Electronic Journal of Statistics
Vol. 9 (2015) 1826–1851
ISSN: 1935-7524
DOI: 10.1214/15-EJS1058

Tests for the equality of conditional variance functions in nonparametric regression

Juan Carlos Pardo-Fernández*

Department of Statistics and Operational Research, Universidade de Vigo
Campus Universitario, 36310 Vigo, Spain
e-mail: juancp@uvigo.es

María Dolores Jiménez-Gamero*

Department of Statistics and Operational Research, Universidad de Sevilla
Facultad de Matemáticas, Calle Tarfia s.n., 41012 Sevilla, Spain
e-mail: dolores@us.es

and

Anouar El Ghouch†

Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium
e-mail: anouar.elghouch@uclouvain.be

Abstract: In this paper we are interested in checking whether the conditional variances are equal in $k \geq 2$ location-scale regression models. Our procedure is fully nonparametric and is based on the comparison of the error distributions under the null hypothesis of equality of variances and without making use of this null hypothesis. We propose four test statistics based on empirical distribution functions (Kolmogorov-Smirnov and Cramér-von Mises type test statistics) and two test statistics based on empirical characteristic functions. The limiting distributions of these six test statistics are established under the null hypothesis and under local alternatives. We show how to approximate the critical values using either an estimated version of the asymptotic null distribution or a bootstrap procedure. Simulation studies are conducted to assess the finite sample performance of the proposed tests. We also apply our tests to data on household expenditures.

MSC 2010 subject classifications: Primary 62G10, 62G08; secondary 62G20, 62G09.
Keywords and phrases: Asymptotics, bootstrap, comparison of curves, empirical characteristic function, empirical distribution function, kernel smoothing, local alternatives, regression residuals.

Received July 2014.

*These authors were financially supported by grant MTM2014-55966-P (ERDF support included) of the Spanish Ministry of Economy and Competitiveness.

†This author was financially supported by IAP research network P6/03 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

1. Introduction

When comparing k ($k \geq 2$) populations it is interesting not only comparing the means, but also other characteristics, like the variances. For example, in quality control, it is important to check the uniformity and the stability of the production process under different experimental and practical conditions. In biomedical research, detecting variation in gene expression levels is important for many reasons, for example, to identify experimental and environmental factors that affect a biological process; for a concrete example, see e.g. [19]. Equality of variances, when satisfied, can also be used to develop more powerful and simple ANOVA-type test statistics. Without controlling for the effect of covariates, there are a substantial number of tests available in the literature for the equality of (unconditional) variances of two or more populations. The standard procedures include the classical F-test and Levene's test (see [17]) which is known to be robust to the violation of normality; see [11] for a recent review and some interesting examples and applications.

In this paper, we are interested in the comparison of conditional variances. We assume that in each population, along with the variable of interest or response variable, Y, it is also observed another variable, X, the covariate, so that the mean and the variance of the response variable depend on the values of X. More specifically, let (X_j, Y_j), $1 \leq j \leq k$, be k independent random vectors satisfying general nonparametric regression models

$$Y_j = m_j(X_j) + \sigma_j(X_j)\varepsilon_j, \tag{1}$$

where $m_j(x) = E(Y_j \mid X_j = x)$ is the regression function, $\sigma_j^2(x) = Var(Y_j \mid X_j = x)$ is the conditional variance function and ε_j is the regression error, which is assumed to be independent of X_j. Note that, by construction, $E(\varepsilon_j) = 0$ and $Var(\varepsilon_j)$=1. The covariate X_j is continuous with density function f_j. Since the objective is to compare the variance functions, it is reasonable to assume that the covariates have common support, say R. The regression functions, the variance functions, the distribution of the errors and the distribution of the covariates are completely unknown and no parametric models are assumed for them. Thus, our approach is completely nonparametric. In this conditional setting, the hypothesis of equality of variances is stated in terms of the conditional variance functions, $H_0 : \sigma_1^2(x) = \sigma_2^2(x) = \cdots = \sigma_k^2(x)$ for all $x \in R$, or, equivalently,

$$H_0 : \sigma_j(x)/\sigma_0(x) = 1, \quad \text{for } 1 \leq j \leq k,$$

where $\sigma_0^2(x)$ is the common variance that can be expressed as

$$\sigma_0^2(x) = \sum_{j=1}^{k} \pi_j(x)\sigma_j^2(x),$$

for some positive functions $\pi_1, \ldots, \pi_k$ satisfying $\sum_{j=1}^{k} \pi_j(x) = 1$. The alternative hypothesis is

$$H_1 : \sigma_j(x)/\sigma_0(x) \neq 1, \quad \text{for some } j \in \{1 \ldots, k\}.$$

We will develop several test statistics and study their distribution under H_0 and under local alternatives converging to the null hypothesis at the rate $n^{-1/2}$, n being the total sample size. Specifically, we consider the following local alternative hypothesis

$$H_{1,n} : \sigma_j(x)/\sigma_0(x) = 1 + n^{-1/2}\delta_j(x), \quad \text{for } 1 \le j \le k,$$

for some functions δ_j. To be more precise, in the previous expression we should have written $\sigma_{n,j}(x)/\sigma_{n,0}(x)$ instead of $\sigma_j(x)/\sigma_0(x)$, as this function depends on n. However, to short the notation we suppress this explicit dependence on n. Observe that as n increases $H_{1,n}$ becomes closer and closer to H_0. Also, when $\delta_j(x) = 0$, for $1 \le j \le k$, $H_{1,n}$ reduces to H_0.

Statistical literature concerning the problem of testing for common features in several regression models has mainly focused on testing for common regression curves or testing for common error distributions. The problem of testing for the equality of regression curves in nonparametric settings has been extensively treated; see for example [3, 16, 22, 25, 24, 27], and [12] for a recent review. On the other hand, testing for the equality of error distributions has been addressed in [23]. To the best of our knowledge, the comparison of conditional variance functions has not been studied before. Most papers dealing with testing on the conditional variance function focus on homoscedasticity assumption (see for example [18, 4] and the references therein) or, more in general, on the parametric form of the conditional variance function (see for example [5, 15] and the references therein).

In order to construct a test for testing H_0, several approaches are possible. Here we follow the ideas in [25, 24] for testing the equality of the regression functions, $m_1, \ldots, m_k$, which consist of comparing the distributions of the errors of the regression models. Specifically, let

$$\varepsilon_j = \frac{Y_j - m_j(X_j)}{\sigma_j(X_j)} \tag{2}$$

be the regression error in population j, $1 \le j \le k$. Define

$$\varepsilon_{0j} = \frac{Y_j - m_j(X_j)}{\sigma_0(X_j)} = \varepsilon_j \frac{\sigma_j(X_j)}{\sigma_0(X_j)} \tag{3}$$

to be the error under the null hypothesis, $1 \le j \le k$. Let $F_{\varepsilon_j}(t) = P(\varepsilon_j \le t)$ and $F_{\varepsilon_{0j}}(t) = P(\varepsilon_{0j} \le t)$ be the cumulative distribution function (CDF) of ε_j and ε_{0j}, respectively. The following theorem shows that H_0 is true if and only if the distributions of ε_j and ε_{0j} coincide. The proof can be found in the Appendix.

Theorem 1. *Assume that σ_j is continuous on R and $0 < E(\varepsilon_j^4) < \infty$, for $1 \le j \le k$.*

(i) H_0 is true if and only if the random variables ε_j and ε_{0j} have the same distribution for all $1 \le j \le k$.

(ii) Let $p_1, \ldots, p_k$ be such that $p_j > 0$, $1 \le j \le k$, and $\sum_{j=1}^k p_k = 1$. Let $F_\varepsilon(t) = \sum_{j=1}^k p_j F_{\varepsilon_j}(t)$ and $F_{\varepsilon_0}(t) = \sum_{j=1}^k p_j F_{\varepsilon_{0j}}(t)$. Assume also that

$E(\varepsilon_1^4) = \cdots = E(\varepsilon_k^4)$. Then H_0 is true if and only if $F_\varepsilon(t) = F_{\varepsilon_0}(t)$, for all $t \in \mathbb{R}$.

The assertions in the previous result can be interpreted in terms of the CDF or in terms of any other function characterizing a probability law, such as the characteristic function (CF). In this paper we will consider both cases, that is, to test H_0 we will compare consistent estimators of the CDFs and CFs of the random variables ε_j and ε_{0j}, $1 \le j \le k$.

With this aim, the paper is organized as follows. In Section 2 we introduce the test statistics and explain the testing procedures. Sections 3 and 4 contain the main asymptotic results concerning the empirical CDF-based test statistics and the empirical CF-based test statistics, respectively, and discuss some practical considerations. In Section 5 we explain how the critical values of the proposed test statistics can be approximated. Investigating the finite sample performance of our tests is the topic of Section 6. A data example follows in Section 7 and conclusions are given in Section 8. All proofs of the theoretical results are deferred to the Appendix.

The following notation will be used along the paper: P_0 denotes probability assuming that H_0 is true; E_0 denotes expectation assuming that H_0 is true; 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; if $x \in \mathbb{R}^k$, with $x' = (x_1, \ldots, x_k)$, then $diag(x)$ is the $k \times k$ diagonal matrix whose (i,i) entry is x_i, $1 \le i \le k$; for any complex number $z = a + \mathrm{i}b$, $Re(z) = a$ is its real part, $Im(z) = b$ is its imaginary part, $\bar{z} = a - \mathrm{i}b$ is its conjugate and $|z|$ is its modulus; $N_k(\mu, \Sigma)$ denotes the k-variate normal distribution with mean vector μ and variance-covariance matrix Σ; an unspecified integral denotes integration over the whole real line $\mathbb{R}$; $\sup_t$ stands for $\sup_{t\in\mathbb{R}}$; $I(S)$ denotes the indicator function of a set S.

2. The test statistics

As in the Introduction, let (X_j, Y_j), $1 \le j \le k$, be k independent random vectors satisfying general nonparametric regression models (1). For $1 \le j \le k$, let ε_j and ε_{0j} be as defined in (2) and (3), respectively. As justified in Theorem 1, to test for H_0 we will compare consistent estimators of the CDFs and CFs of the random variables ε_j and ε_{0j}, $1 \le j \le k$, and also consistent estimators of the CDFs F_ε and $F_{\varepsilon 0}$ and of their associated CFs. Since neither ε_j nor ε_{0j} are observable, the inference must be based on residuals. Next we construct them.

Let (X_{jl}, Y_{jl}), $1 \le l \le n_j$, be independent and identically distributed (iid) observations from (X_j, Y_j), $1 \le j \le k$, and let $n = \sum_{j=1}^k n_j$. Along the paper it will be assumed that $n_j/n \to p_j > 0$, $1 \le j \le k$. In order to estimate the errors, we first need to estimate the regression functions, $m_j(x) = E(Y_j|X_j = x)$, the variance functions, $\sigma_j^2(x) = E[\{Y_j - m_j(x)\}^2|X_j = x]$, and the common variance function under H_0, $\sigma_0^2(x)$. With this aim we use nonparametric estimators based on kernel smoothing techniques. Let K denote a nonnegative kernel function defined on $\mathbb{R}$, let $0 < h_n \equiv h \to 0$ be the bandwidth or smoothing parameter and $K_h(x) = h^{-1}K(x/h)$. We use the following estimators for the functions m_j, σ_j^2 and σ_0^2:

$$\hat{m}_j(x) = \sum_{l=1}^{n_j} w_{jl}(x)Y_{jl},\ \hat{\sigma}_j^2(x) = \sum_{l=1}^{n_j} w_{jl}(x)Y_{jl}^2 - \hat{m}_j^2(x),\ \hat{\sigma}_0^2(x) = \sum_{j=1}^{k} \pi_j(x)\hat{\sigma}_j^2(x).$$

The quantities w_{jl} are either the local-linear weights given by

$$w_{jl}(x) = \frac{K_h(X_{jl} - x)\left\{S_{2,n_j}(x) - (X_{jl} - x)S_{1,n_j}\right\}}{S_{0,n_j}(x)S_{2,n_j}(x) - S_{1,n_j}^2(x)},$$

with $S_{k,n_j}(x) = \sum_{l=1}^{n_j}(X_{jl}-x)^k K_h(X_{jl}-x)$, $k = 0, 1, 2$, or the Nadaraya-Watson weights

$$w_{jl}(x) = \frac{K_h(X_{jl} - x)}{\sum_{v=1}^{n_j} K_h(X_{jv} - x)}.$$

Both are particular cases of local-polynomial weighting (see [8]). Under the model assumptions that will be stated in the next section, the results in this article are valid for local-linear and for Nadaraya-Watson (local-constant) estimators. Note that we have implicitly assumed that the functions $\pi_1, \ldots, \pi_k$ do not depend on unknowns. The theory also apply to the case where they depend on unknowns, replacing π_j by $\hat{\pi}_j$ in the expression $\hat{\sigma}_0^2(x)$, whenever $\hat{\pi}_j$ converges to π_i fast enough. Later we will discuss this issue in more detail.

Based on these estimators, for each population j, $1 \le j \le k$, we construct two samples of residuals,

$$\hat{\varepsilon}_{jl} = \frac{Y_{jl} - \hat{m}_j(X_{jl})}{\hat{\sigma}_j(X_{jl})} \qquad \text{and} \qquad \hat{\varepsilon}_{0jl} = \frac{Y_{jl} - \hat{m}_j(X_{jl})}{\hat{\sigma}_0(X_{jl})}, \tag{4}$$

$1 \le l \le n_j$. Then we can construct the corresponding empirical CDFs (ECDFs),

$$\hat{F}_{\varepsilon_j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} I(\hat{\varepsilon}_{jl} \le t) \qquad \text{and} \qquad \hat{F}_{\varepsilon_{0j}}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} I(\hat{\varepsilon}_{0jl} \le t),$$

and empirical CFs (ECFs),

$$\hat{\varphi}_{\varepsilon_j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} \exp(\mathrm{i}t\hat{\varepsilon}_{jl}) \qquad \text{and} \qquad \hat{\varphi}_{\varepsilon_{0j}}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} \exp(\mathrm{i}t\hat{\varepsilon}_{0jl}),$$

respectively. These ECDFs are consistent kernel-based nonparametric estimators of the population CDFs $F_{\varepsilon_j}(t)$ and $F_{\varepsilon_{0j}}(t)$, respectively (see Theorem 2 below). Analogously, the above ECFs are consistent kernel-based nonparametric estimators of the population CFs $\varphi_{\varepsilon_j}(t) = E\{\exp(\mathrm{i}t\varepsilon_j)\}$ and $\varphi_{\varepsilon_{0j}}(t) = E\{\exp(\mathrm{i}t\varepsilon_{0j})\}$, respectively (see Theorem 6 below). We can also consider the following ECDFs

$$\hat{F}_{\varepsilon}(t) = \frac{1}{n}\sum_{j=1}^{k}\sum_{l=1}^{n_j} I(\hat{\varepsilon}_{jl} \le t) \qquad \text{and} \qquad \hat{F}_{\varepsilon_0}(t) = \frac{1}{n}\sum_{j=1}^{k}\sum_{l=1}^{n_j} I(\hat{\varepsilon}_{0jl} \le t),$$

and ECFs

$$\hat{\varphi}_{\varepsilon}(t) = \frac{1}{n}\sum_{j=1}^{k}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\hat{\varepsilon}_{jl}) \qquad \text{and} \qquad \hat{\varphi}_{\varepsilon_0}(t) = \frac{1}{n}\sum_{j=1}^{k}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\hat{\varepsilon}_{0jl}),$$

which estimate the functions $F_{\varepsilon}(t) = \sum_{j=1}^{k} p_j F_{\varepsilon_j}(t)$, $F_{\varepsilon_0}(t) = \sum_{j=1}^{k} p_j F_{\varepsilon_{0j}}(t)$, $\varphi_{\varepsilon}(t) = \sum_{j=1}^{k} p_j \varphi_{\varepsilon_j}(t)$ and $\varphi_{\varepsilon_0}(t) = \sum_{j=1}^{k} p_j \varphi_{\varepsilon_{0j}}(t)$, respectively.

To test for H_0, we will construct Kolmogorov-Smirnov type statistics and Cramér-von Mises type statistics to compare the ECDFs, and weighted L_2-distances to compare the ECFs. More precisely, the considered statistics are

$$\begin{aligned}
T_{KS}^{1} &= \sum_{j=1}^{k}\sqrt{n_j}\sup_t |\hat{F}_{\varepsilon_j}(t) - \hat{F}_{\varepsilon_{0j}}(t)|,\\
T_{CM}^{1} &= \sum_{j=1}^{k} n_j \int \{\hat{F}_{\varepsilon_j}(t) - \hat{F}_{\varepsilon_{0j}}(t)\}^2 d\hat{F}_{\varepsilon_{0j}}(t),\\
T_{KS}^{2} &= \sqrt{n}\sup_t |\hat{F}_{\varepsilon}(t) - \hat{F}_{\varepsilon_0}(t)|,\\
T_{CM}^{2} &= n\int \{\hat{F}_{\varepsilon}(t) - \hat{F}_{\varepsilon_0}(t)\}^2 d\hat{F}_{\varepsilon_0}(t),\\
T_1 &= \sum_{j=1}^{k} n_j \int \left|\hat{\varphi}_{\varepsilon_j}(t) - \hat{\varphi}_{\varepsilon_{0j}}(t)\right|^2 w(t)dt,\\
T_2 &= n\int |\hat{\varphi}_{\varepsilon}(t) - \hat{\varphi}_{\varepsilon_0}(t)|^2 w(t)dt,
\end{aligned}$$

where w is a positive weight function that is needed to guarantee consistency (see Section 4). Note that in the case of T_1 and T_2, $|\cdot|$ represents the modulus of a complex number. In Section 3 we will study the asymptotic properties of the statistics T_{KS}^1, T_{CM}^1, T_{KS}^2 and T_{CM}^2 and in Section 4 we will deal with T_1 and T_2.

3. Asymptotics for ECDF-based test statistics

This section studies some asymptotic properties of the ECDF-based test statistics T_{KS}^1, T_{CM}^1, T_{KS}^2 and T_{CM}^2. To derive them we will need some commonly assumed regularity assumptions. First let us define $F_j(t|x) = P(Y_j \le t|X_j = x)$ and $F_j(x) = P(X_j \le x)$, for $1 \le j \le k$.

Assumption (A1): For $1 \le j \le k$,

- **(i)** X_j is absolutely continuous with compact support R and density f_j.
- **(ii)** f_j, m_j, σ_j and π_j are twice continuously differentiable on R.
- **(iii)** $\inf_{x\in R} f_j(x) \ge c > 0$ and $\inf_{x\in R}\sigma_j(x) \ge d > 0$, for some $c, d \in \mathbb{R}$.
- **(iv)** $E(\varepsilon_j^4) < \infty$.
- **(v)** $nh_n^4 \to 0$ and $nh_n^{3+2\delta}(\log h_n^{-1})^{-1} \to \infty$, for some $\delta > 0$.

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

Assumption (A2): For $1 \leq j \leq k$, $F_j(t|x)$ is continuous in (x,t) and differentiable with respect to t, $\frac{\partial}{\partial t}F_j(t|x) = F_j'(t|x)$ is continuous in (x,t) and $\sup_{x,t}|t^2 F_j'(t|x)| < \infty$. The same holds for all other partial derivatives of $F_j(t|x)$ with respect to x and t up to order two.

From now on we will name Assumption A to be the set of Assumptions (A1)–(A2). Assumption A (skipping (A1)(iv)) was also considered in [25] to derive asymptotic properties of some ECDF-based tests designed to detect differences between the conditional mean functions. This assumption is mainly needed to guarantee the uniform consistency of the estimators $\hat{f}_j$, $\hat{\sigma}_j$, $\hat{m}_j$ and $\hat{\sigma}_0$. Also note that Assumption (A2), which will only be needed for the asymptotics related to ECDF-based tests, implies that ε_j has a density, denoted by f_{ε_j}.

We first give the following result that justifies the use of the test statistics T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM} for testing H_0.

Theorem 2. *Suppose that Assumption A holds. Then,* $\hat{F}_{\varepsilon_{0j}}(t) = F_{\varepsilon_{0j}}(t) + o_p(1)$ *and* $\hat{F}_{\varepsilon_j}(t) = F_{\varepsilon_j}(t) + o_p(1)$*, uniformly in* t*,* $1 \leq j \leq k$*.*

Corollary 3. *Suppose that Assumption A holds. Then,*

$$\frac{1}{\sqrt{n}}T^1_{KS} \xrightarrow{P} \sum_{j=1}^{k} \sqrt{p_j}\sup_t |F_{\varepsilon_j}(t) - F_{\varepsilon_{0j}}(t)|,$$

$$\frac{1}{n}T^1_{CM} \xrightarrow{P} \sum_{j=1}^{k} p_j \int \{F_{\varepsilon_j}(t) - F_{\varepsilon_{0j}}(t)\}^2 dF_{\varepsilon_{0j}}(t),$$

$$\frac{1}{\sqrt{n}}T^2_{KS} \xrightarrow{P} \sup_t |F_{\varepsilon}(t) - F_{\varepsilon_0}(t)|,$$

$$\frac{1}{n}T^2_{CM} \xrightarrow{P} \int \{F_{\varepsilon}(t) - F_{\varepsilon_0}(t)\}^2 dF_{\varepsilon_0}(t).$$

Observe that all considered test statistics converge in probability to non-negative quantities. Under the assumptions in Theorem 1, such quantities are 0 if and only if H_0 is true. Therefore it seems reasonable to reject the null hypothesis for *large* values of these test statistics. Now, to determine what a large value means in each case, we must calculate the null distribution of the test statistic, or at least an approximation to it. Since the null distributions are unknown, we study their asymptotic null distributions.

Theorem 4. *Suppose that Assumption A holds. Then, under* $H_{1,n}$*,*

$$\sqrt{n_j}\left\{\hat{F}_{\varepsilon_j}(t) - \hat{F}_{\varepsilon_{0j}}(t)\right\} = \frac{1}{2} t f_{\varepsilon_j}(t)(p_j^{1/2}\Delta_j + Z_{n,j}) + o_p(1),$$

uniformly in t*, where* $\Delta_j = 2E\left[\delta_j(X_j)\right]$*, and*

$$Z_{n,j} = \sqrt{n_j}\sum_{v=1}^{k}\frac{1}{n_v}\sum_{l=1}^{n_v}\left\{I(v=j) - \pi_v(X_{vl})\frac{f_j(X_{vl})}{f_v(X_{vl})}\right\}\left(\varepsilon_{vl}^2 - 1\right). \qquad (5)$$

The next Corollary, derived mainly by applying the multivariate Central Limit Theorem to $Z_n = (Z_{n,1}, \ldots, Z_{n,k})'$, gives the asymptotic distribution of our ECDF-based test statistics under H_0 and $H_{1,n}$.

Corollary 5. *Suppose that Assumption A holds. Then, under $H_{1,n}$,*

$$
\begin{aligned}
T^1_{KS} &\xrightarrow{\mathcal{L}} \frac{1}{2}\sum_{j=1}^{k} |Z_j + p_j^{1/2}\Delta_j| \sup_t |t f_{\varepsilon_j}(t)|,\\
T^1_{CM} &\xrightarrow{\mathcal{L}} \frac{1}{4}\sum_{j=1}^{k} (Z_j + p_j^{1/2}\Delta_j)^2 \int t^2 f^2_{\varepsilon_j}(t) dF_{\varepsilon_j}(t),\\
T^2_{KS} &\xrightarrow{\mathcal{L}} \frac{1}{2}\sup_y |Z(t) + \Delta(t)|,\\
T^2_{CM} &\xrightarrow{\mathcal{L}} \frac{1}{4}\int \{Z(t) + \Delta(t)\}^2 dF_{\varepsilon}(t),
\end{aligned}
$$

where $Z(t) = \sum_{j=1}^k p_j^{1/2} t f_{\varepsilon_j}(t) Z_j$ and $\Delta(t) = \sum_{j=1}^k p_j t f_{\varepsilon_j}(t)\Delta_j$, with $(Z_1, \ldots, Z_k)' \sim N_k(0, \Sigma)$, $\Sigma = (\sigma_{jv})$ being the $k \times k$-matrix whose elements are

$$
\begin{aligned}
\sigma_{jv} \quad = \quad & (p_j p_v)^{1/2} \sum_{l=1}^{k} \frac{E\{(\varepsilon_l^2 - 1)^2\}}{p_l} \times \qquad (6)\\
& \times E\left[\left\{\pi_l(X_l)\frac{f_j(X_l)}{f_l(X_l)} - I(l = j)\right\}\left\{\pi_l(X_l)\frac{f_v(X_l)}{f_l(X_l)} - I(l = v)\right\}\right]
\end{aligned}
$$

$1 \le j, v \le k$.

Let T denote any of the test statistics T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM}. Since H_0 can be seen as a special case of $H_{1,n}$ with $\delta_j = 0$, $1 \le j \le k$, the asymptotic distribution of T under the null hypothesis trivially follows by setting $\Delta_j = 0$. For example, under H_0,

$$
T^1_{KS} \xrightarrow{\mathcal{L}} \frac{1}{2}\sum_{j=1}^{k} |Z_j| \sup_t |t f_{\varepsilon_j}(t)|, \text{ and } T^2_{KS} \xrightarrow{\mathcal{L}} \frac{1}{2}\sup_t |Z(t)|.
$$

Let $\alpha \in (0, 1)$ be arbitrary but fixed. As an immediate consequence of Theorem 1 and Corollaries 3 and 5, the test that rejects H_0 when $T \ge t_\alpha$, where t_α is the $1 - \alpha$ percentile of the null distribution of T or any consistent estimator of it, is consistent against all fixed alternatives. It is also able to detect local alternatives converging to the null at the rate $n^{-1/2}$, whenever $\Delta_j \ne 0$ for some $1 \le j \le k$.

So far we have assumed that the weight functions $\pi_1, \ldots, \pi_k$ are known. Nevertheless we did not make any restriction on them except the fact that they are positive and sum up to one. In our simulation study, see Section 6, we take $\pi_j = n_j/n$. This simple choice is shown to work reasonably well for all the investigated examples. Another possibility is to choose the π_j's from the data. For example, as for the problem of testing the equality of regression curves, see

[25, 24], one may take $\pi_j(x) = p_j f_j(x)/f_{mix}(x)$, with $f_{mix}(x) = \sum_{j=1}^{k} p_j f_j(x)$. For this choice, since $f_1, \dots f_k$ are unknown, the functions $\pi_1, \dots, \pi_k$ must be estimated. A careful reading of the proofs reveals that all the results in this paper continue to be true whenever $\pi_1, \dots, \pi_k$ are replaced by estimators $\hat{\pi}_1, \dots, \hat{\pi}_k$ satisfying $\sup_{x \in R} |\pi_j(x) - \hat{\pi}_j(x)| = o_p(n^{-1/4})$, $1 \le j \le k$.

4. Asymptotics for ECF-based test statistics

In order to study the limit behaviour of the test statistics T_1 and T_2 we also need some regularity conditions. Recall that to derive the asymptotic properties for the ECDF-based test statistics we assumed that the regression errors have a twice differentiable CDF. Analogously, to derive the asymptotic properties for the ECF-based test statistics we need that the regression errors has a twice differentiable CF, which is tantamount to assume that the regression errors has finite second order moment. But this assumption is implicit in the the definition of the regression models (1). As a consequence, the assumptions required to derive the asymptotics for ECF-based test statistics will be weaker than those assumed in Section 3, in the sense that no restriction on the distribution of the errors will be imposed, such as the existence of a density. Specifically, we mainly need to assume that Assumption (A1) holds. The motivation behind the test statistics T_1 and T_2 is in the following result.

Theorem 6. *Suppose that Assumption (A1) holds and that $w \ge 0$ is such that $\int t^2 w(t)dt < \infty$. Then, $n^{-1}T_i = \tau_i + o_p(1)$, $i = 1, 2$, where*

$$\tau_1 = \sum_{j=1}^{k} p_j \int \left|\varphi_{\varepsilon_j}(t) - \varphi_{\varepsilon_{0j}}(t)\right|^2 w(t)dt, \qquad \tau_2 = \int \left|\varphi_{\varepsilon}(t) - \varphi_{\varepsilon_0}(t)\right|^2 w(t)dt.$$

Thus, T_1 and T_2 converge in probability to non-negative quantities. Since two distinct CFs can be equal in a finite interval (see, for example, [10], p. 479), a general way to ensure that $\tau_1 > 0$ and $\tau_2 > 0$ whenever $\sigma_r \ne \sigma_s$, for some $1 \le r, s \le k$, $r \ne s$, is to take $w(t) > 0$, for all $t \in \mathbb{R}$. For instance, one can take w as the pdf of a normal law. Now, the reasoning made just after Corollary 3 can be repeated for the test statistics T_1 and T_2. So our next goal is to determine the asymptotic distribution of T_1 and T_2. With this aim we first give a result that provides an asymptotic approximation for $\sqrt{n_j}\{\hat{\varphi}_{\varepsilon_j}(t) - \hat{\varphi}_{\varepsilon_{0j}}(t)\}$, $1 \le j \le k$. Let $\varphi'_{\varepsilon_j}(t) = \frac{\partial}{\partial t}\varphi_{\varepsilon_j}(t) = \frac{\partial}{\partial t}Re\varphi_{\varepsilon_j}(t) + \mathrm{i}\frac{\partial}{\partial t}Im\varphi_{\varepsilon_j}(t) = \mathrm{i}E\left[\varepsilon_j \exp(\mathrm{i}t\varepsilon_j)\right]$, which exists because $E(|\varepsilon_j|) < \infty$, $1 \le j \le k$.

Theorem 7. *Suppose that Assumption (A1) holds. Then, under $H_{1,n}$,*

$$\sqrt{n_j}\left\{\hat{\varphi}_{\varepsilon_j}(t) - \hat{\varphi}_{\varepsilon_{0j}}(t)\right\} = \frac{1}{2}t\varphi'_{\varepsilon_j}(t)(p_j^{1/2}\Delta_j - Z_{n,j}) + tR_{1j}(t) + t^2R_{2j}(t),$$

where $\sup_t |R_{sj}(t)| = o_p(1)$, $s = 1, 2$, and $Z_{n,j}$ and Δ_j, $1 \le j \le k$, are defined as in Theorem 4.

Corollary 8. *Suppose that Assumption (A1) holds and that $w \geq 0$ is such that $\int t^4 w(t)dt < \infty$. Then, under $H_{1,n}$,*

$$T_1 \xrightarrow{\mathcal{L}} \frac{1}{4}\sum_{j=1}^{k}(Z_j + p_j^{1/2}\Delta_j)^2 \int t^2 |\varphi'_{\varepsilon_j}(t)|^2 w(t)dt,$$

$$T_2 \xrightarrow{\mathcal{L}} \frac{1}{4}\int |V(t) + W(t)|^2 w(t)dt,$$

where $V(t) = \sum_{j=1}^{k} p_j^{1/2} t\varphi'_{\varepsilon_j}(t) Z_j$, $W(t) = \sum_{j=1}^{k} p_j t\varphi'_{\varepsilon_j}(t)\Delta_j$, and $(Z_1, \dots, Z_k)' \sim N_k(0, \Sigma)$.

Similar comments to those made after Corollary 5 for the test statistics T^1_{KS}, T^1_{CM}, T^2_{KS} and T^2_{CM} can be done for T_1 and T_2.

Before ending this section, we give a brief discussion on the choice the weight function w. It has been seen that taking $w > 0$ ensures that the test that rejects H_0 for large values of T_1 or T_2 is consistent against any fixed alternative. It also ensures that T_1 converges in law, under H_0, to a non-degenerate distribution (see Section 5). From a theoretical point of view, any positive function w satisfying $\int t^4 w(t)dt < \infty$ can be used. From a practical point of view, the ease of computation of T_1 and T_2 is closely related to the choice of w. In fact, an alternative and more useful expression for T_1 and T_2 is given by (see Lemma 1 in [2])

$$T_1 = \sum_{j=1}^{k}\frac{1}{n_j}\left\{\sum_{l,s=1}^{n_j} I_w(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{js}) + \sum_{l,s=1}^{n_j} I_w(\hat{\varepsilon}_{0jl} - \hat{\varepsilon}_{0js}) - 2\sum_{l,s=1}^{n_j} I_w(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{0js})\right\},$$

$$T_2 = \frac{1}{n}\sum_{j,v=1}^{k}\sum_{l=1}^{n_j}\sum_{s=1}^{n_v}\left\{I_w(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{vs}) + I_w(\hat{\varepsilon}_{0jl} - \hat{\varepsilon}_{0vs}) - 2I_w(\hat{\varepsilon}_{jl} - \hat{\varepsilon}_{0vs})\right\},$$

where

$$I_w(t) = \int \cos(tx)w(x)dx. \tag{7}$$

These expressions are specially appealing when one wishes to employ the bootstrap to approximate the null distribution, which requires to evaluate the test statistic in a high number of artificial samples. Another point that should be taken into account is the fact that the ECF estimates more accurately the population CF around $t = 0$. Consequently, w should put most of the weight near the origin. For the problem of testing the equality of mean regression curves, [24] take w to be the standard normal density. We also considered this choice for w in our simulation study.

5. Estimation of the null distribution

The results in Corollaries 5 and 8 reveal that the asymptotic null distributions of the proposed test statistics are in all cases unknown because they depend on unknown quantities. Therefore, the asymptotic null distribution cannot be directly used to approximate the null distribution of these statistics. Two solutions can be considered: (a) approximate the null distributions by a bootstrap procedure, or (b) construct an approximation of the asymptotic null distribution. The first approach was also considered in [25] for the problem of testing the equality of conditional mean functions. They employed a bootstrap procedure based on smoothed residuals, whose theoretical justification can be found in [21]. The same bootstrap procedure could be used to approximate the null distribution of the test statistics studied in this paper.

The second possibility is to approximate the null distribution by means of an estimator of the asymptotic null distribution of the test statistic. This estimator is usually called a bootstrap-in-the-limit estimator. Let us first consider the test statistic T^1_{CM}. According to Corollary 5, under H_0, $4T^1_{CM} \xrightarrow{\mathcal{L}} W_1 := \sum_{j=1}^k \alpha_j \chi^2_{1,j}$, where $\chi^2_{1,1}, \ldots, \chi^2_{1,k}$ are independent chi-square random variables with one degree of freedom and $\alpha_1, \ldots, \alpha_k$ are the eigenvalues of $\mathcal{A}\Sigma$, $\mathcal{A} = diag(a_1, \ldots, a_k)$, $a_j = \int t^2 f^2_{\varepsilon_j}(t) dF_{\varepsilon_j}(t)$, $1 \le j \le k$. Before employing a bootstrap-in-the-limit estimator we must be sure that the asymptotic null distribution is non-degenerate. Since, under our assumptions, $a_j > 0$ and $\sigma_{jj} > 0$, $1 \le j \le k$, we have that $\sum_{j=1}^k \alpha_j = \text{trace}(\mathcal{A}\Sigma) = \sum_{j=1}^k a_j \sigma_{jj} > 0$, and therefore its asymptotic null distribution is non-degenerate. The quantities α_j in W_1 are unknown but can be estimated consistently from the data, say by $\hat{\alpha}_j$, the eigenvalues of $\hat{\mathcal{A}}\hat{\Sigma}$, using a plug-in principle and kernel smoothing methods. In such a case,

$$\sup_t \left| P_0\{T^1_{CM} \le t\} - P_*(\hat{W}_1 \le t) \right| \xrightarrow{P} 0,$$

where $\hat{W}_1 = \sum_{j=1}^k \hat{\alpha}_j \chi^2_{1j}$. Analogously, one could also estimate the null distribution of T^2_{CM}, T_1 and T_2.

As for T_1, Corollary 8 says that $4T_1$ converges in law to $W_2 := \sum_{j=1}^k \beta_j \chi^2_{1,j}$, where $\chi^2_{1,1}, \ldots, \chi^2_{1,k}$ are as before and $\beta_1, \ldots, \beta_k$ are the eigenvalues of $\mathcal{B}\Sigma$, $\mathcal{B} = diag(b_1, \ldots, b_k)$, with $b_j = \int t^2 |\varphi'_{\varepsilon_j}(t)|^2 w(t) dt$, $1 \le j \le k$. Since $\sigma_{jj} > 0$, $1 \le j \le k$, and $\sum_{j=1}^k \beta_j = \sum_{j=1}^k b_j \sigma_{jj}$, to ensure that W_2 is non-degenerate we must have that $b_j > 0$ for some $1 \le j \le k$. Since $E(\varepsilon_j) = \varphi'_{\varepsilon_j}(0) = 0$ and $E(\varepsilon_j^2) = -\varphi''_{\varepsilon_j}(0) = 1$, where $\varphi''_{\varepsilon_j}(t) = \frac{\partial^2}{\partial t^2} Re\varphi_{\varepsilon_j}(t) + \mathrm{i}\frac{\partial^2}{\partial t^2} Im\varphi_{\varepsilon_j}(t)$, it readily follows that $|\varphi'_{\varepsilon_j}(t)| > 0$, for all $t \in (-\delta, 0) \cup (0, \delta)$, for some $\delta > 0$. Thus, if the weight function w is positive in an open neighborhood of the origin, we have that $b_j > 0$, $1 \le j \le k$, which implies that W_2 is non-degenerate.

Interestingly, if all the covariates have the same distribution, $f_1 = \cdots = f_k$, $E[(\varepsilon_1^2 - 1)^2] = \cdots = E[(\varepsilon_k^2 - 1)^2] := \theta$ and $\pi_j(x) = p_j$, $1 \le j \le k$, then

$$\Sigma = \theta(I_k - pp'), \quad p' = (\sqrt{p_1}, \ldots, \sqrt{p_k}). \tag{8}$$

It is easy to see that the matrix $I_k - pp'$ has two different eigenvalues: 0, with multiplicity 1, and 1, with multiplicity $k-1$. Therefore, if the laws of the errors also satisfy $a_1 = \cdots = a_k$ (for instance, if all errors have the same distribution), then $4(\theta a_1)^{-1} T^1_{CM} \xrightarrow{\mathcal{L}} \chi^2_{k-1}$, which coincides with the null distribution of the classical Levene's test for equality of variances in two or more groups. To get a consistent null distribution estimator of T^1_{CM} in this case, it suffices to estimate θ and a_1 consistently, which is a quite easy task. The same is true for T_1 if $b_1 = \cdots = b_k$.

The asymptotic behaviours of T^2_{CM} and T_2 is somewhat different. From Corollary 5, $4T^2_{CM}$ converges in law to $W_3 := \sum_{j=1}^{k} \gamma_j \chi^2_{1,j}$, where $\chi^2_{1,1}, \ldots, \chi^2_{1,k}$ are as before and $\gamma_1, \ldots, \gamma_k$ are the eigenvalues of $\mathcal{C}\Sigma$, $\mathcal{C} = diag(p)\mathcal{M}diag(p)$, with p as defined in (8) and $\mathcal{M} = (m_{rs})$, $m_{rs} = \sum_{v=1}^{k} p_v \int t^2 f_{\varepsilon_r}(t) f_{\varepsilon_s}(t) f_{\varepsilon_v}(t)dt$, $1 \le r, s \le k$. Note that if Σ is as in (8), since $(I_k - pp')p = 0$ we have that trace$(\mathcal{C}\Sigma) = 0$, and thus $W_3 = 0$. That is to say that, in this case, the asymptotic null distribution of T^2_{CM} is degenerate. The same happens to T_2. Since in practice Σ is unknown, in order to estimate the null distribution of T^2_{CM} and T_2 it is preferable to use the bootstrap procedure mentioned in the first paragraph of this section.

6. Finite sample performance

This section is devoted to the study of the practical performance of the proposed test statistics in terms of level approximation and power. With that purpose, we consider the following variance models in a two-population ($k = 2$) framework:

(L1) $\sigma_1^2(x) = \sigma_2^2(x) = 0.25$
(L2) $\sigma_1^2(x) = \sigma_2^2(x) = \left(\frac{7}{6}0.50x + \frac{1}{2}0.50\right)^2$
(P1) $\sigma_1^2(x) = 0.25$; $\sigma_2^2(x) = 0.50$
(P2) $\sigma_1^2(x) = 0.25$; $\sigma_2^2(x) = 0.75$
(P3) $\sigma_1^2(x) = 0.25$; $\sigma_2^2(x) = \left(\frac{7}{8}\sqrt{0.50}x + \frac{1}{2}\sqrt{0.50}\right)^2$
(P4) $\sigma_1^2(x) = \left(\frac{7}{6}0.50x + \frac{1}{2}0.50\right)^2$; $\sigma_2^2(x) = \left(\frac{7}{8}\sqrt{0.50}x + \frac{1}{2}\sqrt{0.50}\right)^2$

In all cases the regression functions are $m_1(x) = m_2(x) = x$. The distributions of the covariates X_1 and X_2 are $Beta(1.5, 2)$ and $Beta(2, 1.5)$, respectively, and the regression errors ε_1 and ε_2 are $N(0, 1)$. Models (L1) and (L2) are under the null hypothesis, so they will be used to study the level approximation. On the other hand, the power will be investigated through models (P1), (P2), (P3) and (P4). The variance functions are chosen to make the homoscedastic models (P1 and P2) and the heteroscedastic models (P3 and P4) somehow comparable: in the four models $E[\sigma_1(X_1)] = 0.50$, while and $E[\sigma_2(X_2)] = \sqrt{0.50}$ for models (P1), (P3) and (P4), and $E[\sigma_2(X_2)] = \sqrt{0.75}$ for model (P2).

The weight function w required to construct the ECF-based test statistics is the density of a standard normal and we choose $\pi_j = n_j/n$, $j = 1, 2$. The tables below display the observed proportion of rejections in 1000 simulated data sets with significance level $\alpha = 0.05$ (other significance levels were also considered and similar results were obtained).

Nonparametric estimation of the regression functions is performed by local-linear estimation, while the estimation of the conditional variance functions is done with the local-constant (Nadaraya-Watson) estimator, as it guarantees the positiveness of the estimation. The application of these smoothing techniques requires the specification of a kernel function and a smoothing pa-

Table 1
Observed rejection frequencies in 1000 simulated data sets for the tests based on the critical values obtained from the asymptotic null distribution of T^1_{CM} and T_1

model	(n_1, n_2)	$h:$	cv	0.10	0.20	0.30	cv	0.10	0.20	0.30
			T^1_{CM}				T_1			
(L1)	$(100, 100)$		0.065	0.069	0.061	0.071	0.046	0.046	0.043	0.046
	$(200, 100)$		0.036	0.037	0.039	0.036	0.047	0.048	0.050	0.049
	$(200, 200)$		0.033	0.033	0.038	0.034	0.050	0.052	0.051	0.052
(L2)	$(100, 100)$		0.102	0.093	0.092	0.092	0.074	0.070	0.075	0.076
	$(200, 100)$		0.040	0.041	0.043	0.046	0.061	0.062	0.063	0.065
	$(200, 200)$		0.056	0.056	0.054	0.055	0.067	0.069	0.070	0.072
(P1)	$(100, 100)$		0.909	0.907	0.902	0.908	0.905	0.907	0.901	0.900
	$(200, 100)$		0.927	0.926	0.923	0.924	0.960	0.957	0.955	0.957
	$(200, 200)$		0.992	0.992	0.992	0.991	0.993	0.993	0.993	0.993
(P2)	$(100, 100)$		0.994	0.991	0.987	0.988	0.999	0.997	0.995	0.995
	$(200, 100)$		0.994	0.992	0.991	0.990	0.995	0.995	0.994	0.994
	$(200, 200)$		0.999	0.997	0.997	0.997	1.000	0.998	0.998	0.997
(P3)	$(100, 100)$		0.776	0.775	0.773	0.778	0.768	0.766	0.770	0.771
	$(200, 100)$		0.783	0.781	0.775	0.780	0.888	0.883	0.880	0.882
	$(200, 200)$		0.934	0.937	0.931	0.932	0.964	0.965	0.964	0.964
(P4)	$(100, 100)$		0.676	0.673	0.664	0.669	0.636	0.631	0.635	0.635
	$(200, 100)$		0.650	0.645	0.636	0.640	0.691	0.681	0.684	0.689
	$(200, 200)$		0.852	0.852	0.854	0.854	0.887	0.887	0.884	0.884

rameter or bandwidth. For the kernel, we choose the kernel of Epanechnikov $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$. On the other hand, the optimal choice of the smoothing parameter in testing frameworks is not a solved problem (see, for example, the discussion about this topic in [12]). To study the impact of the smoothing parameters in our tests, we will show results obtained under fixed values and also for values obtained by cross-validation. From some unreported simulations, we have learned that taking the same bandwidth in all populations is recommended. In the case of the cross-validation (indicated by cv in the tables), the regular least-squares method was applied to find the smoothing parameters to estimate σ_j^2, $j = 1, 2$, and then the average of the two obtained quantities is used to perform the estimation. A similar procedure is used to obtain the cross-validation bandwidth to estimate the regression functions m_j. On the other hand, in the case of fixed bandwidths we take values 0.1, 0.2 and 0.3 (recall that the support of the covariates is [0,1]) to estimate both the regression and the variance functions.

We first study the behaviour of the tests based on the approximation of the asymptotic null distribution. We will only study the tests based on T^1_{CM} and T_1 because, as explained in Section 5, the asymptotic null distribution of these statistics is a non-degenerate combination of chi-square random variables. Since we are dealing with approximations based on asymptotics, we consider moderate sample sizes (100 and 200). The obtained results are displayed in Table 1. In terms of level approximation, the behaviour of both statistics is reasonable for model (L1). For model (L2), the level is clearly overestimated for samples sizes (100,100), specially in the case of T^1_{CM}. The approximation improves as the sample sizes increase. In terms of power, both statistics present a similar behaviour. The choice of the smoothing parameter does not seems to have an important impact, neither in the approximation of the level, nor in the values of the power.

Another possibility to obtain critical values is by means of bootstrap. In particular, in the current setup, as in other related papers (see for example [25, 6]) a smoothed bootstrap of residuals is recommended. We have applied this bootstrap mechanism with 200 bootstrap replications to the six test statistics proposed in Section 2. Tables 2 and 3 display the observed rejection probabilities for the ECDF-based tests and for the ECF-based tests, respectively. In this case, smaller sample sizes (50 and 100) are employed. The approximation of the level (models L1 and L2) is good for the tests statistics based on L_2-distances T^1_{CM}, T^2_{CM}, T_1 and T_2. On the other hand, the Kolmogorov-Smirnov type statistics are a bit conservative, specially T^2_{KS}. Regarding the power, T^1_{CM}, T^1_{KS} and T_1 achieve better results than T^2_{CM}, T^2_{KS} and T_2, respectively. Globally, the tests based on L_2-distances (for example, T^1_{CM} and T_1) produce very similar results, and they outperform the Kolmogorov-Smirnov-type statistics. As before, the choice of the smoothing parameters does not have much impact on the rejection frequencies.

The models considered so far for the regression functions, the conditional variances and the distributions of the errors are somewhat simple. One may wonder if the proposed procedures still work for more complicated models. With this purpose, we consider the regression functions $m_1(x) = x + \sin(2\pi x)$ and $m_2(x) = \sin(2\pi x)$ and two new models for the conditional variance functions:

(L3) $\sigma_1^2(x) = \sigma_1^2(x) = 0.5^2 \exp(2x)$.
(P5) $\sigma_1^2(x) = 0.5^2 \exp(2x)$ and $\sigma_2^2(x) = 0.7^2 \exp(2x)$.

Model (L3) is under the null hypothesis and model (P5) is under the alternative. For the regression errors, in addition to the normal distribution, we also consider laws with heavier tails, $\sqrt{5/3}\,\varepsilon_j \sim t_5$ and $\sqrt{7/5}\,\varepsilon_j \sim t_7$, and an asymmetric distribution, $\varepsilon_j + 1 \sim Exponential(1)$. From the results in the previous tables, we have learned that the tests based on the statistics T^1_{CM} and T_1 exhibit the best results in terms of power. Moreover, the critical values for these two statistics can be approximated from their asymptotic null distributions. Because of these reasons, we have conducted a similar study to that in Table 1. For the new models described above, T_1 exhibits better behavior than T^1_{CM} in most cases. Table 4 reports the obtained results, which, for the sake of brevity are restricted to T_1 with models (L2), (L3), (P4) and (P5) and cross-validation bandwidths. For non-normal errors, larger sample sizes are required in order to get a reasonable approximation of the nominal level. The distribution of the errors also affects the power, in the sense that lower values are observed for heavy-tailed and asymmetric laws when compared to the normal case.

Table 2
Observed rejection frequencies in 1000 simulated data sets for the tests based on the test statistics T^1_{CM}, T^2_{CM}, T^1_{KS} and T^2_{KS}. The critical values obtained by bootstrap

model	(n_1, n_2)	$h:$ cv	0.10	0.20	0.30	cv	0.10	0.20	0.30
		T^1_{CM}				T^2_{CM}			
(L1)	$(50, 50)$	0.045	0.047	0.053	0.048	0.044	0.037	0.050	0.053
	$(100, 50)$	0.038	0.036	0.041	0.036	0.049	0.051	0.039	0.046
	$(100, 100)$	0.044	0.043	0.039	0.048	0.045	0.046	0.047	0.047
(L2)	$(50, 50)$	0.072	0.068	0.070	0.070	0.058	0.049	0.057	0.063
	$(100, 50)$	0.040	0.044	0.045	0.045	0.049	0.046	0.052	0.058
	$(100, 100)$	0.058	0.049	0.053	0.058	0.061	0.059	0.070	0.066
(P1)	$(50, 50)$	0.549	0.530	0.531	0.536	0.383	0.372	0.388	0.368
	$(100, 50)$	0.647	0.657	0.634	0.641	0.573	0.569	0.563	0.568
	$(100, 100)$	0.885	0.889	0.874	0.884	0.687	0.691	0.690	0.686
(P2)	$(50, 50)$	0.905	0.906	0.911	0.902	0.769	0.771	0.783	0.763
	$(100, 50)$	0.951	0.954	0.947	0.944	0.900	0.896	0.893	0.895
	$(100, 100)$	0.996	0.995	0.995	0.993	0.974	0.976	0.969	0.973
(P3)	$(50, 50)$	0.395	0.397	0.397	0.402	0.289	0.291	0.297	0.287
	$(100, 50)$	0.503	0.504	0.501	0.496	0.393	0.417	0.413	0.412
	$(100, 100)$	0.705	0.713	0.709	0.707	0.521	0.510	0.517	0.512
(P4)	$(50, 50)$	0.326	0.310	0.319	0.308	0.216	0.225	0.221	0.221
	$(100, 50)$	0.356	0.359	0.355	0.356	0.318	0.312	0.332	0.336
	$(100, 100)$	0.555	0.569	0.574	0.567	0.406	0.394	0.394	0.402
		T^1_{KS}				T^2_{KS}			
(L1)	$(50, 50)$	0.032	0.034	0.034	0.037	0.025	0.028	0.026	0.028
	$(100, 50)$	0.029	0.038	0.035	0.029	0.035	0.034	0.029	0.020
	$(100, 100)$	0.034	0.040	0.040	0.038	0.025	0.028	0.022	0.026
(L2)	$(50, 50)$	0.050	0.047	0.051	0.049	0.029	0.028	0.032	0.032
	$(100, 50)$	0.038	0.040	0.035	0.043	0.033	0.033	0.029	0.037
	$(100, 100)$	0.054	0.045	0.042	0.050	0.046	0.033	0.037	0.048
(P1)	$(50, 50)$	0.409	0.409	0.401	0.397	0.210	0.183	0.208	0.208
	$(100, 50)$	0.586	0.593	0.593	0.598	0.360	0.358	0.362	0.372
	$(100, 100)$	0.802	0.812	0.800	0.809	0.457	0.431	0.452	0.469
(P2)	$(50, 50)$	0.788	0.806	0.795	0.784	0.497	0.481	0.515	0.506
	$(100, 50)$	0.927	0.934	0.920	0.915	0.754	0.756	0.732	0.732
	$(100, 100)$	0.992	0.988	0.988	0.992	0.862	0.871	0.869	0.873
(P3)	$(50, 50)$	0.296	0.290	0.301	0.289	0.155	0.127	0.145	0.146
	$(100, 50)$	0.454	0.444	0.449	0.433	0.216	0.215	0.212	0.225
	$(100, 100)$	0.631	0.634	0.630	0.609	0.269	0.254	0.280	0.262
(P4)	$(50, 50)$	0.224	0.239	0.224	0.220	0.099	0.104	0.104	0.101
	$(100, 50)$	0.313	0.319	0.326	0.314	0.173	0.183	0.183	0.177
	$(100, 100)$	0.464	0.486	0.456	0.470	0.226	0.258	0.240	0.243

Table 3

Observed rejection frequencies in 1000 simulated data sets for the tests based on the test statistics T_1 and T_2. The critical values obtained by bootstrap

model	(n_1, n_2)	$h:$	cv	0.10	0.20	0.30	cv	0.10	0.20	0.30
			T_1				T_2			
(L1)	$(50, 50)$		0.046	0.046	0.052	0.045	0.040	0.051	0.045	0.047
	$(100, 50)$		0.035	0.032	0.035	0.031	0.041	0.044	0.041	0.041
	$(100, 100)$		0.040	0.038	0.040	0.041	0.048	0.053	0.046	0.052
(L2)	$(50, 50)$		0.070	0.067	0.076	0.069	0.055	0.061	0.056	0.058
	$(100, 50)$		0.043	0.045	0.043	0.041	0.040	0.041	0.034	0.037
	$(100, 100)$		0.058	0.056	0.052	0.060	0.063	0.063	0.057	0.059
(P1)	$(50, 50)$		0.562	0.566	0.553	0.556	0.361	0.358	0.371	0.367
	$(100, 50)$		0.693	0.686	0.679	0.679	0.466	0.480	0.488	0.490
	$(100, 100)$		0.895	0.896	0.887	0.894	0.543	0.541	0.546	0.547
(P2)	$(50, 50)$		0.928	0.926	0.926	0.917	0.726	0.716	0.738	0.738
	$(100, 50)$		0.968	0.971	0.961	0.956	0.850	0.857	0.869	0.859
	$(100, 100)$		0.997	0.997	0.995	0.993	0.937	0.936	0.935	0.938
(P3)	$(50, 50)$		0.427	0.431	0.424	0.440	0.298	0.294	0.300	0.301
	$(100, 50)$		0.581	0.580	0.565	0.566	0.367	0.356	0.376	0.381
	$(100, 100)$		0.743	0.737	0.744	0.742	0.432	0.421	0.412	0.400
(P4)	$(50, 50)$		0.328	0.326	0.331	0.319	0.194	0.192	0.206	0.203
	$(100, 50)$		0.381	0.380	0.383	0.383	0.238	0.240	0.249	0.248
	$(100, 100)$		0.573	0.574	0.581	0.572	0.306	0.308	0.309	0.313

Table 4

Observed rejection frequencies in 1000 simulated data sets for the tests based on the critical values obtained from the asymptotic null distribution of T_1 for several error distributions and cross-validation bandwidths

model	(n_1, n_2)	errors:	$N(0,1)$	$t_5/\sqrt{5/3}$	$t_7/\sqrt{7/5}$	$Exp(1)-1$
(L2)	$(100, 100)$		0.065	0.065	0.063	0.082
	$(200, 100)$		0.052	0.086	0.083	0.088
	$(200, 200)$		0.051	0.067	0.070	0.061
	$(400, 200)$		0.050	0.083	0.070	0.045
	$(400, 400)$		0.041	0.064	0.046	0.051
(L3)	$(100, 100)$		0.100	0.077	0.066	0.088
	$(200, 100)$		0.062	0.082	0.086	0.090
	$(200, 200)$		0.066	0.073	0.073	0.070
	$(400, 200)$		0.055	0.077	0.074	0.041
	$(400, 400)$		0.051	0.065	0.051	0.063
(P4)	$(100, 100)$		0.494	0.374	0.378	0.264
	$(200, 100)$		0.542	0.348	0.390	0.289
	$(200, 200)$		0.796	0.532	0.614	0.391
	$(400, 200)$		0.859	0.582	0.664	0.432
	$(400, 400)$		0.968	0.776	0.876	0.596
(P5)	$(100, 100)$		0.917	0.669	0.729	0.518
	$(200, 100)$		0.936	0.740	0.831	0.587
	$(200, 200)$		0.992	0.872	0.946	0.755
	$(400, 200)$		0.997	0.927	0.975	0.847
	$(400, 400)$		0.999	0.966	0.996	0.939

Table 5
p-values for testing for the equality of the conditional variance functions for the data set concerning expenditures of Dutch households

	asymptotic		bootstrap					
h	T^1_{CM}	T_1	T^1_{CM}	T^2_{CM}	T^1_{KS}	T^2_{KS}	T_1	T_2
0.20	0.011	0.395	0.207	0.101	0.123	0.135	0.428	0.735
0.25	0.028	0.380	0.321	0.296	0.378	0.350	0.387	0.688
0.30	0.028	0.368	0.349	0.269	0.386	0.302	0.387	0.615
0.35	0.045	0.384	0.417	0.280	0.344	0.263	0.377	0.509
0.40	0.049	0.435	0.443	0.274	0.450	0.247	0.435	0.470
0.45	0.078	0.509	0.588	0.247	0.439	0.241	0.536	0.435
0.50	0.083	0.597	0.602	0.239	0.477	0.237	0.647	0.419

7. Application to data

To illustrate our testing procedure we will use a data set concerning monthly expenditures of Dutch households. The variable 'log of the total monthly expenditure' is considered as a covariate and 'log of the expenditure on food' is considered as the response. See [7, 25] for more details on these data. In the latter paper, the equality of the regression curves of households of 2, 3 and 4 members was tested and the equality between the regression curves of 3-member households (43 observations) and 4-member households (73 observations) was accepted. Here we move one step further in the comparison of the regression models and test for the equality of the conditional variance functions. Table 5 shows the p-values obtained from the asymptotic null distribution for T_1 and T^1_{CM} and from the bootstrap for the six test statistics with fixed bandwidths ranging from 0.20 and 0.50 (the support of the covariates is approximately between 9.5 and 11.5). The results are quite homogeneous, as all test statistics, except the asymptotic version of T^1_{CM}, lead to the acceptance of the equality of the conditional variance functions. As we have seen in the simulations presented in Section 6, the approximation of the asymptotic null distribution of the T^1_{CM} is not satisfactory, specially for small sample sizes. Since here we are working with 43 and 73 observations, the results for this test statistic are not reliable, and we should only consider its bootstrap version.

8. Conclusions

In this paper, we have constructed and studied six tests for the equality of k conditional variances. To do so, we compare the ECDF and ECF of the error terms estimated nonparametrically under H_0 and H_1. Under some regularity conditions, the proposed tests are consistent against any fixed alternative and are able to detect contiguous alternatives converging to the null at a rate $n^{-1/2}$. The assumptions needed to derive these properties are weaker for the ECF-based test statistics. Specifically, no requirement is imposed on the distributions of the errors. An approximation of the asymptotic null distribution has been proposed and the performance of each test has been evaluated by means of some simulations. The proposed approximation works, in the sense of providing type I errors close to the nominal values, specially when the sample sizes are large (at least 200). For smaller sample sizes it is recommended to approximate the null distribution through a bootstrap mechanism.

The estimation of the conditional variance functions has been also studied in the econometric literature when the data present dependence structure (see for example [9, 28, 20]). The proposed tests could be extended to this setting by assuming mixing conditions on the data and using appropriate results for the nonparametric estimators for the variance and regression functions in the same line of [6], who tested for a constant variation coefficient in regression models with stationary data and used the results about kernel estimators with dependent data given in [13].

Although the results in this paper are presented for local-constant and local-linear weights variance estimators, in practice, it is well-known that the local-linear variance estimator may take negative values. In our simulations and application to real data we used the local-constant estimator to estimate the conditional variance functions, as it guarantees the positiveness of the estimate. There are other possibilities to obtain positive estimators, such us, for instance, the local-exponential estimator studied by [28]. Under suitably adapted conditions, the procedures and the results in this paper can be extended for the local-exponential and other conditional variance estimators.

The above extensions, as well as others motivated by recent applications (for instance, in casual inference, [14]) constitute fields of future research.

Appendix

We now sketch the proofs of the results stated in Sections 1–4. With this aim we first give some preliminary results, some of them are of independent interest.

A.1. Preliminary results

Under Assumption (A1), and consequently under Assumption A, we have that, for $1 \leq j \leq k$, $\sup_{x\in R} |\hat{m}_j(x) - m_j(x)| = o_p(n_j^{-1/4})$, $\sup_{x\in R} |\hat{\sigma}_j(x) - \sigma_j(x)| = o_p(n_j^{-1/4})$, and $\sup_{x\in R} |\hat{f}_j(x) - f_j(x)| = o_p(n_j^{-1/4})$. This, together with some routine calculations, show that

$$\sup_{x\in R} \left| \hat{\sigma}_j^2(x) - \sigma_j^2(x) - \frac{1}{n_j f_j(x)} \sum_{s=1}^{n_j} K_h(X_{js} - x) \left[\{Y_{js} - m_j(x)\}^2 - \sigma_j^2(x) \right] \right|$$
$$= o_p(n^{-1/2}). \tag{9}$$

Also, from the equality

$$\hat{\sigma}_j(x) - \sigma_j(x) = \frac{\hat{\sigma}_j^2(x) - \sigma_j^2(x)}{2\sigma_j(x)} - \frac{\{\hat{\sigma}_j(x) - \sigma_j(x)\}^2}{2\sigma_j(x)},$$

it follows that,

$$\sup_{x\in R}\left|\hat{\sigma}_j(x)-\sigma_j(x)-\frac{\hat{\sigma}_j^2(x)-\sigma_j^2(x)}{2\sigma_j(x)}\right| = o_p(n^{-1/2}). \tag{10}$$

So, by the definition of $\hat{\sigma}_0(x)$ and $\sigma_0(x)$, we also have that $\sup_{x\in R}|\hat{\sigma}_0(x)-\sigma_0(x)| = o_p(n^{-1/4})$ and

$$\sup_{x\in R}\left|\hat{\sigma}_0(x)-\sigma_0(x)-\sum_{j=1}^{k}\pi_j(x)\frac{\hat{\sigma}_j^2(x)-\sigma_j^2(x)}{2\sigma_0(x)}\right| = o_p(n^{-1/2}). \tag{11}$$

Lemma 9. *Suppose that Assumption (A1) holds. Then,*

(i) *For* $1 \le j \le k$,

$$\int \frac{\hat{\sigma}_j(x)-\sigma_j(x)}{\sigma_j(x)} f_j(x)dx = \frac{1}{2n_j}\sum_{s=1}^{n_j}\left(\varepsilon_{js}^2-1\right) + o_p(n^{-1/2}).$$

(ii)

$$\begin{aligned}&\int \frac{\hat{\sigma}_0(x)-\sigma_0(x)}{\sigma_0(x)} f_j(x)dx\\ =\quad & \sum_{v=1}^{k}\frac{1}{2n_v}\sum_{s=1}^{n_v}\pi_v(X_{vs})\frac{f_j(X_{vs})}{f_v(X_{vs})}\frac{\sigma_v^2(X_{vs})}{\sigma_0^2(X_{vs})}\left(\varepsilon_{vs}^2-1\right) + o_p(n^{-1/2}).\end{aligned}$$

Proof. From (9) and (10), we get

$$\begin{aligned}&\int \frac{\hat{\sigma}_j(x)-\sigma_j(x)}{\sigma_0(x)} f_j(x)dx\\ =\quad & \frac{1}{2n_j}\sum_{s=1}^{n_j}\int K_h(X_{js}-x)\left[\{Y_{js}-m_j(x)\}^2/\sigma_j^2(x)-1\right]dx + o_p(n^{-1/2}).\end{aligned}$$

Part (i) follows from the above equality by making the change of variable $U_{js} = (X_{js}-x)/h$ and applying a Taylor's development. Part (ii) can be proved similarly by using (9) and (11). □

Lemma 10. *Let* $\tilde{\varphi}_{\varepsilon_j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{jl})$, $\hat{\varphi}_{\varepsilon_j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\hat{\varepsilon}_{jl})$, *and similarly define* $\tilde{\varphi}_{\varepsilon_{0j}}(t)$ *and* $\hat{\varphi}_{\varepsilon_{0j}}(t)$. *Suppose that Assumption (A1) holds. Then, for* $1 \le j \le k$,

(i)

$$\hat{\varphi}_{\varepsilon_j}(t) \quad = \quad \tilde{\varphi}_{\varepsilon_j}(t) + \mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{jl})\frac{m_j(X_{jl})-\hat{m}_j(X_{jl})}{\sigma_j(X_{jl})}$$

$$+\,\mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{jl})\frac{\sigma_j(X_{jl})-\hat\sigma_j(X_{jl})}{\sigma_j(X_{jl})}\varepsilon_{jl}$$
$$+\,tR_{j,1}(t)+t^2R_{j,2}(t),$$

with $\sup_t|R_{j,s}(t)|=o_p(n^{-1/2}),\ s=1,2.$

(ii)

$$\begin{aligned}\hat\varphi_{\varepsilon_{0j}}(t) &= \tilde\varphi_{\varepsilon_{0j}}(t)+\mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{0jl})\frac{m_j(X_{jl})-\hat m_j(X_{jl})}{\sigma_0(X_{jl})}\\ &\quad+\mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{0jl})\frac{\sigma_0(X_{jl})-\hat\sigma_0(X_{jl})}{\sigma_0(X_{jl})}\varepsilon_{0jl}\\ &\quad+tR_{0j,1}(t)+t^2R_{0j,2}(t),\end{aligned}$$

with $\sup_t|R_{0j,s}(t)|=o_p(n^{-1/2}),\ s=1,2.$

Proof. Using a Taylor's development, we get

$$\hat\varphi_{\varepsilon_j}(t)-\tilde\varphi_{\varepsilon_j}(t)=\mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}(\hat\varepsilon_{jl}-\varepsilon_{jl})\exp(\mathrm{i}t\varepsilon_{jl})+t^2R_j(t)\frac{1}{n_j}\sum_{l=1}^{n_j}(\hat\varepsilon_{jl}-\varepsilon_{jl})^2,$$

with $\sup_t|R_j(t)|=O_p(1)$. Part (i) follows from the equality

$$\begin{aligned}\hat\varepsilon_j-\varepsilon_j &= \frac{m_j(X_j)-\hat m_j(X_j)}{\hat\sigma_j(X_j)}+\frac{\sigma_j(X_j)-\hat\sigma_j(X_j)}{\hat\sigma_j(X_j)}\varepsilon_j\\ &= \frac{m_j(X_j)-\hat m_j(X_j)}{\sigma_j(X_j)}+\frac{\{m_j(X_j)-\hat m_j(X_j)\}\{\sigma_j(X_j)-\hat\sigma_j(X_j)\}}{\sigma_j(X_j)\hat\sigma_j(X_j)}\\ &\quad+\frac{\sigma_j(X_j)-\hat\sigma_j(X_j)}{\sigma_j(X_j)}\varepsilon_j+\frac{\{\sigma_j(X_j)-\hat\sigma_j(X_j)\}^2}{\sigma(X_j)\hat\sigma_j(X_j)}\varepsilon_j.\end{aligned}$$

Similarly, one can prove (ii). □

Lemma 11. *Let g be a bounded function. Suppose Assumption (A1) holds. Then,*

$$\begin{aligned}&\frac{\mathrm{i}t}{\sqrt{n_j}}\sum_{l=1}^{n_j}\varepsilon_{jl}\exp(\mathrm{i}t\varepsilon_{jl})g(X_{jl})\frac{\hat\sigma_v(X_{jl})-\sigma_v(X_{jl})}{\sigma_v(X_{jl})}\\ =\ &\frac{t}{2}\varphi'_{\varepsilon_j}(t)\frac{\sqrt{n}_j}{n_v}\sum_{s=1}^{n_v}(\varepsilon_{vs}^2-1)g(X_{vs})\frac{f_j(X_{vs})}{f_v(X_{vs})}+tR_{j,v}(t),\end{aligned}$$

with $\sup_t|R_{j,v}(t)|=o_p(1)$, $1\le j,v\le k$.

Proof. From (9) and (10),

$$\frac{\mathrm{i}t}{\sqrt{n_j}}\sum_{l=1}^{n_j}\exp(\mathrm{i}t\varepsilon_{jl})\varepsilon_{jl}g(X_{jl})\frac{\hat\sigma_v(X_{jl})-\sigma_v(X_{jl})}{\sigma_v(X_{jl})}=\frac{\mathrm{i}t}{2}\sqrt{n_j}H_{jv}+tR_1(t),$$

where $\sup_t |R_1(t)| = o_p(1)$ and

$$H_{jv}(t) = \frac{1}{n_j n_v} \sum_{l=1}^{n_j} \sum_{s=1}^{n_v} U_v(X_{jl}, \varepsilon_{jl}; X_{vs}, \varepsilon_{vs}; t),$$

with

$$U_v(X_1, \varepsilon_1; X_2, \varepsilon_2; t)$$
$$= \varepsilon_1 \exp(\mathrm{i}t\varepsilon_1) \frac{g(X_1)}{f_v(X_1)} K_h(X_1 - X_2) \left[\frac{\{m_v(X_2) + \varepsilon_2 \sigma_v(X_2) - m_v(X_1)\}^2}{\sigma_v^2(X_1)} - 1 \right].$$

• If $j \neq v$, then, for every t, $H_{jv}(t)$ is a two sample U-statistic of degree $(1, 1)$ with kernel $U_v(X_{jl}, \varepsilon_{jl}; X_{vs}, \varepsilon_{vs}; t)$. Its Hájek projection, $H'_{jv}(t)$, is given by

$$H'_{jv}(t) = -\mathrm{i}\varphi'_{\varepsilon_j}(t) \frac{1}{n_v} \sum_{s=1}^{n_v} (\varepsilon_{vs}^2 - 1) g(X_{vs}) \frac{f_j(X_{vs})}{f_v(X_{vs})} + R'_{jv}(t)$$

where $\sup_t |R'_{jv}(t)| = O_p(h^2)$. Moreover,

$$Var\{H_{jv}(t) - H'_{jv}(t)\} \leq \frac{1}{n_j n_v} E\{U_h^2(X_j, \varepsilon_j; X_v, \varepsilon_v; t)\} = O(n_j^{-1} n_v^{-1} h^{-1}).$$

Therefore,

$$\sqrt{n_j} H_{jv}(t) = -\mathrm{i}\varphi'_{\varepsilon_j}(t) \frac{\sqrt{n_j}}{n_v} \sum_{s=1}^{n_v} (\varepsilon_{vs}^2 - 1) g(X_{vs}) \frac{f_j(X_{vs})}{f_v(X_{vs})} + R_{jv}(t), \tag{12}$$

with $\sup_t |R_{jv}(t)| = o_p(1)$.
• If $j = v$, then

$$H_{jj}(t) = \frac{K(0)}{n_j^2 h} \sum_{l=1}^{n_j} \varepsilon_{jl} \exp(\mathrm{i}t\varepsilon_{jl})(\varepsilon_{jl}^2 - 1) \frac{g(X_{jl})}{f_j(X_{jl})} + \frac{n_j - 1}{2n_j} H_j(t),$$

where, for every t, $H_j(t)$ is a one sample U-statistic of degree 2 with kernel $U_j(X_{jl}, \varepsilon_{jl}; X_{js}, \varepsilon_{js}; t) + U_j(X_{js}, \varepsilon_{js}; X_{jl}, \varepsilon_{jl}; t)$. Arguments very similar to those employed for the case $j \neq v$ can be used to show that

$$\sqrt{n_j} H_j(t) = -2\mathrm{i}\varphi'_{\varepsilon_j}(t) \frac{1}{\sqrt{n_j}} \sum_{s=1}^{n_j} (\varepsilon_{js}^2 - 1) g(X_{js}) + R_j(t),$$

with $\sup_t |R_j(t)| = o_p(1)$. Since

$$\sqrt{n_j} \frac{K(0)}{n_j^2 h} \left| \sum_{l=1}^{n_j} \varepsilon_{jl} \exp(\mathrm{i}t\varepsilon_{jl})(\varepsilon_{jl}^2 - 1) \frac{g(X_{jl})}{f_j(X_{jl})} \right| \leq \frac{M}{\sqrt{nh^2}} \frac{1}{n_j} \sum_{l=1}^{n_j} |\varepsilon_{jl}|^3,$$

for some positive constant M, we conclude that $H_{jj}(t)$ also satisfies (12) with $j = v$. This proves the result. □

A.2. Proofs of main results

Proof of Theorem 1. *(i)* The direct implication is trivial. To prove the converse implication, assume that ε_{0j} and ε_j have the same distribution. They will also share the same moments. Now, because of the independence of ε_j and X_j, we get

$$E(\varepsilon_{0j}^2) = E(\varepsilon_j^2) \Rightarrow E\{\sigma_j^2(X_j)/\sigma_0^2(X_j)\} = 1, \text{ and}$$
$$E(\varepsilon_{0j}^4) = E(\varepsilon_j^4) \Rightarrow E\{\sigma_j^4(X_j)/\sigma_0^4(X_j)\} = 1.$$

Hence, $E[(\frac{\sigma_j^2(X_j)}{\sigma_0^2(X_j)} - 1)^2] = 0$, for $1 \le j \le k$, and so we deduce that H_0 holds.

(ii) Let ε (respectively, ε_0) be a random variable with CDF F_ε (respectively, F_{ε_0}). Note that ε (respectively, ε_0) is the mixture of the random variables $\{\varepsilon_j, 1 \le j \le k\}$ (respectively, $\{\varepsilon_{j0}, 1 \le j \le k\}$) with probabilities $\{p_j, 1 \le j \le k\}$. As for part *(i)*, using the fact that $E(\varepsilon_j^2) = 1$ and $E(\varepsilon_j^4) = E(\varepsilon_1^4) > 0$, for $1 \le j \le k$, we obtain

$$E(\varepsilon_0^2) = E(\varepsilon^2) \Rightarrow \sum_{j=1}^{k} p_j E\{\sigma_j^2(X_j)/\sigma_0^2(X_j)\} = 1, \text{ and}$$
$$E(\varepsilon_0^4) = E(\varepsilon^4) \Rightarrow \sum_{j=1}^{k} p_j E\{\sigma_j^4(X_j)/\sigma_0^4(X_j)\} = 1.$$

Hence, $\sum_{j=1}^k p_j E[(\frac{\sigma_j^2(X_j)}{\sigma_0^2(X_j)} - 1)^2] = 0$. Since $p_j > 0$, we conclude that H_0 is true. □

Proof of Theorem 2. From the proof of Theorem 1 in [1],

$$\begin{aligned}\hat{F}_{\varepsilon_{0j}}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} I(\varepsilon_{0jl} \le t) + t f_{\varepsilon_{0j}}(t) \int \frac{\hat{\sigma}_0(x) - \sigma_0(x)}{\sigma_0(x)} f_j(x)dx \\ + f_{\varepsilon_{0j}}(t) \int \frac{\hat{m}_j(x) - m_j(x)}{\sigma_0(x)} f_j(x)dx + o_p(n^{-1/2}),\end{aligned} \tag{13}$$

and

$$\begin{aligned}\hat{F}_{\varepsilon_j}(t) = \frac{1}{n_j}\sum_{l=1}^{n_j} I(\varepsilon_{jl} \le t) + t f_{\varepsilon_j}(t) \int \frac{\hat{\sigma}_j(x) - \sigma_j(x)}{\sigma_j(x)} f_j(x)dx \\ + f_{\varepsilon_j}(t) \int \frac{\hat{m}_j(x) - m_j(x)}{\sigma_j(x)} f_j(x)dx + o_p(n^{-1/2}),\end{aligned} \tag{14}$$

uniformly in t, where $f_{\varepsilon_{0j}}$ denotes the density corresponding to $F_{\varepsilon_{0j}}$. The desired results follow directly from (13) and (14). □

Proof of Theorem 4. From the proof of Lemma 1 in [1], we have that

$$\frac{1}{n_j}\sum_{l=1}^{n_j} I(\varepsilon_{jl} \le t) = \frac{1}{n_j}\sum_{l=1}^{n_j} I(\varepsilon_{0jl} \le t) + F_{\varepsilon_j}(t) - F_{\varepsilon_{0j}}(t) + o_p(n^{-1/2}), \tag{15}$$

uniformly in t. Observe that

$$\frac{\sigma_0(x)}{\sigma_j(x)} = 1 - n^{-1/2}\frac{\sigma_0(x)}{\sigma_j(x)}\delta_j(x) = 1 - n^{-1/2}\delta_j(x) + n^{-1}\frac{\sigma_0(x)}{\sigma_j(x)}\delta_j^2(x).$$

Using this and a Taylor's development leads to

$$F_{\varepsilon_{0j}}(t) = E\left[F_{\varepsilon_j}\left\{t\frac{\sigma_0(X_j)}{\sigma_j(X_j)}\right\}\right] = F_{\varepsilon_j}(t) - n^{-1/2}tf_{\varepsilon_j}(t)E\left[\delta_j(X_j)\right] + o(n^{-1/2}), \tag{16}$$

uniformly in t,

$$\sup_t |f_{\varepsilon_j}(t) - f_{\varepsilon_{0j}}(t)| = O(n^{-1/2}), \text{ and } \sup_t |tf_{\varepsilon_j}(t) - tf_{\varepsilon_{0j}}(t)| = o(1). \tag{17}$$

From (13)–(17), after some routine calculations, using the fact that $\sigma_j(X_j)/\sigma_0(X_j) = 1 + n^{-1/2}\delta_j(X_j)$, we obtain that,

$$\sqrt{n_j}\left\{\hat{F}_{\varepsilon_j}(t) - \hat{F}_{\varepsilon_{0j}}(t)\right\} = p_j^{1/2}tf_{\varepsilon_j}(t)E\left[\delta_j(X_j)\right] + \frac{t}{2}f_{\varepsilon_j}(t)Z_{n,j} + o_p(1),$$

uniformly in t, where $Z_{n,j}$ is defined in (5). □

Proof of Theorem 6. First observe that

$$\begin{aligned} & \hat{\varphi}_{\varepsilon_j}(t) - \hat{\varphi}_{\varepsilon_{0j}}(t) \\ = \; & \left\{\hat{\varphi}_{\varepsilon_j}(t) - \tilde{\varphi}_{\varepsilon_j}(t)\right\} - \left\{\hat{\varphi}_{\varepsilon_{0j}}(t) - \tilde{\varphi}_{\varepsilon_{0j}}(t)\right\} + \left\{\tilde{\varphi}_{\varepsilon_j}(t) - \tilde{\varphi}_{\varepsilon_{0j}}(t)\right\}. \end{aligned} \tag{18}$$

We have that

$$\int |\hat{\varphi}_{\varepsilon_j}(t) - \tilde{\varphi}_{\varepsilon_j}(t)|^2 w(t) \leq \frac{2}{n_j}\sum_l (\hat{\varepsilon}_{jl} - \varepsilon_{jl})^2 \int t^2 w(t)dt = o_p(1),$$

and, similarly, $\int |\hat{\varphi}_{\varepsilon_{0j}}(t) - \tilde{\varphi}_{\varepsilon_{0j}}(t)|^2 w(t) = o_p(1)$. On the other hand,

$$\begin{aligned} & \int \left|\tilde{\varphi}_{\varepsilon_j}(t) - \tilde{\varphi}_{\varepsilon_{0j}}(t)\right|^2 w(t)dt \\ = \; & \frac{1}{n_j^2}\sum_{r,s=1}^{n_j}\left\{I_w(\varepsilon_{jr} - \varepsilon_{js}) + I_w(\varepsilon_{0jr} - \varepsilon_{0js}) - 2I_w(\varepsilon_{jr} - \varepsilon_{0js})\right\}, \end{aligned}$$

with I_w as defined in (7), is a V-statistic of degree 2 with a bounded kernel and thus (see [26]) it converges to its expected value $\int \left|\varphi_{\varepsilon_j}(t) - \varphi_{\varepsilon_{0j}}(t)\right|^2 w(t)dt$. We conclude that $\frac{1}{n}T_1 \xrightarrow{p} \tau_1$. The limit of $\frac{1}{n}T_2$ can be similarly derived. □

Proof of Theorem 7. Using a Taylor's development, we get

$$\tilde{\varphi}_{\varepsilon_{0j}}(t) - \tilde{\varphi}_{\varepsilon_j}(t) = \mathrm{i}\frac{t}{n_j}\sum_{l=1}^{n_j}(\varepsilon_{0jl} - \varepsilon_{jl})\exp(\mathrm{i}t\varepsilon_{jl}) + t^2 R_{0j}(t)\frac{1}{n_j}\sum_{l=1}^{n_j}(\varepsilon_{0jl} - \varepsilon_{jl})^2,$$

with $\sup_t |R_{0j}(t)| = O_p(1)$. This, together with the fact that

$$\varepsilon_{0j} - \varepsilon_j = \left\{ \frac{\sigma_j(X_j)}{\sigma_0(X_j)} - 1 \right\} \varepsilon_j = n^{-1/2} \delta_j(X_j) \varepsilon_j, \qquad (19)$$

leads to

$$\sqrt{n_j} \left\{ \tilde{\varphi}_{\varepsilon_{0j}}(t) - \tilde{\varphi}_{\varepsilon_j}(t) \right\} = p_j^{1/2} t \varphi_{\varepsilon_j}^{'}(t) E\left[\delta_j(X_j)\right] + t R_{0j}^{(1)}(t) + t^2 R_{0j}^{(2)}(t), \qquad (20)$$

with $\sup_t |R_{0j}^{(s)}(t)| = o_p(1), s = 1, 2$.

Now using (11), (10), (19) and Lemmas 11 and 10, we obtain that

$$\begin{aligned}
& \left\{ \hat{\varphi}_{\varepsilon_j}(t) - \tilde{\varphi}_{\varepsilon_j}(t) \right\} - \left\{ \hat{\varphi}_{\varepsilon_{0j}}(t) - \tilde{\varphi}_{\varepsilon_{0j}}(t) \right\} \\
= \; & \mathrm{i} \frac{t}{n_j} \sum_{l=1}^{n_j} \exp(\mathrm{i} t \varepsilon_{jl}) \frac{\hat{m}_j(X_{jl}) - m_j(X_{jl})}{\sigma_j(X_{jl})} \left\{ \frac{\sigma_j(X_{jl})}{\sigma_0(X_{jl})} - 1 \right\} \\
& + \frac{t}{2} \varphi_{\varepsilon_j}'(t) \sum_{v=1}^{k} \frac{1}{n_v} \sum_{s=1}^{n_v} (\varepsilon_{vs}^2 - 1) \pi_v(X_{vs}) \frac{f_j(X_{vs})}{f_v(X_{vs})} \left\{ \frac{\sigma_j^2(X_{vs})}{\sigma_0^2(X_{vs})} - 1 \right\} \\
& - \frac{t}{2} \varphi_{\varepsilon_j}'(t) \frac{1}{\sqrt{n_j}} Z_{n,j} + t R_{0j,1}^{(2)}(t) + t^2 R_{0j,2}^{(2)}(t) \\
= \; & -\frac{t}{2} \varphi_{\varepsilon_j}'(t) \frac{1}{\sqrt{n_j}} Z_{n,j} + t R_{0j,1}^{(2)}(t) + t^2 R_{0j,2}^{(2)}(t), \qquad (21)
\end{aligned}$$

where $Z_{n,j}$ is given by (5) and $\sup_t |R_{0j,s}^{(2)}(t)| = o_p(n^{-1/2})$, $s = 1, 2$.

Combining (18), (20) and (21), we conclude that

$$\sqrt{n_j} \left\{ \hat{\varphi}_{\varepsilon_j}(t) - \hat{\varphi}_{\varepsilon_{0j}}(t) \right\} = p_j^{1/2} t \varphi_{\varepsilon_j}^{'}(t) E\left[\delta_j(X_j)\right] - \frac{t}{2} \varphi_{\varepsilon_j}'(t) Z_{n,j} + t R_{1j}(t) + t^2 R_{2j}(t),$$

with $\sup_t |R_{sj}(t)| = o_p(1)$, $s = 1, 2$. □

Acknowledgements

The authors thank the Editor and an anonymous referee for their comments and suggestions.

References

[1] Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **28** 549–567. MR1858417

[2] Alba-Fernández, V., Jiménez-Gamero, M. D. and Muñoz-García, J. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics & Data Analysis*, **52** 3730–3748. MR2427377

[3] Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statistics and Probability Letters*, **17** 199–204. MR1229937

[4] Dette, H. and Marchlewski, M. (2010). A robust test for homoscedasticity in nonparametric regression. *Journal of Nonparametric Statistics*, **22** 723–736. MR2682218

[5] Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in non-parametric regression. *Journal of the Royal Statistical Society, Series B*, **69** 903–917. MR2368576

[6] Dette, H., Pardo-Fernández, J. C. and Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics*, **36** 782–799. MR2573308

[7] Einmahl, J. and Van Keilegom, I. (2008). Specification tests in nonparametric regression. *Journal of Econometrics*, **143** 88–102. MR2384434

[8] Fan, J. and Gijbels. I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London. MR1383587

[9] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85** 645–660. MR1665822

[10] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Vol. 2*. Wiley, New York. MR0270403

[11] Gastwirth, J. L., Gel, Y. and Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, **24** 343–360. MR2757435

[12] González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of Goodness-of-Fit tests for regression models. *Test*, **22** 361–411. MR3093195

[13] Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24** 726–748. MR2409261

[14] Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142** 615–635. MR2416821

[15] Koul, H. L. and Song, W. (2010). Conditional variance model checking. *Journal of Statistical Planning and Inference*, **140** 1056–1072. MR2574668

[16] Kulasekera, K. B. (1995). Comparison of regression curves using quasi-residuals. *Journal of the American Statistical Association*, **90** 1085–1093. MR1354025

[17] Levene, H. (1960). Robust tests for equality of variances. In *Contributions to Probability and Statistics* (I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow and H. B. Mann, editors), 278–292. Stanford University Press, Stanford. MR0120709

[18] Liero, H. (2003). Testing homoscedasticity in nonparametric regression. *Journal of Nonparametric Statistics*, **15** 31–51. MR1958958

[19] Mathur, S. K. and Dolo, S. (2008). A new efficient statistical test for detecting variability in the gene expression data. *Statistical Methods in Medical Research*, **17** 405–419. MR2526661

[20] Mishra, S., Su, L. and Ullah, A. (2010). Semiparametric estimator of time series conditional variance. *Journal Business and Economic Statistics*, **28** 256–274. MR2681200

[21] Neumeyer, N. (2009). Smooth residual bootstrap for empirical processes of non-parametric regression residuals. *Scandinavian Journal of Statistics*, **36** 204–228. MR2528982

[22] Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves: An empirical process approach. *The Annals of Statistics*, **31** 880–920. MR1994734

[23] Pardo-Fernández, J. C. (2007). Comparison of error distributions in nonparametric regression. *Statistics and Probability Letters*, **77** 350–356. MR2339039

[24] Pardo-Fernández, J. C., Jiménez-Gamero, M. D. and El Ghouch, A. (2015). A nonparametric ANOVA-type test for regression curves based of characteristic functions. *Scandinavian Journal of Statistics*, **42** 197–213. MR3318032

[25] Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of k regression curves. *Statistica Sinica*, **17** 1115–1137. MR2397389

[26] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[27] Srihera, R. and Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis*, **101** 2039–2059. MR2671200

[28] Ziegelmann, F. A. (2002). Nonparametric estimation of volatility functions: The local exponential estimator. *Econometric Theory*, **18** 985–991. MR1918331