null hypothesis.

Remark 3.3. If model (1.1) is homoscedastic and the null hypothesis is simple, which implies that $\sigma(x) = \sigma$, $\forall x$, for some known $\sigma > 0$, as observed in Remark 3.1, we can use the residuals $\tilde{\varepsilon} = Y_j - \hat{m}(X_j)$, $1 \le j \le n$. In this setting, the result in Theorem 3.1 (with weaker assumptions) is also true with the following simpler expression for $Z_1(\varepsilon; t, \theta, \psi) = Z_1(\varepsilon; t)$,

$$Z_1(\varepsilon; t) = \cos(t\varepsilon) - R(t) + \sin(t\varepsilon) - I(t) - t\varepsilon R(t) + t\varepsilon I(t),$$

where $R(t)$ and $I(t)$ denote the real and the imaginary parts of the CF of the law in the null hypothesis.

Remark 3.4. When the null hypothesis is simple, the asymptotic null distribution of the test statistic $T_{n,\omega}(\hat{\theta})$ does not depend on unknown parameters. So, in this case the asymptotic null distribution could be used to approximate the null distribution. The simulations carried out (reported in Section 5) reveal that, for small to moderate sample sizes, the WB provides a better fit.

Remark 3.5. Theorem 3 in [11] shows that the PB null distribution estimator of $T_{n,\omega}(\hat{\theta})$ satisfies a result which is similar to that stated in Corollary 3.1 for the WB estimator. Nevertheless, although the tests Ψ_* and the one obtained by approximating $t_{n,\omega,\alpha}$ through its PB estimator, are both of them consistent against all fixed alternatives, their powers will be different for finite sample sizes.

So far we have assumed that the weight function does not depend on θ, but in some cases it does. Such dependence is motivated by the recommendations in Epps and Pulley [8], who suggest to choose $\omega(t)$ giving high weight where the ECF is a relatively precise estimator of the population CF. It entails taking $\omega(t) = \nu\{|c(t; \hat{\theta})|\}$, for some ν, a nonnegative increasing function. For example, if $\int |c(t;\theta)|^2 dt < \infty$, one could choose $\omega(t) = |c(t; \hat{\theta})|^2 / \int |c(x; \hat{\theta})|^2 dx$, which is the choice for ω in Epps and Pulley [8] (see also Epps [7]). In addition, as observed in Jiménez-Gamero *et al.* [12], such choice for $\omega(t)$ may have some computational advantages when the density (under the null hypothesis) of $\varepsilon_1 - \varepsilon_2$, $\varepsilon_1 - \varepsilon_2 + \varepsilon_3$ and $\varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4$ is known since from expression (14) in [12], the test statistic (2.1) can be expressed as

$$\frac{1}{f_{\varepsilon_1-\varepsilon_2}(0;\hat{\theta})}\left\{\frac{1}{n}\sum_{j,k=1}^{n} f_{\varepsilon_1-\varepsilon_2}(\hat{\varepsilon}_j - \hat{\varepsilon}_k;\hat{\theta}) - 2\sum_{j=1}^{n} f_{\varepsilon_1-\varepsilon_2+\varepsilon_3}(\hat{\varepsilon}_j;\hat{\theta}) + n f_{\varepsilon_1-\varepsilon_2+\varepsilon_3-\varepsilon_4}(0;\hat{\theta})\right\},$$

where $f_U(x;\theta)$ is the density function of U.

If the weight function ω depends on θ, $\omega(t) = \omega(t;\theta)$, then the test statistic (2.1) becomes

$$T_{n,\hat{\omega}}(\hat{\theta}) = n\int |c_n(t) - c(t;\hat{\theta})|^2 \omega(t;\hat{\theta})dt = n\|c_n(t) - c(t;\hat{\theta})\|^2_{\hat{\omega}},$$

where the subindex $\hat{\omega}$ means that the weight function depends on $\hat{\theta}$, that is, $\omega(t) = \omega(t;\hat{\theta})$. To deal with this case we will assume that the weight function is smooth as a function of θ, as expressed in the next assumption.

(**A.8**) $|\omega(t;\theta_1) - \omega(t;\theta)| \leq \omega_0(t;\theta_1)\|\theta - \theta_1\|$, $\forall\,\theta$ in an open neighborhood of θ_1, with $\omega_0(t;\theta_1)$ satisfying $\int \omega_0(t;\theta_1)dt < \infty$.

If assumption (A.8) holds, assumptions (A.2), (A.7) hold with $\omega(t) = \omega_0(t;\theta)$ and H_0 is true, then

$$T_{n,\hat{\omega}}(\hat{\theta}) = T^1_{n,\omega}(\hat{\theta}) + o_p(1),$$

with $T^1_{n,\omega}(\hat{\theta}) = n\int |c_n(t) - c(t,\hat{\theta})|^2\omega(t;\theta_1)dt$.

Let $T^*_{3,n,\omega}(\hat{\theta}) = \|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_1(\hat{\varepsilon}_j; t, \hat{\theta}, \psi_n)\xi_j\|^2_{\hat{\omega}}$ and

$$\Psi_{1*} = \begin{cases} 1, \text{ if } \quad T_{n,\hat{\omega}}(\hat{\theta}) \geq t^*_{3,n,\omega,\alpha}, \\ 0, \text{ otherwise,} \end{cases}$$

where $t^*_{3,n,\omega,\alpha}$ is the $1-\alpha$ percentile of the conditional distribution of $T^*_{3,n,\omega}(\hat{\theta})$. Now, proceeding as in the case where ω does not depend on the parameter θ, we state the following result.

Theorem 3.2. *Suppose that* $\|\hat{\theta} - \theta_1\| = o_p(1)$, *for some* $\theta_1 \in \Theta$, θ_1 *being the true parameter value if* H_0 *is true, that assumptions (A.1)–(A.8) and* (3.2) *hold, where both (A.2) and (A.7) hold with* $\omega(t) = \omega_0(t;\theta_1)$ *and* $\omega(t) = \omega(t;\theta_1)$.

(**a**) *If* H_0 *is true, then*

$$\sup_x \left| P_*\left\{T^*_{3,n,\omega}(\hat{\theta}) \leq x\right\} - P_{\theta_1}\left\{T_{n,\hat{\omega}}(\hat{\theta}) \leq x\right\}\right| \xrightarrow{P} 0.$$

(**b**) *If* H_0 *is not true and* (3.3) *holds with* $\omega(t) = \omega(t;\theta_1)$, *then* $P(\Psi_{1*} = 1) \to 1$.

The observation in Remark 3.1 also applies in this case.

Remark 3.6. The results stated up to now keep on being true if instead of using the raw multipliers, $\xi_1, ..., \xi_n$, we use the centered multipliers, $\xi_1 - \bar{\xi}, ..., \xi_n - \bar{\xi}$, as suggested in [2, 15], where $\bar{\xi} = \frac{1}{n}\sum_{j=1}^{n}\xi_j$.

Remark 3.7. In practice, to calculate the WB approximation to the null distribution of $T_{n,\omega}(\hat{\theta})$ (analogously for $T_{n,\hat{\omega}}(\hat{\theta})$) we proceed as follows:

1. Calculate the residuals $\hat{\varepsilon}_1, ..., \hat{\varepsilon}_n$ (or $\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_n$, if the model is homoscedastic).
2. Calculate $\hat{\theta}$ and the observed value of the test statistic $T_{n,\omega}(\hat{\theta})_{obs}$.
3. Calculate $m_{jk} = \langle Z_1(\hat{\varepsilon}_j; t, \hat{\theta}, \psi_n),\, Z_1(\hat{\varepsilon}_k; t, \hat{\theta}, \psi_n)\rangle_\omega$, $1 \le j \le k \le n$, and take $m_{jk} = m_{kj}$.
4. For some large integer B, repeat the following steps for every $b \in \{1, ..., B\}$:
   (a) Generate n IID variables $\xi_1, ..., \xi_n$ with mean 0 and variance 1.
   (b) Calculate $T^{*b}_{2,n,w}(\hat{\theta}) = \frac{1}{n}\sum_{j,k}\xi_j\xi_k m_{jk}$ (or $T^{*b}_{2,n,\omega}(\hat{\theta}) = \frac{1}{n}\sum_{j,k}(\xi_j - \bar{\xi}) \cdot (\xi_k - \bar{\xi}) m_{jk}$, as noted in Remark 3.6).
5. Approximate the p-value by $\hat{p} = \frac{1}{B}\sum_{b=1}^{B} I\{T^{*b}_{2,n,\omega}(\hat{\theta}) > T_{n,\omega}(\hat{\theta})_{obs}\}$.

4. PARAMETER ESTIMATORS

The maximum likelihood estimator (MLE) satisfies Assumption (A.1) for observable random variables. In our case, the errors are not observable. It seems reasonable to replace the errors by the residuals in the likelihood and then maximize in θ the resulting function. Specifically, assume that the CDF $F(x;\theta)$ has a Radon–Nikodym derivative $f(x;\theta)$ with respect to some σ-finite measure over $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the class of Borel sets of $\mathbb{R}$. To estimate θ we treat the residuals as it they were the true errors and consider

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta}\sum_{j=1}^{n}\log f(\hat{\varepsilon}_j;\theta).$$

Theorem 3.1 in Heuchenne and Van Keilegom [6] shows that (under certain conditions) $\hat{\theta}_{ML}$ satisfies (A.1) with $\psi(\varepsilon;\theta) = \psi_{ML}(\varepsilon;\theta)$ given by

$$\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta) + \frac{\varepsilon^2 - 1}{2}\rho_2(\theta), \tag{4.1}$$

where $\rho_1(\theta) = E_\theta\{\rho'(\varepsilon;\theta)\}$, $\rho_2(\theta) = E_\theta\{\varepsilon\rho'(\varepsilon;\theta)\}$, $\rho(\varepsilon;\theta) = -A(\theta)^{-1}\nabla \log f(\varepsilon;\theta)$, $A(\theta) = (A_{rs}(\theta))$ and

$$A_{rs}(\theta) = E_\theta\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right), \quad 1 \le s,r \le p.$$

In view of (4.1), a natural choice for $\psi_n(\varepsilon;\theta)$ is $\psi_n(\varepsilon;\theta) = \psi_{n,ML}(\varepsilon;\theta)$ with

$$\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta) + \frac{\varepsilon^2-1}{2}\hat{\rho}_2(\theta),$$

where

$$\begin{aligned}
\rho_n(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\nabla \log f(\varepsilon;\theta),\\
\hat{\rho}_1(\theta) &= \frac{1}{n}\sum_{j=1}^n \rho_n'(\hat{\varepsilon}_j;\theta),\\
\hat{\rho}_2(\theta) &= \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j\rho_n'(\hat{\varepsilon}_j;\theta),\\
\rho_n'(\varepsilon;\theta) &= -\hat{A}_n(\theta)^{-1}\frac{\partial}{\partial\varepsilon}\nabla \log f(\varepsilon;\theta),\\
\hat{A}_n(\theta) &= (\hat{A}_{n,rs}(\theta)),\\
\hat{A}_{n,rs}(\theta) &= \frac{1}{n}\sum_{j=1}^n \frac{\partial}{\partial\theta_r}\log f(\hat{\varepsilon}_j;\theta)\frac{\partial}{\partial\theta_s}\log f(\hat{\varepsilon}_j;\theta), \quad 1 \le s,r \le p.
\end{aligned}$$

The next theorem shows that $\psi_{n,ML}(\varepsilon;\theta)$ satisfies (3.2). Let $A_F(\theta) = (A_{F,rs}(\theta))$, with $A_{F,rs}(\theta) = E\left(\frac{\partial}{\partial\theta_r}\log f(\varepsilon;\theta)\,\frac{\partial}{\partial\theta_s}\log f(\varepsilon;\theta)\right)$, $1 \le s,r \le p$, $\rho_{1,F}(\theta) = E\{\rho_F'(\varepsilon;\theta)\}$, $\rho_{2,F}(\theta) = E\{\varepsilon\rho_F'(\varepsilon;\theta)\}$ and $\rho_F(\varepsilon;\theta) = -A_F(\theta)^{-1}\nabla \log f(\varepsilon;\theta)$.

Theorem 4.1. *Suppose that $\|\hat{\theta} - \theta_1\| = o_p(1)$, for some $\theta_1 \in \Theta$, θ_1 being the true parameter value if H_0 is true, and that assumptions (A.3)–(A.6), (A.9) hold, then $\psi_{n,ML}(\varepsilon;\theta)$ satisfies*

$$\frac{1}{n}\sum_{j=1}^n \|\psi_{n,ML}(\hat{\varepsilon}_j;\hat{\theta}) - \psi_1(\varepsilon_j;\theta_1)\|^2 \xrightarrow{P} 0,$$

with $\psi_1(\varepsilon;\theta) = \rho_F(\varepsilon;\theta) + \varepsilon\rho_{1,F}(\theta) + \frac{\varepsilon^2-1}{2}\rho_{2,F}(\theta)$.

Clearly, $\psi_1(\varepsilon_j;\theta)$ in Theorem 4.1 satisfies $\psi_1(\varepsilon_j;\theta_1) = \psi_{ML}(\varepsilon;\theta_1)$ when H_0 is true.

Remark 4.1. If model (1.1) is homoscedastic then the expressions for $\psi_{ML}(\varepsilon;\theta)$ and $\psi_{n,ML}(\varepsilon;\theta)$ simplify to $\psi_{ML}(\varepsilon;\theta) = \rho(\varepsilon;\theta) + \varepsilon\rho_1(\theta)$ and $\psi_{n,ML}(\varepsilon;\theta) = \rho_n(\varepsilon;\theta) + \varepsilon\hat{\rho}_1(\theta)$, respectively.

Another estimator that is commonly used is the method of moment estimator (MME). Although these estimators are not usually optimal, they are frequently used because their calculation is less time consuming than that of MLEs. MMEs satisfy Assumption (A.1) for observable random variables. As noticed before, in our setting the errors are not observable. Next, we study if (A.1) still holds when the errors are replaced by the residuals. Assume that, under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1, ..., g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \le r \le p$, $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$ and $\mu_{0,s} = E_{\theta_0}(\varepsilon^s)$, $\forall s$. The first moment has not been included because, by construction, it is known and equal to 0. In heteroscedastic models the second order moment is also known (thus in this case $\mu_0 = (\mu_{0,3}, ..., \mu_{0,k})^T$), but it is not in homoscedastic models (thus in this case $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$). Nevertheless, we will work with $\mu_0 = (\mu_{0,2}, ..., \mu_{0,k})^T$, by implicitly understanding that in heteroscedastic models $g(\mu_{0,2}, ..., \mu_{0,k}) = g(\mu_{0,3}, ..., \mu_{0,k})$. Let $\hat{\theta}_{MM} = g(\hat{\mu})$, with $\hat{\mu} = (\hat{\mu}_2, ..., \hat{\mu}_k)^T$, $\hat{\mu}_s = \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j^s$, $\forall s$. The next theorem states that, under certain conditions, assumption (A.1) holds for $\hat{\theta}_{MM}$. Let $\nabla g_r(x) = \left(\frac{\partial}{\partial x_2} g_r(x), ..., \frac{\partial}{\partial x_k} g_r(x)\right)^T$, $1 \le r \le p$, and let $\nabla g(x)$ be the $p \times (k-1)$-matrix with rows $\nabla g_1(x)^T, ..., \nabla g_p(x)^T$, for any $x = (x_2, ..., x_k)^T \in \mathbb{R}^{k-1}$.

Theorem 4.2. *Suppose that assumptions (A.3)–(A.6) hold, that g is continuously differentiable at μ_0, that $\mu_{0,2k} < \infty$ and that H_0 is true, then*

$$\sqrt{n}(\hat{\theta}_{MM} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \psi_{MM}(\varepsilon_j; \mu_0) + o_p(1),$$

where $\psi_{MM}(\varepsilon; \mu_0) = \nabla g(\mu_0) v$, $v = (v_2, ..., v_k)^T$, $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon - \mu_{0,s}\frac{\varepsilon^2 - 1}{2}$, $2 \le s \le k$.

In the light of the result in Theorem 4.2, to approximate $\psi_{MM}(\varepsilon; \mu)$ we could replace the population moments by their empirical counterparts based on the residuals. The next theorem shows that this approximation for $\psi_{MM}(\varepsilon; \theta)$ satisfies (3.2). Let $\mu_{F,s} = E(\varepsilon^s)$ and $\mu_F = (\mu_{F,2}, ..., \mu_{F,k})^T$.

Theorem 4.3. *Suppose that assumptions (A.3)–(A.6), (A.10) hold and that $\mu_{F,2k} < \infty$, then*

$$\frac{1}{n} \sum_{j=1}^n \|\psi_{MM}(\hat{\varepsilon}_j; \hat{\mu}) - \psi_{MM}(\varepsilon_j; \mu_F)\|^2 \xrightarrow{P} 0.$$

Clearly, $\psi_{MM}(\varepsilon_j; \mu_F) = \psi_{MM}(\varepsilon_j; \mu_0)$ when H_0 is true.

Remark 4.2. If model (1.1) is homoscedastic then the expressions for $\psi_{MM}(\varepsilon; \mu)$ simplifies to $\psi_{MM}(\varepsilon; \mu_0) = \nabla g(\mu_0) v$, $v = (v_2, ..., v_k)^T$, $v_s = \varepsilon^s - \mu_{0,s} - \mu_{0,s-1}\varepsilon$, $2 \le s \le k$.

5. FINITE SAMPLE PERFORMANCE

With the aim of studying the finite sample performance of the proposed procedure, two simulation experiments were carried out: first, a homoscedastic regression model was considered, and then a heteroscedastic regression model. The main goal of these experiments is to compare the approximations provided by the asymptotic null distribution (when the null hypothesis is simple), the PB (as described in [11]) and the WB proposed in this paper, in three senses: closeness of the approximation under the null, the power for fixed alternatives of the resulting test and the consumed time (for the PB and the WB). This section reports and summarizes the numerical results obtained. All computations were performed using programs written in the R language [20].

In both models the hypotheses $H_0 : \varepsilon \sim N(0, \theta)$, that corresponds to testing that the error distribution is normal with CF $\exp(-0.5\theta t^2)$, and $H_0 : \varepsilon \sim \mathcal{L}(0, \theta)$, that corresponds to testing that the error distribution is Laplace with CF $\frac{1}{1+\theta t^2}$, were studied. As in Hušková and Meintanis [11], and following the recommendations in Epps and Pulley [8], the weight functions considered were: $\omega(t; \theta) = \exp(-\lambda\theta t^2)$, when testing normality, and $\omega(t; \theta) = (1 + \theta t^2)^4 \exp(-\lambda t^2)$, when testing for the Laplace distribution. For the homoscedastic model two cases were considered: θ known and θ unknown. In this second case, the parameter was estimated by a MME. Specifically, $\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} \hat{\varepsilon}_j^2$, for testing normality, and $\hat{\theta} = \frac{1}{2n} \sum_{j=1}^{n} \hat{\varepsilon}_j^2$, for the Laplace distribution. To estimate the regression function and the conditional variance, the Epanechnikov kernel $K(u) = 0.75 \times (1 - u^2)$ was employed.

As for the choice of the bandwidth, in a recent review about GOF problems in nonparametric regression, González-Manteiga and Crujeiras [9] say that the bandwidth selection for tests based on smoothing is a "really tough problem" and "it is far from being solved" (see also the discussions of Sperlich [21] and de Uña-Álvarez [3] to the mentioned article). Because of this reason, to choose h, we proceeded as in the simulation study in Pardo-Fernández *et al.* [18]: we took $h = c \times n^a$, where c and a are real constants and n is the sample size; to determine c, a and λ some preliminary simulations were performed with the purpose of finding values giving type I error close to the nominal. For all tried combinations of $c \in (1, 1.8)$, $a \in (-0.50, -0.25)$ and $\lambda \in (0.03, 0.54)$ good results were obtained for the WB. Here we only report the results for $c = 1.2$, $a = -0.375$ and $\lambda = 0.04$.

The error distribution were generated from: the normal distribution (denoted as N in the tables), the Laplace distribution (denoted as LP), the logistic distribution (denoted as LG), the Gumbel distribution (denoted as G), the beta distribution with parameters $a = 1$ and $b = 0.5$ (denoted as β), the chi-squared distribution with 3 degrees of freedom (denoted as χ^2_3) and the Student t distribution with 5 degrees of freedom (denoted as t_5). All aforementioned distributions were conveniently centered and scaled to have mean 0 and variance 1.

To approximate the p-value, 1000 replications were generated for both the PB and the WB. For the WB, the raw multipliers and the centered multipliers were considered, denoted by WB1 and WB2 in the tables, respectively. The multipliers were generated from a univariate standard normal distribution. As for the asymptotic distribution (when the null hypothesis is simple, denoted as A in the tables), it is rather difficult to calculate because it coincides with that of $\sum_{j\geq 1}\lambda_j\chi^2_{1,j}$, where $\chi^2_{1,1}, \chi^2_{1,2}, \ldots$ are independent chi-squared variables with one degree of freedom, the set $\{\lambda_j,\, j\geq 1\}$ are the non-null eigenvalues of the integral equation $\int C(t,s)G_j(t)dt = \lambda_j G_j(s)$, with corresponding eigenfunctions $\{G_j(\cdot),\, j\geq 1\}$, $C(t,s)$ is the covariance kernel of $Z_1(\varepsilon;t)$ (see Remarks 3.2 and 3.3 for the expression of $Z_1(\varepsilon;t)$), and determining the eigenvalues of an integral equation is tricky. Because of this reason, we approximated it by generating 10,000 samples of size 1000 obeying H_0 and calculated the test statistic at each sample, obtaining 10,000 values. The empirical CDF of these 10,000 values was taken as an approximation to the asymptotic null distribution.

1000 samples with size $n = 25$ were generated from each distribution and the fractions of p-values less than or equal to 0.05 and 0.1 were calculated. The experiment was repeated for $n = 50$, 100.

5.1. Homoscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + \varepsilon_j, \quad 1\leq j\leq n,$$

where X_j follows the uniform $(0,1)$ distribution. We first considered that θ is known. Since the model is homoscedastic and the null hypothesis is simple, the simplifications in Remark 3.3 can be applied. Table 1 displays the results obtained for the type I error and the power for testing normality and Table 2 for testing GOF to the Laplace distribution. Looking at these tables it can be concluded that, in terms of type I error, both the PB and the WB behave very close to the nominal levels, while the asymptotic approximation is a bit conservative, specially for testing GOF for the Laplace distribution. As for the power, the test based on the WB approximation seems to be a bit more powerful than one based on the PB. In most cases (all but alternatives β and χ^2_3 in Table 2) the WB approximation is also more powerful than one based on the asymptotic approximation.

Tables 3 and 4 show the results when θ is assumed to be unknown. In this case, the simplifications in Remark 3.1 can be applied. Looking at these tables it can be concluded that, in terms of the type I error, as before, both the PB and the WB behave very close to the nominal levels. As for the power, for $n = 25, 50$ in some cases the WB is more powerful than the PB, but in others cases the opposite is observed; for $n = 100$ the test based on the WB approximation seems to be a bit more powerful than one based on the PB.

Table 1: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.60	6.10	4.10	6.40	4.00	5.00	4.12	4.84	5.20	4.74	4.12	4.74
	8.20	11.50	10.20	12.30	9.00	10.04	9.24	10.48	10.20	9.64	9.34	10.40
LP	25.50	36.10	57.80	64.80	45.40	56.30	86.60	88.30	76.60	77.70	98.90	99.00
	35.90	48.70	70.60	74.70	57.40	68.50	90.40	91.00	83.60	83.20	99.60	99.70
LG	10.30	57.60	56.40	63.10	12.70	88.10	87.30	89.30	17.80	99.90	100.00	100.00
	18.10	70.40	72.00	76.00	20.60	93.20	94.30	95.10	27.80	99.90	100.00	100.00
G	18.40	33.50	45.80	52.00	36.70	61.80	87.30	89.30	71.70	90.70	100.00	100.00
	30.60	46.30	62.80	67.70	49.80	74.40	94.30	95.10	81.80	96.70	100.00	100.00
β	54.10	37.50	76.20	83.10	87.50	61.20	98.40	99.00	99.70	85.30	100.00	100.00
	65.20	49.40	87.60	89.60	92.70	69.90	99.60	98.80	99.90	90.70	100.00	100.00
χ^2_3	48.60	44.20	76.50	82.50	84.60	73.40	98.40	98.60	99.90	94.50	100.00	100.00
	61.30	57.30	87.80	89.60	92.70	83.10	99.10	99.30	99.90	97.00	100.00	100.00
t_5	15.50	44.50	49.10	55.00	24.50	74.00	87.30	89.20	39.30	97.50	99.90	99.90
	25.00	59.50	63.00	67.90	35.40	84.70	93.70	94.90	51.10	99.50	100.00	100.00

Table 2: (Homoscedastic model, simple null hypothesis) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	3.70	22.60	17.20	19.10	4.20	42.60	38.10	39.30	8.30	69.70	68.60	69.20
	8.40	30.90	25.00	27.20	9.10	51.80	48.10	50.20	14.60	77.60	78.10	78.20
LP	2.70	4.70	3.60	4.20	3.80	4.80	3.80	3.80	3.90	5.50	4.40	4.50
	7.30	9.40	7.70	8.90	8.20	10.60	8.00	9.20	8.90	9.20	9.00	9.10
LG	4.20	25.60	18.90	20.60	4.70	40.60	36.90	37.50	5.90	69.90	70.00	70.70
	7.50	35.00	28.50	31.20	9.30	48.80	46.60	47.50	11.90	78.30	78.30	79.00
G	6.00	23.30	17.70	18.80	11.60	41.60	36.60	38.20	27.10	67.10	68.20	68.90
	10.90	31.70	25.90	28.20	20.70	50.60	47.50	48.60	40.40	77.10	77.10	77.80
β	35.50	12.80	13.40	15.30	78.60	19.30	30.90	32.60	99.30	36.20	66.00	66.60
	48.80	19.20	21.80	24.30	86.20	27.70	43.20	44.50	99.60	46.60	74.00	75.60
χ^2_3	17.50	20.00	16.00	17.80	44.40	34.00	32.60	33.80	92.30	61.50	65.60	66.50
	27.20	27.60	24.00	25.70	59.10	44.00	43.70	44.60	96.90	72.00	76.20	76.70
t_5	3.20	21.50	16.20	18.10	5.40	39.30	35.00	36.70	8.70	71.60	70.80	71.40
	8.00	31.10	24.60	27.10	10.00	49.60	45.90	48.00	14.10	79.80	80.10	80.70

Table 3: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the normality null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$			$n = 50$			$n = 100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	6.50	5.60	7.30	5.20	4.80	5.60	5.40	5.10	5.20
	10.70	10.90	14.50	10.00	9.90	11.10	9.20	9.20	9.60
LP	29.90	15.30	21.50	33.60	40.40	43.50	38.30	80.50	81.60
	40.50	26.20	30.60	44.10	56.30	58.70	54.70	90.40	91.00
LG	30.30	44.10	50.50	47.80	86.60	89.00	94.90	99.90	99.90
	40.30	60.60	65.80	63.90	93.80	94.60	98.50	99.99	99.99
G	29.10	18.50	21.50	35.70	42.50	43.50	51.80	80.50	83.60
	43.50	29.20	30.60	51.30	58.30	59.70	66.10	90.40	95.90
β	18.00	16.40	20.40	23.30	39.40	42.70	67.30	80.80	82.10
	25.40	27.10	30.50	32.80	55.30	56.60	72.10	89.70	91.60
χ^2_3	37.30	51.40	53.80	58.90	77.30	80.70	83.10	89.90	91.30
	48.50	63.20	64.20	67.80	85.40	87.20	91.50	97.70	98.80
t_5	40.40	14.50	21.50	52.90	38.90	42.40	76.80	82.30	83.10
	58.70	28.70	31.40	69.20	53.70	56.00	88.20	89.50	90.30

Table 4: (Homoscedastic model, composite null hypothesis) Percentage of rejections for the Laplace null hypothesis at significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$			$n = 50$			$n = 100$		
	PB	WB1	WB2	PB	WB1	WB2	PB	WB1	WB2
N	53.20	56.90	58.80	62.80	64.40	66.20	69.30	71.40	77.20
	66.30	68.20	71.10	74.50	75.40	76.60	80.60	80.90	81.20
LP	4.30	3.80	4.50	4.60	4.60	4.40	5.00	4.70	4.90
	9.20	8.30	9.20	10.30	9.30	10.40	9.50	9.80	9.50
LG	52.40	48.20	50.50	60.40	58.50	60.20	74.60	77.50	78.50
	65.30	62.00	65.70	72.10	71.70	73.90	90.80	93.20	93.70
G	52.20	47.20	50.30	50.40	51.10	58.70	63.80	65.50	66.90
	64.30	60.70	64.60	62.20	61.50	73.20	80.40	82.30	83.10
β	50.50	57.00	62.90	55.80	60.60	65.60	76.40	83.50	87.70
	63.60	71.50	76.50	72.30	74.20	77.30	87.70	95.60	98.80
χ^2_3	37.50	67.30	70.60	41.40	78.50	80.10	43.50	88.00	88.40
	51.50	79.60	82.30	54.10	91.30	93.20	59.60	97.30	98.30
t_5	33.30	42.20	44.60	38.10	44.80	44.80	44.10	51.20	52.00
	46.40	52.80	56.70	52.30	56.40	58.90	60.90	65.00	65.80

5.2. Heteroscedastic model

The reported results correspond to the model

$$Y_j = X_j + X_j^2 + (X_j + 0.5)\varepsilon_j, \quad 1 \le j \le n,$$

where X_j follows the uniform $(0,1)$ distribution. Since the model is heteroscedastic and the null hypothesis is simple, the simplifications in Remark 3.2 can be applied. Table 5 displays the results obtained for the type I error and the power for testing normality and Table 6 for testing GOF to the Laplace distribution. Similar conclusions to those given for Tables 1 and 2 can be also expressed in this case.

Table 5: (Heteroscedastic model) Percentage of rejections for the normality null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	4.50	6.00	5.40	6.50	4.90	5.36	4.82	5.92	4.90	5.32	5.08	5.74
	10.30	10.80	10.20	12.50	10.50	10.70	10.12	11.74	9.40	10.30	10.64	11.24
LP	16.40	43.00	60.00	64.30	34.20	60.00	87.10	88.50	65.00	75.60	99.50	99.50
	23.40	54.00	70.40	73.70	44.50	71.40	92.10	92.80	73.80	82.30	99.80	99.80
LG	7.40	57.60	56.40	63.10	8.50	91.90	94.70	95.20	12.60	99.80	100.00	100.00
	12.10	70.40	72.00	76.00	15.00	95.90	97.40	98.10	20.20	99.90	100.00	100.00
G	19.40	39.10	56.40	63.10	36.90	68.10	94.70	95.20	67.20	93.60	100.00	100.00
	29.90	55.00	72.00	76.00	49.90	80.30	97.40	98.10	76.10	97.60	100.00	100.00
β	43.00	16.10	57.60	63.30	86.20	77.00	99.80	99.80	99.90	95.20	100.00	100.00
	56.20	26.10	70.00	74.70	92.30	86.20	100.00	100.00	100.00	97.60	100.00	100.00
χ^2_3	50.90	41.60	85.50	89.10	83.00	71.30	99.70	99.70	99.20	95.70	100.00	100.00
	61.80	54.80	92.50	93.90	91.00	83.10	99.90	99.90	99.70	98.80	100.00	100.00
t_5	9.20	51.00	59.10	65.40	15.90	80.20	92.90	94.30	27.90	99.00	100.00	100.00
	16.20	65.70	71.60	76.50	23.40	89.30	97.70	98.00	36.80	99.90	100.00	100.00

Table 6: (Heteroscedastic model) Percentage of rejections for the Laplace null hypothesis at the significance levels 5% (upper entry) and 10% (lower entry).

	$n = 25$				$n = 50$				$n = 100$			
	A	PB	WB1	WB2	A	PB	WB1	WB2	A	PB	WB1	WB2
N	2.00	31.80	25.00	27.10	2.60	55.30	51.20	52.50	2.80	86.10	85.70	86.20
	4.90	40.30	34.70	37.10	7.40	64.80	61.80	62.90	7.80	90.50	91.20	91.40
LP	2.10	4.60	3.70	4.60	3.00	5.70	4.00	4.40	3.60	4.40	4.00	4.40
	6.80	10.00	8.00	9.60	7.30	11.50	9.20	10.20	7.80	9.10	8.40	9.00
LG	2.10	33.80	27.10	29.30	2.30	54.80	50.80	52.30	3.10	85.00	84.40	84.70
	6.30	43.80	37.60	40.20	6.80	64.40	61.60	62.50	7.00	89.30	89.60	89.90
G	2.10	31.30	23.50	25.50	2.80	53.90	50.20	51.50	3.00	85.30	85.10	85.60
	6.70	41.10	34.40	37.10	6.80	65.10	62.70	63.70	7.50	91.10	91.10	91.50
β	3.00	19.20	18.40	21.00	6.00	33.50	43.20	45.90	27.60	56.70	81.20	81.50
	8.00	27.40	29.10	31.70	14.60	43.70	55.30	56.80	39.60	68.50	87.30	87.80
χ^2_3	2.70	22.30	18.60	20.80	3.40	43.10	42.90	44.50	5.60	78.40	81.30	81.90
	7.10	30.80	27.30	30.10	7.60	54.50	54.10	56.60	12.70	84.10	87.30	87.70
t_5	2.90	30.60	22.80	24.50	3.90	56.80	53.20	53.90	4.60	84.30	83.90	84.30
	6.30	41.50	33.70	38.00	6.50	66.70	64.30	65.20	9.40	90.20	90.40	90.70

5.3. Time consumed

Table 7 compares the PB and the WB (with raw and centered multipliers) in terms of the required CPU time. This table shows the CPU time consumed in seconds to get a p-value for testing GOF for the normal and the Laplace distributions in the homoscedastic (for both single and composite null hypothesis) and the heteroscedastic models with sample sizes $n = 25$, 50, 100, 200. Looking at this table it becomes evident that the WB is more efficient than the PB, in terms of the required computing time, specially for larger sample sizes. The difference in time when using the raw and the centered multipliers is rather small.

Table 7: CPU time consumed for the calculation of one p-value in seconds for testing normality and Laplace distribution for the homoscedastic model and composite null hypothesis (upper entry), the heteroscedastic model (middle entry) and the homoscedastic model and single null hypothesis (lower entry).

n	Normal distribution			Laplace distribution		
	PB/WB1	WB1	WB2	PB/WB1	WB1	WB2
25	2.72	0.71	0.74	3.49	1.00	1.01
	7.45	0.33	0.35	7.17	0.54	0.60
	4.42	0.31	0.34	5.34	0.50	0.55
50	5.61	0.71	0.70	7.51	1.08	1.09
	30.88	0.17	0.22	38.15	0.26	0.25
	15.63	0.19	0.19	23.68	0.28	0.25
100	12.15	0.84	0.86	23.40	1.11	1.12
	52.80	0.25	0.27	74.33	0.42	0.45
	30.64	0.25	0.26	64.56	0.37	0.39
200	27.56	1.25	1.27	76.37	1.54	1.58
	66.19	0.59	0.62	127.80	0.83	0.83
	41.14	0.56	0.58	117.51	0.78	0.76

The gain in computational efficiency of the WB over the PB stems from the fact that one does not have to re-estimate the parameters at each iteration, which slows down the process considerably. Note that in the WB the parameter θ, the regression function $m(.)$ and the conditional variance function $\sigma(\cdot)$ are estimated only one time. For the WB approximation, once the set $\{m_{jk},\ 1 \le j \le k \le n\}$ is computed, the WB replicates $T^{*1}_{2,n,\omega}(\hat{\theta}), ..., T^{*B}_{2,n,\omega}(\hat{\theta})$ can be calculated very rapidly.

6. CONCLUSIONS

This paper proposes a WB approximation for the null distribution of a test statistic for testing GOF to the error distribution in nonparametric models. It provides a consistent estimator. The WB and the PB share this property. Nevertheless, from a computational point of view, the WB approximation is more efficient, in the sense of requiring less computation time. The numerical examples support these attributes. In addition, in cases were the asymptotic null distribution does not depend on unknown quantities, the simulations carried out declare that, for small to moderate sample sizes, the WB provides a better fit than the asymptotic distribution.

To derive the results in this paper we considered certain estimators for the regression function and the conditional variance function. In addition, we assumed that the covariate was univariate. The results could be extended by considering other estimators (such as other local polynomial estimators) as well as covariates with higher dimension. The null distribution of other test statistics (for example, those based on the empirical CDF) could be similarly approximated.

7. APPENDIX

7.1. Assumptions

(**A.2**) The weight function ω satisfies

$$\omega(t) = \omega(-t), \quad \forall t, \tag{7.1}$$

$\omega(t) \geq 0$, $\forall t$, and $\int t^4 \omega(t) dt < \infty$.

There is no restriction in assuming that the weight function $\omega(t)$ satisfies (7.1) because otherwise by defining $\omega_1(t) = 0.5\{\omega(t) + \omega(-t)\}$, which satisfies (7.1), we have that $T_{n,\omega}(\hat{\theta}) = T_{n,\omega_1}(\hat{\theta})$.

(**A.3**) $\varepsilon_1, ..., \varepsilon_n$ are IID with $E(\varepsilon_j^4) < \infty$ and $\varepsilon_1, ..., \varepsilon_n$ and $X_1, ..., X_n$ are independent.

Recall that by construction we have that $E(\varepsilon_j) = 0$ and $\mathrm{Var}(\varepsilon_j) = 1$.

(**A.4**)
- (**i**) X has a compact support S.
- (**ii**) f_X, m and σ are twice continuously differentiable on S.
- (**iii**) $\inf_{x \in S} f_X(x) > 0$ and $\inf_{x \in S} \sigma(x) > 0$.

(**A.5**) $nh_n^4 \to 0$, $nh_n^2/\ln n \to \infty$.

(**A.6**) K is a twice continuously differentiable symmetric pdf with compact support.

Assumptions (A.4)–(A.6) are mainly needed to guarantee the uniform consistency of the kernel estimators $\hat{f}_X(\cdot)$, $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ for $f_X(\cdot)$, $m(\cdot)$ and $\sigma(\cdot)$, respectively.

(**A.7**) The first partial derivatives $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $1 \le r \le p$, exist and are continuous functions $\forall t \in \mathbb{R}$, $\forall \theta$ in an open neighborhood of θ_1. In addition, $R'(t;\theta)$, $I'(t;\theta)$, $R_{(r)}(t;\theta)$, $I_{(r)}(t;\theta)$, $tR'(t;\theta)$, $tI'(t;\theta)$, $tR_{(r)}(t;\theta)$, $tI_{(r)}(t;\theta)$, $1 \le r \le p$, are bounded by functions in $L_2(\omega)$, $\forall \theta$ in an open neighborhood of θ_1.

The following assumption will be used for the maximum likelihood estimator of the parameter.

(**A.9**) The following functions exist $\forall \theta$ in an open neighborhood of θ_1:

$$u_r(x;\theta) = \frac{\partial}{\partial\theta_r} \log f(x;\theta) ,$$

$$u_{1,r}(x;\theta) = \frac{\partial^2}{\partial x \partial\theta_r} \log f(x;\theta) , \quad u_{0,r,s}(x;\theta) = \frac{\partial^2}{\partial\theta_r \partial\theta_s} \log f(x;\theta) ,$$

$$u_{2,r}(x;\theta) = \frac{\partial^3}{\partial x^2 \partial\theta_r} \log f(x;\theta) , \quad u_{1,r,s}(x;\theta) = \frac{\partial^3}{\partial x \partial\theta_r \partial\theta_s} \log f(x;\theta) ,$$

and satisfy

$$|u_{1,r}(a_1 + a_2 x;\theta)| \le b_{1,r}(x), \quad \text{with } x b_{1,r}(x),\, b_{1,r}(x) \in L_2(F) ,$$

$$|u_{0,r,s}(a_1 + a_2 x;\theta)| \le b_{0,r,s}(x) \in L_2(F) ,$$

$$|u_{2,r}(a_1 + a_2 x;\theta)| \le b_{2,r}(x) \in L_2(F) ,$$

$$|u_{1,r,s}(a_1 + a_2 x;\theta)| \le b_{1,r,s}(x) \in L_2(F) ,$$

$\forall a_1, a_2, \theta$ such that $|a_1|, |a_2 - 1|, |\theta - \theta_1| \le \delta$, for some small δ, $1 \le r, s \le p$. In addition, the following expectations exist:

$$E\big\{u_r(\varepsilon;\theta_1)\, u_s(\varepsilon;\theta_1)\big\} ,$$

$$E\big\{\varepsilon u_{1,r}(\varepsilon;\theta_1)\big\} ,$$

$1 \le r, s \le p$.

The following assumption will be used for the method of moment estimator of the parameter, which assumes that under the null hypothesis, $\theta_0 = g(\mu_0)$, for some known function $g = (g_1, ..., g_p)^T$, $g_r : \mathbb{R}^{k-1} \to \mathbb{R}$, $1 \le r \le p$:

(**A.10**) g_r is twice continuously differentiable at a neighborhood of μ_F, $1 \le r \le p$.

7.2. Proofs

We now sketch the proofs of the results stated in the previous sections, as well as some preliminary results. Along this section M denotes a generic positive constant taking many different values.

Lemma 7.1. *Suppose that assumptions (A.3)–(A.6) hold, then*

(a) $\frac{1}{n}\sum_{j=1}^{n}(\varepsilon_j-\hat{\varepsilon}_j)^2=o_p(1)$.

(b) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2-\varepsilon_j^2)^2=o_p(1)$.

(c) $\frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j^2-1)^2=O_p(1)$.

(d) $\frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j^2=O_p(1)$.

Proof: First, observe that under the considered assumptions (see, for example, Masry [16])

$$\sup_{x\in S}|\hat{m}(x)-m(x)| \;=\; o_p(n^{-1/4}), \tag{7.2}$$

$$\sup_{x\in S}|\hat{\sigma}(x)-\sigma(x)| \;=\; o_p(n^{-1/4}). \tag{7.3}$$

The difference between the residuals and the errors can be written as follows

$$\hat{\varepsilon}_j-\varepsilon_j=\varepsilon_j\left(\frac{\sigma(X_j)-\hat{\sigma}(X_j)}{\hat{\sigma}(X_j)}\right)+\left(\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)}\right). \tag{7.4}$$

The results in (a)–(d) follow from (7.2)–(7.4). □

Lemma 7.2. *If $\|\hat{\theta}-\theta_1\|=o_p(1)$ and (A.7) holds, then*

(a) $\|t\{R'(t;\hat{\theta})-R'(t;\theta_1)\}\|_\omega^2=o_p(1)$,
$\|t\{I'(t;\hat{\theta})-I'(t;\theta_1)\}\|_\omega^2=o_p(1)$.

(b) $\int\|\nabla R(t;\hat{\theta})-\nabla R(t;\theta_1)\|^2\omega(t)dt=o_p(1)$,
$\int\|\nabla I(t;\hat{\theta})-\nabla I(t;\theta_1)\|^2\omega(t)dt=o_p(1)$.

(c) $\|R(t;\hat{\theta})-R(t;\theta_1)\|_\omega^2=o_p(1)$,
$\|I(t;\hat{\theta})-I(t;\theta_1)\|_\omega^2=o_p(1)$.

(d) $\|t\{R(t;\hat{\theta})-R(t;\theta_1)\}\|_\omega^2=o_p(1)$,
$\|t\{I(t;\hat{\theta})-I(t;\theta_1)\}\|_\omega^2=o_p(1)$.

Proof: (a) From (A.7) $tR'(t;\theta)\in L_2(\omega)$, $\forall\theta$ in a neighborhood of θ_1. Since $\hat{\theta}\xrightarrow{P}\theta_1$, the integral $\int\{R'(t;\hat{\theta})-R'(t;\theta_1)\}^2t^2\omega(t)dt$ is finite with probability tending to 1. Thus, $\forall\epsilon>0, \exists M=M(\epsilon)>0$ such that

$$\int_{\mathbb{R}\backslash[-M,M]}\{R'(t;\hat{\theta})-R'(t;\theta_1)\}^2t^2\omega(t)dt<\epsilon, \tag{7.5}$$

with probability tending to 1. $tR'(t;\theta)$ is a uniformly continuous function in $[-M, M] \times B_\delta(\theta_1) = C$, where $B_\delta(\theta_1) = \{\theta : \|\theta - \theta_1\| \leq \delta\}$. Thus, $\forall \epsilon > 0$, $\exists\ \rho = \rho(\epsilon) > 0$ such that $\forall (t_a, \theta_a), (t_b, \theta_b) \in C$ satisfying $\|(t_a, \theta_a) - (t_b, \theta_b)\| < \rho$, we have $|t_1 R'(t_a; \theta_a) - t_2 R'(t_b; \theta_b)| < \epsilon/\iota$, with $\iota = \int \omega(t)dt$. As a consequence

$$\int_{-M}^{M} \{R'(t;\hat{\theta}) - R'(t;\theta_1)\}^2 t^2 \omega(t)dt < \epsilon, \tag{7.6}$$

with probability tending to 1. As ϵ is arbitrary, the result in (a) for the real part follows from (7.5) and (7.6). The proof for the imaginary part is parallel.

(b) The proof of this part is quite similar to that of part (a).

Parts (c) and (d) can be proven by applying the mean value theorem. □

Proof of Theorem 2.1: W^* can be expressed as $W^* = W_1 + W_2 + 2W_3$, where $W_3^2 \leq W_1 W_2$, $W_1 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n Z_0(\varepsilon_j; t, \theta_1)\xi_j\|_\omega^2$, $W_2 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n \{Z_0(\hat{\varepsilon}_j; t, \hat{\theta}) - Z_0(\varepsilon_j; t, \theta_1)\}\xi_j\|_\omega^2$. From the results in [4],

$$\sup_x |P_* \{W_1 \leq x\} - P\{W_0 \leq x\}| \xrightarrow{a.s.} 0.$$

Thus, to show the result it suffices to see that $W_2 = o_{p_*}(1)$ in probability. With this aim, observe that W_2 can be expressed as $W_2 = \sum_{j=1}^4 S_j + \sum_{j\neq k} S_{jk}$, with $S_{jk}^2 \leq S_j S_k$, $1 \leq j, k \leq 4$. In the proof of Theorem 3.1 it is given the expression of S_j and it is also proven that $S_j = o_{p_*}(1)$ in probability, $1 \leq j \leq 4$. This proves the result. □

Proof of Theorem 3.1: $T^*_{2,n,\omega}(\hat{\theta})$ can be expressed as $T^*_{2,n,\omega}(\hat{\theta}) = D_1 + D_2 + 2D_3$, where $D_1 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n Z_2(\varepsilon_j; t, \theta_1)\xi_j\|_\omega^2$, $D_2 = \|\frac{1}{\sqrt{n}} \sum_{j=1}^n \{Z_2(\hat{\varepsilon}_j; t, \hat{\theta}) - Z_2(\varepsilon_j; t, \theta_1)\}\xi_j\|_\omega^2$, $D_3^2 \leq D_1 D_2$. From the results in [4],

$$\sup_x |P_* \{D_1 \leq x\} - P\{T_2 \leq x\}| \xrightarrow{a.s.} 0.$$

Thus, to show the result it suffices to see that $D_2 = o_{p_*}(1)$ in probability. With this aim, observe that D_2 can be expressed as

$$D_2 = \sum_{j=1}^{10} S_j + \sum_{k<j} S_{jk},$$

with $S_{jk}^2 \leq S_j S_k$, $1 \leq j, k \leq 10$,

$$S_1 = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\cos(t\varepsilon_j) - \cos(t\hat{\varepsilon}_j)\}\xi_j\|_\omega^2,$$

$$S_2 = \|\tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\sin(t\varepsilon_j) - \sin(t\hat{\varepsilon}_j)\}\xi_j\|_\omega^2,$$

$$S_3 = \|\tfrac{1}{\sqrt{n}} \{R(t;\hat{\theta}) - R(t;\theta_1)\} \left(\textstyle\sum_{j=1}^n \xi_j\right) \|_\omega^2,$$

$$S_4 = \| \tfrac{1}{\sqrt{n}} \{I(t;\hat\theta) - I(t;\theta_1)\} \left(\textstyle\sum_{j=1}^n \xi_j\right) \|_\omega^2,$$
$$S_5 = \| \tfrac{t}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\hat\varepsilon_j R(t;\hat\theta) - \varepsilon_j R(t;\theta_1)\}\xi_j \|_\omega^2,$$
$$S_6 = \| \tfrac{t}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\hat\varepsilon_j I(t;\hat\theta) - \varepsilon_j I(t;\theta_1)\}\xi_j \|_\omega^2,$$
$$S_7 = \| \tfrac{t}{2\sqrt{n}} \textstyle\sum_{j=1}^n \{(\hat\varepsilon_j^2 - 1) R'(t;\hat\theta) - (\varepsilon_j^2 - 1) R'(t;\theta_1)\}\xi_j \|_\omega^2,$$
$$S_8 = \| \tfrac{t}{2\sqrt{n}} \textstyle\sum_{j=1}^n \{(\hat\varepsilon_j^2 - 1) I'(t;\hat\theta) - (\varepsilon_j^2 - 1) I'(t;\theta_1)\}\xi_j \|_\omega^2,$$
$$S_9 = \| \tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\psi_n^T(\hat\varepsilon_j;\hat\theta)\nabla R(t;\hat\theta) - \psi_1^T(\varepsilon_j;\theta)\nabla R(t;\theta_1)\}\xi_j \|_\omega^2,$$
$$S_{10} = \| \tfrac{1}{\sqrt{n}} \textstyle\sum_{j=1}^n \{\psi_n^T(\hat\varepsilon_j;\hat\theta)\nabla I(t;\hat\theta) - \psi_1^T(\varepsilon_j;\theta)\nabla I(t;\theta_1)\}\xi_j \|_\omega^2.$$

We will show that $S_j = o_{p_*}(1)$ in probability, $1 \le j \le 10$. By the mean value theorem,

$$S_1 = \frac{1}{n}\sum_{j,k=1}^n \xi_j\xi_k(\varepsilon_j - \hat\varepsilon_j)(\varepsilon_k - \hat\varepsilon_k)\int t^2 \sin(t\,\widetilde{\varepsilon}_j)\sin(t\,\widetilde{\varepsilon}_k)\omega(t)dt,$$

where $\widetilde{\varepsilon}_j = \alpha_j\varepsilon_j + (1-\alpha_j)\hat\varepsilon_j$, for some $\alpha_j \in (0,1)$. Then, from Lemma 7.1 (a),

$$E_*(S_1) \le \frac{1}{n}\sum_{j=1}^n (\varepsilon_j - \hat\varepsilon_j)^2 \int t^2\omega(t)dt = o_p(1),$$

which implies $S_1 = o_{p_*}(1)$ in probability. Analogously, $S_2 = o_{p_*}(1)$ in probability.

Since $S_3 = \left(\frac{1}{\sqrt{n}}\sum_{j=1}^n \xi_j\right)^2 \|R(t;\hat\theta) - R(t;\theta_1)\|_\omega^2$, the central limit theorem and Lemma 7.2 (c) imply that $S_3 = o_{p_*}(1)$ in probability. Analogously, $S_4 = o_{p_*}(1)$ in probability.

Observe that $S_5 = S_{51} + S_{52} + 2S_{53}$, with $S_{53}^2 \le S_{51}S_{52}$,

$$S_{51} = \tfrac{1}{n}\textstyle\sum_{j,k=1}^n (\hat\varepsilon_j - \varepsilon_j)(\hat\varepsilon_k - \varepsilon_k)\xi_j\xi_k \|tR(t;\hat\theta)\|_\omega^2,$$
$$S_{52} = \tfrac{1}{n}\textstyle\sum_{j,k=1}^n \varepsilon_j\varepsilon_k\xi_j\xi_k \|t\{R(t;\hat\theta) - R(t;\theta_1)\}\|_\omega^2.$$

From Lemma 7.1 (a) and Assumption (A.2), it follows that $E_*(S_{51}) = o_p(1)$ and thus $S_{51} = o_{p_*}(1)$, in probability. From Lemma 7.2 (d), it follows that $E_*(S_{52}) = o_p(1)$ and thus $S_{52} = o_{p_*}(1)$, in probability. Therefore, $S_5 = o_{p_*}(1)$, in probability. Analogously, $S_6 = o_{p_*}(1)$, in probability.

Observe that $S_7 = S_{71} + S_{72} + 2S_{73}$, with $S_{73}^2 \le S_{71}S_{72}$,

$$S_{71} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^n (\hat\varepsilon_j^2 - 1)(\hat\varepsilon_k^2 - 1)\xi_j\xi_k \|t\{R'(t;\hat\theta) - R'(t;\theta_1)\}\|_\omega^2,$$
$$S_{72} = \tfrac{1}{4}\tfrac{1}{n}\textstyle\sum_{j,k=1}^n (\hat\varepsilon_j^2 - \varepsilon_j^2)(\hat\varepsilon_k^2 - \varepsilon_k^2)\xi_j\xi_k \|tR'(t;\theta_1)\|_\omega^2.$$

From Lemma 7.1 (c) and Lemma 7.2 (a), it follows that $E_*(S_{71}) = o_p(1)$ and thus $S_{71} = o_{p_*}(1)$, in probability.

From Lemma 7.1 (b) and (A.7), it follows that $E_*(S_{72}) = o_p(1)$ and thus $S_{72} = o_{p_*}(1)$, in probability. Therefore, $S_7 = o_{p_*}(1)$, in probability. Analogously, $S_8 = o_{p_*}(1)$, in probability.

Observe that $S_9 = S_{91} + S_{92} + 2S_{93}$, with $S_{93}^2 \le S_{91}S_{92}$,

$$S_{91} = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n \{\psi_n(\hat\varepsilon_j;\hat\theta) - \psi_1(\varepsilon_j;\theta_1)\}^T \nabla R(t;\hat\theta)\xi_j\|_\omega^2,$$
$$S_{92} = \|\tfrac{1}{\sqrt{n}}\textstyle\sum_{j=1}^n \psi_1(\varepsilon_j;\theta_1)^T \{\nabla R(t;\hat\theta) - \nabla R(t;\theta_1)\}\xi_j\|_\omega^2.$$

From (3.2) and (A.7), it follows that $E_*(S_{91}) = o_p(1)$ and thus $S_{91} = o_{p_*}(1)$, in probability. From (A.1) and Lemma 7.2 (b), it follows that $E_*(S_{92}) = o_p(1)$ and thus $S_{92} = o_{p_*}(1)$, in probability. Therefore, $S_9 = o_{p_*}(1)$, in probability. Analogously, $S_{10} = o_{p_*}(1)$, in probability. This completes the proof. □

Proof of Corollary 3.2: From Theorem 3.1 it follows that $T^*_{2,n,\omega}(\hat\theta) = O_{p_*}(1)$ in probability. From Theorem 2 in [11], $\frac{T_{n,\omega}(\theta)}{n} \xrightarrow{P} \kappa > 0$. These two facts imply the result. □

Lemma 7.3. *Suppose that* $\|\hat\theta - \theta_1\| = o_p(1)$, *for some* $\theta_1 \in \Theta$, *and that assumptions (A.3)–(A.6), (A.9) hold, then*

(**a**) $\frac{1}{n}\sum_{j=1}^n \|\nabla \log f(\hat\varepsilon_j;\hat\theta) - \nabla \log f(\varepsilon_j;\theta_1)\|^2 = o_p(1)$.

(**b**) $\hat A_{n,rs}(\hat\theta) = A_{F,rs}(\theta_1) + o_p(1)$, $1 \le r, s \le p$.

(**c**) $\hat\rho_1(\hat\theta) = \rho_{1,F}(\theta_1) + o_p(1)$.

(**d**) $\hat\rho_2(\hat\theta) = \rho_{2,F}(\theta_1) + o_p(1)$.

Proof: (a) From the mean value theorem and (A.9),

$$\begin{aligned}
&\tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{\frac{\partial}{\partial\theta_r}\log f(\hat\varepsilon_j;\hat\theta) - \frac{\partial}{\partial\theta_r}\log f(\varepsilon_j;\theta_1)\right\}^2\\
&\quad = \tfrac{1}{n}\textstyle\sum_{j=1}^n \left\{\frac{\partial^2}{\partial\varepsilon\partial\theta_r}\log f(\tilde\varepsilon_j;\tilde\theta)(\hat\varepsilon_j - \varepsilon_j) + \sum_{s=1}^p \frac{\partial^2}{\partial\theta_r\partial\theta_s}\log f(\tilde\varepsilon_j;\tilde\theta)(\hat\theta_s - \theta_{1s})\right\}^2\\
&\quad \le S_{r,1} + S_{r,2} + 2S_{r,3},
\end{aligned}$$

with $S_{r,3}^2 \le S_{r,1}S_{r,2}$, $\tilde\varepsilon_j = (1-\alpha_j)\hat\varepsilon_j + \alpha_j\varepsilon_j$, for some $\alpha_j \in (0,1)$, $1 \le j \le n$, $\tilde\theta = (1-\alpha)\hat\theta + \alpha\theta_1$, for some $\alpha \in (0,1)$,

$$S_{r,1} = \|\hat\theta - \theta_1\|^2 \frac{1}{n}\sum_{j=1}^n\sum_{s=1}^p b_{0,r,s}^2(\varepsilon_j)$$

and

$$S_{r,2} = \frac{1}{n}\sum_{j=1}^n b_{1,r}^2(\varepsilon_j)(\hat\varepsilon_j - \varepsilon_j) = o_p(1).$$

From (A.9), (7.2)–(7.4), it follows that $S_{r,1} = o_p(1)$, $S_{r,2} = o_p(1)$, $1 \le r \le p$. This proves (a).

The proof of parts (b)–(d) follows similar steps to that of part (a). □

Proof of Theorem 4.1: Observe that $\frac{1}{n}\sum_{j=1}^{n} \|\psi_{1n}(\hat{\varepsilon}_j;\hat{\theta}) - \psi(\varepsilon_j;\theta_1)\|^2 \le D_1 + D_2 + D_3 + D_4$, with $D_4^2 \le \sum_{j\ne k} D_j D_k$,

$$D_1 = \frac{1}{n}\sum_{j=1}^{n} \|\hat{A}_n(\hat{\theta})^{-1}\nabla \log f(\hat{\varepsilon}_j;\hat{\theta}) - A_F(\theta_1)^{-1}\nabla \log f(\varepsilon_j;\theta_1)\|^2,$$

$$D_2 = \frac{1}{n}\sum_{j=1}^{n} \|\hat{\varepsilon}_j\hat{\rho}_1(\hat{\theta}) - \varepsilon_j\rho_{F,1}(\theta_1)\|^2,$$

$$D_3 = \frac{1}{n}\sum_{j=1}^{n} \|\frac{\hat{\varepsilon}_j^2 - 1}{2}\hat{\rho}_2(\hat{\theta}) - \frac{\varepsilon_j^2 - 1}{2}\rho_{F,2}(\theta_1)\|^2.$$

By using the results in Lemmas 7.1 and 7.3 one obtain $D_j = o_p(1)$, $1 \le j \le 3$, and hence the result. □

Proof of Theorem 4.2: From (7.2)–(7.4),

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\hat{\varepsilon}_j^s = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^s + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s-1}\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)} + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s}\frac{\sigma(X_j)-\hat{\sigma}(X_j)}{\hat{\sigma}(X_j)} + o_p(1). \tag{7.7}$$

Taking into account the following facts

(m.1) $\sup_{x\in S}\left|\frac{\hat{m}(x)-m(x)}{\hat{\sigma}(x)} - \frac{\hat{m}(x)-m(x)}{\sigma(x)}\right| = o_p(n^{-1/2})$,

(m.2) $\sup_{x\in S}\left|\hat{m}(x) - m(x) - \frac{1}{nf_X(x)}\sum_{k=1}^{n_v} K_{h_n}(x-X_k)\sigma(X_k)\varepsilon_k\right| = o_p(n^{-1/2})$,

it follows that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s-1}\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)} = \frac{-1}{n\sqrt{n}}\sum_{j,k=1}^{n}\varepsilon_j^{s-1}\varepsilon_k\frac{\sigma(X_k)}{f_X(X_j)\sigma(X_j)}K_{h_n}(X_j-X_k) + o_p(1).$$

Now, by using projections, we get (see, for example, the proof of Theorem 2 in [18] for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j^{s-1}\frac{m(X_j)-\hat{m}(X_j)}{\hat{\sigma}(X_j)} = -\mu_{F,s-1}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_j + o_p(1). \tag{7.8}$$

Next we deal with the third term in the right-hand side of (7.7). Taking into account the following facts

(s.1) $\sup_{x\in S}\left|\frac{\hat\sigma(x)-\sigma(x)}{\hat\sigma(x)}-\frac{\hat\sigma(x)-\sigma(x)}{\sigma(x)}\right|=o_p(n^{-1/2})$,

(s.2) $\sup_{x\in S}\left|\hat\sigma(x)-\sigma(x)-\frac{\hat\sigma^2(x)-\sigma^2(x)}{2\sigma(x)}\right|=o_p(n^{-1/2})$,

(s.3) $\sup_{x\in S}\Big|\hat\sigma^2(x)-\sigma^2(x)-\frac{1}{nf_X(x)}\sum_{j=1}^n K_{h_n}(X_j-x)$
$\cdot\left[\{Y_j-m(x)\}^2-\sigma^2(x)\right]\Big|=o_p(n^{-1/2})$,

it follows that

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s\frac{\sigma(X_j)-\hat\sigma(X_j)}{\hat\sigma(X_j)}=$$

$$=\frac{1}{2n\sqrt{n}}\sum_{j,k=1}^n \varepsilon_j^s\frac{1}{f_X(X_j)\sigma^2(X_j)}K_{h_n}(X_j-X_k)\left[\sigma^2(X_j)-\{Y_k-m(X_j)\}^2\right]+o_p(1).$$

Now, by using projections, we get (see, for example, the proof of Lemma 11 in [19] for a similar development)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n \varepsilon_j^s\frac{\sigma(X_j)-\hat\sigma(X_j)}{\hat\sigma(X_j)}=-\frac{\mu_{F,s}}{2}\frac{1}{\sqrt{n}}\sum_{j=1}^n(\varepsilon_j^2-1)+o_p(1). \tag{7.9}$$

The result follows from (7.7)–(7.9). □

Proof of Theorem 4.3: Notice that

$$\hat\mu_s-\mu_{F,s}=\frac{1}{n}\sum_{j=1}^n(\hat\varepsilon_j^s-\varepsilon_j^s)+\frac{1}{n}\sum_{j=1}^n(\varepsilon_j^s-\mu_{F,s}).$$

From (7.2)–(7.4), the first term in the right-hand side of the above equality is $o_p(1)$; from the SLLN, the second term in the right-hand side of the above equality is $o(1)$ a.s. Therefore $\hat\mu_s-\mu_{F,s}=o_p(1)$, $2\le s\le k$. The result follows from this fact and (A.10). □

ACKNOWLEDGMENTS

The authors thank the anonymous referee and the AE for their helpful suggestions and constructive comments. G.I. Rivas-Martínez acknowledges financial support from Fundación Carolina, Universidad Nacional de Asunción and Universidad de Sevilla. M.D. Jiménez-Gamero acknowledges financial support from grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness, and grant MTM2017-89422-P of the Spanish Ministry of Economy, Industry and Competitiveness, ERDF support included.

REFERENCES

[1] Bickel, P.J. and Doksum, K.A. (2001). *Mathematical Statistics*, Prentice Hall, New Jersey.

[2] Burke, M.D. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap, *Statistics & Probability Letters*, **46**, 13–20.

[3] de Uña-Álvarez, J. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 414–418.

[4] Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for U-statistics, *Journal of Multivariate Analysis*, **51**, 392–413.

[5] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York.

[6] Heuchenne, C. and Van Keilegom, I. (2010). Goodness of fit tests for the error distribution in nonparametric regression, *Computational Statistics & Data Analysis*, **54**, 1942–1951.

[7] Epps, T.W. (2005). Tests for location-scale families based on the empirical characteristic function, *Metrika*, **62**, 99–114.

[8] Epps, T.W. and Pulley, L.B. (1983). A test for normality based on the empirical characteristics function, *Biometrika*, **70**, 723–726.

[9] González-Manteiga, W. and Crujeiras, R. (2013). An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 361–411.

[10] Hušková, M. and Janssen, P. (1993). Consistency of the generalized bootstrap for degenerate U-statistics, *The Annals of Statistics*, **21**, 1811–1823.

[11] Hušková, M. and Meintanis, S.G. (2010). Test for the error distribution in nonparametric possiby heteroscedatic regression models, *Test*, **19**, 92–112.

[12] Jiménez-Gamero, M.D.; Alba-Fernández, V.; Muñoz-García, J. and Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristics functions, *Computational Statistics & Data Analysis*, **53**, 3957–3971.

[13] Jiménez-Gamero, M.D. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 412–413.

[14] Jiménez-Gamero, M.D. and Kim, H-M. (2015). Fast goodness-of-fit test based on the characteristic function, *Computational Statistics & Data Analysis*, **89**, 172–191.

[15] Kojadinovic, I. and Yan, J. (2012). Goodness-of-fit testing based on a weighted bootstrap: A fast sample alternative to the parametric bootstrap, *The Canadian Journal of Statistics*, **40**, 480–500.

[16] Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates, *Journal of Time Series Analysis*, **17**, 571–600.

[17] Neumeyer, N.; Dette, H. and Nagel, E-R. (2006). Bootstrap test for the error distribution in linear and nonparametric regression models, *Australian & New Zealand Journal of Statistics*, **48**, 129–156.

[18] Pardo-Fernández, J.C.; Jiménez-Gamero, M.D. and El Ghouch, A. (2015). A nonparametric ANOVA-type test for regression curves based on characteristic functions, *Scandinavian Journal of Statistics*, **42**, 197–213.

[19] Pardo-Fernández, J.C.; Jiménez-Gamero, M.D. and El Ghouch, A. (2015). Tests for the equality of conditional variance functions in nonparametric regression, *Electronic Journal of Statistics*, **9**, 1826–1851.

[20] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, URL: `http://www.R-project.org/`, Vienna, Austria.

[21] Sperlich, S. (2013). Comments on: An updated review of Goodness-of-Fit tests for regression models, *Test*, **22**, 419–427.

[22] White, H. (1982). Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.

[23] Zhu, LX. (2005). *Lecture Notes in Statistics. Nonparametric Monte Carlo test and their applications*, Springer, Berlin.