Lithuanian Mathematical Journal, *Vol.* 47, *No.* 3, 2007

NEW TESTS OF HETEROSKEDASTICITY IN LINEAR REGRESSION MODEL

A. Račkauskas

Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania
(e-mail: alfredas.rackauskas@mif.vu.lt)

D. Zuokas

Vilnius University, Naugarduko 24, LT-03225 Vilnius, Lithuania
(e-mail: danas.zuokas@mif.vu.lt)

Abstract. In this paper, we present a class of tests for heteroskedasticity of various types in the linear regression model. These tests are based on the limit behavior of the polygonal process constructed from squared residuals. The law of test statistics under the null hypothesis is established, and the consistency is proved. By means of simulations these tests are compared with two classical tests (likelihood-ratio and Breusch–Pagan) for two types of heteroskedasticity (changed-segment and the type where the error variance is proportional to one of the components of the design matrix).

Keywords: heteroskedasticity, FCLT, changed segment.

INTRODUCTION

Classical results in the econometric theory show that if coefficients of the linear regression model are estimated by the OLS method when error term is not homoskedastic, one obtains an estimator which, although remains unbiased, consistent, and asymptotically normal, is no longer a minimum-variance unbiased estimator. Therefore, for example, one cannot construct reliable confidence intervals. Thus, diagnostic testing for heteroskedasticity has to be undertaken before further analysis starts. For an informative discussion on some aspects of tests as well as for the state-of-the-art of the heteroskedasticity problem, we refer to two recent and extensive surveys by Dufour *et al.* [3] and Godfrey *et al.* [5].

In this paper, we propose a new class of tests for testing the null hypothesis of equal error variances against a wide range of heteroskedasticity alternatives. These tests are based on determining the largest weighted difference in variance of estimated error, which is found by taking all possible pairs of complementary subsets of indices. The typical example of test statistics is

$$\max_{\ell} q(n,\ell) \max_{k} \left| \frac{1}{\ell} \sum_{j\in[k+1,k+\ell]} \hat{\varepsilon}_j^{\,2} - \frac{1}{n-\ell} \sum_{j\notin[k+1,k+\ell]} \hat{\varepsilon}_j^{\,2} \right|,$$

Published in Lietuvos Matematikos Rinkinys, Vol. 47, No. 3, pp. 307–327, July–September, 2007. Original article submitted March 30, 2007.

where $\hat{\varepsilon}_k$, $k = 1, 2, \dots, n$, are residuals, and q is a certain weight function controlling the "intensity" of heteroskedasticity, which is to be detected.

To establish the limit behavior of test statistics, we preliminarily investigate the empirical polygonal process constructed from the partial sums of squared residuals. Although usual functional framework for such processes are classical function spaces $C[0, 1]$ or $D[0, 1]$, we work in a stronger topology of Hölder function spaces.

The paper is organized as follows. In Section 1, we consider a model under the null hypothesis of a constant error variance (i.e., homoskedasticity) and state limit theorems for the polygonal process in Hölder spaces. In Section 2, we introduce a new class of tests of Hölder norm type. First, using the results of Section 1, we find the limit law of test statistics under the null hypothesis of homoskedasticity. For these test statistics, we then prove the consistency under various alternatives of heteroskedasticity. Section 3 supports theoretical results by simulations. The proofs of the main results are given in the Appendix.

1. FCLT FOR SQUARED RESIDUALS

The null hypothesis H_0 specifies the regression model

$$y_j = \boldsymbol{f}^{\mathrm{T}}(j/n)\boldsymbol{\beta} + \varepsilon_j, \quad j = 1, \dots, n, \tag{1}$$

where $\boldsymbol{\beta}$ is an unknown coefficients vector of length d, $\boldsymbol{f}\colon [0, 1] \to \mathbb{R}^d$ is a given function,

$$\boldsymbol{f}(t) = \bigl(f_1(t), \dots, f_d(t)\bigr)^{\mathrm{T}}, \quad t \in [0, 1],$$

and the disturbances $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. zero-mean random variables with constant variance σ^2. Hereafter, $^{\mathrm{T}}$ denotes the transpose operation.

Set the $n \times d$ design matrix

$$\boldsymbol{X} = \bigl(\boldsymbol{f}(1/n), \boldsymbol{f}(2/n), \dots, \boldsymbol{f}(n/n)\bigr)^{\mathrm{T}}.$$

Throughout, we assume the following regularity condition:

$$\begin{cases} \text{the matrices} \quad \boldsymbol{A}_n = n^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \quad \text{and} \\ \boldsymbol{A} = \int_0^1 \boldsymbol{f}(t)\boldsymbol{f}^{\mathrm{T}}(t)\,\mathrm{d}t \quad \text{are nondegenerate.} \end{cases} \tag{2}$$

Regression coefficients are estimated using the OLS method,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} = \left(\sum_{j=1}^{n} \boldsymbol{f}(j/n)\boldsymbol{f}^{\mathrm{T}}(j/n)\right)^{-1} \sum_{j=1}^{n} \boldsymbol{f}(j/n) y_j,$$

where $\boldsymbol{y} = (y_1, \dots, y_n)^{\mathrm{T}}$, and the residual vector $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)^{\mathrm{T}}$ is defined by

$$\hat{\varepsilon}_j = y_j - \hat{y}_j = y_j - \boldsymbol{f}^{\mathrm{T}}(j/n)\hat{\boldsymbol{\beta}}, \quad j = 1, \dots, n.$$

Next, we consider the sequence of partial sums of squared residuals, which is defined by $\widehat{V}_0^2 = 0$ and

$$\widehat{V}_k^2 = \sum_{j=1}^{k} \left[\hat{\varepsilon}_j^{\,2} - \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2}\right], \quad k = 1, \dots, n.$$

We introduce the sequence of stochastic processes $\{\hat{\nu}_n(t), t \in [0, 1]\}$, $n = 1, 2, \dots,$ by

$$\hat{\nu}_n(t) = \widehat{V}_{[nt]}^2 + (nt - [nt])\bigl(\widehat{V}_{[nt]+1}^2 - \widehat{V}_{[nt]}^2\bigr), \quad t \in [0, 1],$$

that possess continuous sample paths.

Any kind of heteroskedasticity in error term of model (1) has an impact on these processes, so it is important to establish their limiting behavior in as strong as possible functional framework. The processes $\hat{\nu}_n$ are polygonal lines with vertices at $(j/n, \widehat{V}_j^2)$, $j = 0, \dots, n$, and so they posses more smoothness than simple continuity. Therefore, it is natural to study them in a topology of Hölder spaces which is stronger than that of the usual spaces $L_2(0,1)$, $D[0,1]$, or $C[0,1]$. As a consequence, within this topological framework, one can consider more continuous functionals on $\hat{\nu}_n$. This gives more flexibility building tests for heteroskedasticity.

Let us firstly define the Hölder spaces. For $0 < \alpha < 1$ and a function x: $[0,1] \to \mathbb{R}$, set

$$\omega_\alpha(x,\delta) := \sup_{\substack{s,t\in[0,1]\\ 0<|t-s|<\delta}} \frac{|x(t)-x(s)|}{|t-s|^\alpha}.$$

We associate to α the separable Hölder space of continuous functions

$$H_\alpha^0[0,1] := \Bigl\{ x \in C[0,1];\ \lim_{\delta\to 0} \omega_\alpha(x,\delta) = 0 \Bigr\}$$

equipped with the norm

$$\|x\|_\alpha := |x(0)| + \omega_\alpha(x,1).$$

Next, we introduce an admissible class of functions $\boldsymbol{f}$ for model (1). Let $|\cdot|$ be any fixed norm in $\mathbb{R}^d$ and $p \geqslant 1$. The function $\boldsymbol{g}$: $[0,1] \to \mathbb{R}^d$ is said to have a finite p-variation if

$$v_p(\boldsymbol{g}) := \sup_{(t_k)\subset[0,1]} \sum_{k=1}^m \bigl|\boldsymbol{g}(t_k) - \boldsymbol{g}(t_{k-1})\bigr|^p < \infty,$$

where the upper bound is taken over all possible partitions $0 = t_0 < t_1 < \cdots < t_{m-1} < t_m = 1$ of the interval $[0,1]$. The set of bounded functions $\boldsymbol{g}$: $[0,1] \to \mathbb{R}^d$ of finite p-variation is denoted by $V_p^d[0,1]$ and is equipped with the norm

$$\|\boldsymbol{g}\|_{[p]} = \|\boldsymbol{g}\|_\infty + v_p^{1/p}(\boldsymbol{g}),$$

where $\|\boldsymbol{g}\|_\infty = \sup_{0\leqslant t\leqslant 1} |\boldsymbol{g}(t)|$. In our model, we assume that $\boldsymbol{f} \in V_p^d[0,1]$ for some $1 \leqslant p < 2$. For continuously differentiable $\boldsymbol{f}$: $[0,1] \to \mathbb{R}^d$, this condition is automatically satisfied with $p = 1$. Note also that if the function $\boldsymbol{f}$: $[0,1] \to \mathbb{R}^d$ satisfies

$$\|\boldsymbol{f}\|_\infty + \omega_\alpha(\boldsymbol{f},1) < \infty$$

for some $0 < \alpha < 1$, then $\boldsymbol{f} \in V_p^d[0,1]$ for all $p \geqslant 1/\alpha$.

Before stating the main results in this section, consider the processes $\nu_n = \{\nu_n(t), t \in [0,1]\}$ defined in the same way as $\hat{\nu}_n$ but with disturbances $\{\varepsilon_j, j = 1, \dots, n\}$ instead of $\{\hat{\varepsilon}_j, j = 1, \dots, n\}$. The limiting behavior of ν_n in $H_\alpha^0[0,1]$ can be studied following Račkauskas and Suquet [9]. In this paper, we prove a theorem that shows a similar limit behavior of ν_n and $\hat{\nu}_n$ in the Hölder spaces $H_\alpha^0[0,1]$. Using this result, we then state theorems that establish the convergence in distribution of $\hat{\nu}_n$. The proofs are given in the Appendix.

THEOREM 1. *Let $0 < \alpha < 1/2$ and $1 \leqslant p < 1/(1-\alpha)$. For model* (1)*, assume that the function $\boldsymbol{f}$ is continuous and has finite p-variation. In addition, let condition* (2) *be satisfied. Then*

$$\|\hat{\nu}_n - \nu_n\|_\alpha = o_P(n^{1/2}), \tag{3}$$

provided that the following condition holds:

$$\sup_{t>0} t^{2/(1/2-\alpha)} \mathrm{P}\bigl(|\varepsilon_1| > t\bigr) < \infty. \tag{4}$$

Let $(W_t, t \geqslant 0)$ denote a standard Wiener process, and let $(B_t, t \in [0,1])$ be the corresponding Brownian bridge $B_t = W_t - tW_1$, $t \in [0,1]$.

THEOREM 2. *Under the conditions of Theorem* 1,

$$n^{-1/2}\hat{\delta}_n^{-1}\hat{\nu}_n \xrightarrow[n\to\infty]{\mathcal{D}} B \quad \textit{in the space} \quad H_\alpha^0[0,1] \tag{5}$$

if and only if

$$\lim_{t\to\infty} t^{2/(1/2-\alpha)}\mathrm{P}\big(|\varepsilon_1| \geqslant t\big) = 0. \tag{6}$$

Here $\hat{\delta}_n^2$ is the sample variance of $\hat{\boldsymbol{\varepsilon}}$,

$$\hat{\delta}_n^2 = n^{-1}\sum_{j=1}^{n}\left[\hat{\varepsilon}_j^{\,2} - n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^{\,2}\right]^2. \tag{7}$$

Since the embedding $H_\alpha^0 \hookrightarrow C[0,1]$ is continuous for each $0 < \alpha < 1$, as a corollary to Theorem 2, we obtain the following result.

COROLLARY 3. *For model* (1)*, assume that the function* $\boldsymbol{f}$ *is continuous and has finite* p*-variation for some* $1 \leqslant p < 2$*. Then*

$$n^{-1/2}\hat{\delta}_n^{-1}\hat{\nu}_n \xrightarrow[n\to\infty]{\mathcal{D}} B \quad \textit{in the space} \quad C[0,1], \tag{8}$$

provided that $\mathrm{E}|\varepsilon_1|^q < \infty$ *for some* $q > 4p/(2-p)$.

2. A CLASS OF TEST STATISTICS

We can employ the results of the previous section and the continuous mapping theorem to construct various test statistics. Theorem 1, together with Theorem 2 and Corollary 3, also provide us with limit laws for these statistics under the null hypothesis.

PROPOSITION 4. *The convergence* (5) (*respectively* (8)) *yields*

$$F\big(n^{-1/2}\hat{\delta}_n^{-1}\hat{\nu}_n\big) \xrightarrow[n\to\infty]{\mathcal{D}} F(B) \tag{9}$$

for each continuous functional $F\colon H_\alpha^0[0,1] \to \mathbb{R}$ (*respectively* $F\colon C[0,1] \to \mathbb{R}$).

The following continuous functionals are of interest:

$$F_1(x) = \sup_{0<h<1} h^{-\alpha} \sup_{0\leqslant t<1-h} \big|x(t+h) - x(t)\big|, \quad 0 \leqslant \alpha < 1;$$

$$F_2(x) = \int_0^1 \frac{1}{u}\left[\int_{|t-s|\leqslant u} \frac{|x(t)-x(s)|}{|t-s|^\alpha}\,\mathrm{d}t\,\mathrm{d}s\right]\mathrm{d}u, \quad 0 \leqslant \alpha < 1;$$

$$F_3(x) = \max_{j\geqslant 1}\sum_{r\in D_j}\big|x(r+2^{-j}) - x(r)\big|^p, \quad 1 < p < \infty,$$

where D_j is the set of dyadic numbers of the jth level. Each of these functionals gives a certain class (depending on a parameter) of test statistics. In this paper, we investigate in detail the test statistics corresponding to F_1. Precisely,

$$T_{n,\alpha} := \max_{1\leqslant \ell<n} (\ell/n)^{-\alpha} \max_{0\leqslant k\leqslant n-\ell} \left| \sum_{j=k+1}^{k+\ell} \left[\hat{\varepsilon}_j^{\,2} - \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} \right] \right| = F_1(\hat{\nu}_n), \tag{10}$$

where $0 \leqslant \alpha < 1/2$, and $\hat{\varepsilon}_j$, $j = 1, \dots, n$, are OLS residuals of the regression model under consideration. In [7], it is proved that the Hölder norm of a polygonal process is realized at vertices. Therefore, the equality $T_{n,\alpha} = F_1(\hat{\nu}_n)$ is valid.

2.1. Convergence of $T_{n,\alpha}$ under the null hypothesis

Recall that the null hypothesis H_0 specifies model (1). For $0 \leqslant \alpha < 1/2$, set

$$T_\alpha = \sup_{0<h<1} h^{-\alpha} \sup_{0\leqslant t<1-h} |B_{t+h} - B_t|. \tag{11}$$

The following theorem gives a limit law for $T_{n,\alpha}$ when H_0 holds.

THEOREM 5. *Let $0 < \alpha < 1/2$ and $1 \leqslant p < 1/(1-\alpha)$. For model* (1), *assume that the function $\boldsymbol{f}$ is continuous and has finite p-variation. If*

$$\lim_{t\to\infty} t^{2/(1/2-\alpha)} \mathrm{P}\big(|\varepsilon_1| > t\big) = 0,$$

then

$$n^{-1/2}\hat{\delta}_n^{-1} T_{n,\alpha} \xrightarrow[n\to\infty]{\mathcal{D}} T_\alpha.$$

Proof. One can easily show that

$$n^{-1/2}\hat{\delta}_n^{-1} T_{n,\alpha} = F_1\big(n^{-1/2}\hat{\delta}_n^{-1}\hat{\nu}_n\big),$$

where the functional F_1 is defined above. The result now follows by Proposition 4.

For the selected values of the Hölder exponent $\alpha = j/32$, $j = 0, \dots, 15$, we have generated $R = 2^{14}$ replications of approximations for T_α. The Brownian bridge in each replication was approximated by the partial sum process $\xi_m(0) = 0$,

$$\xi_m(t) = \sum_{j=1}^{[mt]} Z_j + (mt - [mt]) Z_{[mt]+1} - t \sum_{j=1}^{m} Z_j, \quad t \in [0, 1].$$

Here Z_j, $j = 1, \dots, m$, are generated independent standard normal random variables and $m = 2^{17}$. Having this set of replications, we took empirical quantiles of level $1 - c$ to approximate the unknown critical values of significance level c. Table 1 presents the results for three values of significance level.

2.2. Consistency of $T_{n,\alpha}$ under various heteroskedasticity alternatives

For each $n \geqslant 1$, the alternative hypothesis specifies the model

$$y_j = \boldsymbol{f}^{\mathrm{T}}(j/n)\boldsymbol{\beta} + \varepsilon_j^*, \quad j = 1, \dots, n, \tag{12}$$

where the errors ε_j^*, $j = 1, \dots, n$, are independent but may be not identically distributed.

Table 1. Approximations for critical values of T_α

	$c=0.10$	$c=0.05$	$c=0.01$
$\alpha=0$	1.6140	1.7418	2.0188
$\alpha=1/32$	1.6673	1.7963	2.0781
$\alpha=1/16$	1.7243	1.8547	2.1371
$\alpha=3/32$	1.7869	1.9180	2.2070
$\alpha=1/8$	1.8533	1.9857	2.2798
$\alpha=5/32$	1.9265	2.0612	2.3630
$\alpha=3/16$	2.0062	2.1438	2.4475
$\alpha=7/32$	2.0918	2.2346	2.5387
$\alpha=1/4$	2.1916	2.3358	2.6484
$\alpha=9/32$	2.3058	2.4555	2.7683
$\alpha=5/16$	2.4432	2.5891	2.9029
$\alpha=11/32$	2.6096	2.7563	3.0712
$\alpha=3/8$	2.8183	2.9586	3.2811
$\alpha=13/32$	3.0943	3.2333	3.5662
$\alpha=7/16$	3.5006	3.6404	3.9317
$\alpha=15/32$	4.1841	4.3107	4.5828

Note that $T_{n,\alpha}$ is constructed using the residuals $\hat{\boldsymbol{\varepsilon}}^* = (\hat{\varepsilon}_1^*, \dots, \hat{\varepsilon}_n^*)^{\mathrm{T}}$ of model (12) that are obtained using the OLS method. Of course, in practice, we do not know which of models (1) or (12) we deal with. This alternative specification and notation of error term is needed only to establish a theoretical result.

The consistency of the test statistics $T_{n,\alpha}$ can be proved for a wide range of heteroskedasticity alternatives. To see this, first note that, for residuals of model (12), we have $\hat{\varepsilon}_j^* = \varepsilon_j^* + \gamma_j$ with $\gamma_j = \boldsymbol{f}^{\mathrm{T}}(j/n)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ $j = 1, \dots, n$. Then, for integers ℓ and k such that $1 \leqslant \ell < n$ and $0 \leqslant k \leqslant n - \ell$ and a vector $\boldsymbol{x} = (x_1, \dots, x_n)^{\mathrm{T}}$ of real-valued numbers, set

$$\psi_{k,\ell}(\boldsymbol{x}) = \sum_{j=k+1}^{k+\ell} \left[x_j - n^{-1} \sum_{i=1}^{n} x_i \right].$$

If ψ^* stands for ψ_{k^*,ℓ^*} with any $1 \leqslant \ell^* < n$ and $0 \leqslant k^* \leqslant n - \ell^*$, we have

$$\begin{aligned} n^{-1/2} T_{n,\alpha} &\geqslant n^{-1/2+\alpha} (\ell^*)^{-\alpha} \left| \psi^*\big(\hat{\boldsymbol{\varepsilon}}^{*2}\big) \right| \\ &\geqslant n^{-1/2+\alpha} (\ell^*)^{-\alpha} \left[\left| \psi^*\big(\boldsymbol{\varepsilon}^{*2}\big) \right| - \left| 2\psi^*\big(\boldsymbol{\varepsilon}^* \boldsymbol{\gamma}\big) + \psi^*\big(\boldsymbol{\gamma}^2\big) \right| \right] \\ &\geqslant n^{-1/2+\alpha} (\ell^*)^{-\alpha} \left| \psi^*\big(\boldsymbol{\sigma}^2\big) \right| - \Delta_n, \end{aligned}$$

where $\boldsymbol{\sigma}^2 = (\mathrm{E}\varepsilon_1^{*2}, \dots, \mathrm{E}\varepsilon_n^{*2})$ and

$$\Delta_n = n^{-1/2+\alpha} (\ell^*)^{-\alpha} \left[\left| \psi^*\big(\boldsymbol{\varepsilon}^{*2}\big) - \psi^*\big(\boldsymbol{\sigma}^2\big) \right| + 2\left| \psi^*\big(\boldsymbol{\varepsilon}^* \boldsymbol{\gamma}\big) \right| + \left| \psi^*\big(\boldsymbol{\gamma}^2\big) \right| \right]. \tag{13}$$

For certain $\ell^* = \ell_n^*$, we introduce the sequence

$$D_n := n^{-1/2+\alpha} (\ell^*)^{-\alpha} \left| \psi^*\big(\boldsymbol{\sigma}^2\big) \right|.$$

Hence, if

$$D_n \to \infty \quad \text{as} \quad n \to \infty \tag{14}$$

and simultaneously

$$\Delta_n = o_P(D_n), \tag{15}$$

then $n^{-1/2}T_{n,\alpha} \to \infty$ in probability.

Assume that the errors in (12) are specified by

$$\varepsilon_j^* = g(j/n)u_j, \quad j = 1, \dots, n, \tag{16}$$

where u_j, $j = 1, \dots, n$ are i.i.d. zero-mean one-variance random variables, and g: $[0, 1] \to \mathbb{R}$ is a bounded function which determines all possible types of heteroskedasticity. In what follows, we consider error model (16). Assuming this, we have

$$D_n = n^{-1/2+\alpha}(\ell^*)^{-\alpha}\left|\sum_{j=k^*+1}^{k^*+\ell^*}\left[g^2(j/n) - n^{-1}\sum_{i=1}^{n} g^2(i/n)\right]\right|.$$

So, if, for some sequence θ_n, the inequality

$$\left|\sum_{j=k^*+1}^{k^*+\ell^*}\left[g^2(j/n) - n^{-1}\sum_{i=1}^{n} g^2(i/n)\right]\right| \geqslant \ell^*\theta_n \tag{17}$$

holds, then (14) becomes $n^{-1/2+\alpha}(\ell^*)^{1-\alpha}\theta_n \to \infty$.

Next we need to control (15). Recall that $\gamma_j = \boldsymbol{f}^{\mathrm{T}}(j/n)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Then

$$\psi^*(\boldsymbol{\gamma}^2) = \sum_{i,m=1}^{d} (\hat{\beta}_i - \beta_i)(\hat{\beta}_m - \beta_m)\psi^*(f_i f_m)$$

with the sequences $f_k = (f_k(1/n), \dots, f_k(n/n))$, $k = 1, \dots, d$. If, for all $i = 1, \dots, d$ and $j = 1, \dots, n$, we use the rough estimate $|f_i(j/n)| \leqslant \|f_i\|_\infty$, then we obtain

$$\left|\psi^*(\boldsymbol{\gamma}^2)\right| \leqslant 2\ell^*\left(\sum_{i=1}^{d} |\hat{\beta}_i - \beta_i| \cdot \|f_i\|_\infty\right)^2 \leqslant c\ell^*|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|^2. \tag{18}$$

Since $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})| = O_P(n^{-1/2})$ even in the heteroskedastic model, we have $|\psi^*(\boldsymbol{\gamma}^2)| = O_P(\ell^*/n)$ for the third term in the parentheses in (13).

For the first term, we have

$$\begin{aligned}
&\mathrm{E}\big(\psi^*(\boldsymbol{\varepsilon}^{*2}) - \psi^*(\boldsymbol{\sigma}^2)\big)^2 \\
&\quad = (\mathrm{E}u_1^4 - 1)\left[(1 - 2\ell^*/n)\sum_{j=k^*+1}^{k^*+\ell^*} g^4(j/n) + (\ell^*/n)^2\sum_{j=1}^{n} g^4(j/n)\right] \\
&\quad \leqslant \ell^*(\mathrm{E}u_1^4 - 1)\|g\|_\infty^4.
\end{aligned}$$

Hence, $\psi^*(\boldsymbol{\varepsilon}^{*2}) - \psi^*(\boldsymbol{\sigma}^2) = O_P(\ell^{*1/2})$.

Since

$$\left|\psi^*(\boldsymbol{\varepsilon}^*\boldsymbol{\gamma})\right| \leqslant \left|\sum_{j=k^*+1}^{k^*+\ell^*}\left[u_j g(j/n)\boldsymbol{f}^{\mathrm{T}}(j/n) - n^{-1}\sum_{i=1}^{n} u_i g(i/n)\boldsymbol{f}^{\mathrm{T}}(i/n)\right]\right| \cdot |\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|$$

and

$$\mathrm{E}\left|\sum_{j=k^*+1}^{k^*+\ell^*}\left[u_j g(j/n)\boldsymbol{f}^{\mathrm{T}}(j/n) - n^{-1}\sum_{i=1}^{n} u_i g(i/n)\boldsymbol{f}^{\mathrm{T}}(i/n)\right]\right|^2 \leqslant c\ell^*,$$

we have $|\psi^*(\boldsymbol{\varepsilon}^*\boldsymbol{\gamma})| = \mathrm{O_P}((\ell^*/n)^{1/2})$.

Collecting all estimates, we finally obtain

$$n^{-1/2}T_{n,\alpha} \geqslant n^{-1/2+\alpha}(\ell^*)^{1-\alpha}\theta_n - \mathrm{O_P}\big((\ell^*/n)^{1/2-\alpha}\big).$$

It remains to deal with $\hat{\delta}_n$. Note, however, that, under model (12), $\hat{\delta}_n^2$ is defined by (7) but with $\hat{\boldsymbol{\varepsilon}}^*$ instead of $\hat{\boldsymbol{\varepsilon}}$ and so includes the function g. In the Appendix, it is proved (see Proposition 8) that, for any bounded g, $\hat{\delta}_n$ converges in probability to some finite limit, provided that $\mathrm{E}u_1^4 < \infty$. Hence, we have proved the following result.

PROPOSITION 6. *Assume that model* (12) *is specified with* $\varepsilon_j^* = g(j/n)u_j$, $j = 1, \ldots, n$, *where* (u_j) *are i.i.d. zero-mean random variables with finite* $\mathrm{E}u_1^4$, *and* $g\colon [0,1] \to \mathbb{R}$ *is a bounded function. If there exist intervals* $[k^*+1, k^*+\ell^*] \subset [1,n]$ *such that* (17) *holds with a sequence* θ_n *and*

$$\ell^* \to \infty, \quad \ell^*/n \to 0, \quad n^{-1/2+\alpha}(\ell^*)^{1-\alpha}\theta_n \to \infty, \tag{19}$$

then

$$n^{-1/2}\hat{\delta}_n^{-1}T_{n,\alpha} \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Example 1. If, for example, $g(j/n) = \tau$ for $j \in \{k^*+1, \ldots, k^*+\ell^*\}$ and $g(j/n) = \sigma$ for $j \notin \{k^*+1, \ldots, k^*+\ell^*\}$, then we have

$$\sum_{j=k^*+1}^{k^*+\ell^*}\left[g^2(j/n) - n^{-1}\sum_{i=1}^{n} g^2(i/n)\right] = \ell^*(1-\ell^*/n)[\tau^2-\sigma^2]. \tag{20}$$

This example corresponds to a changed-segment type alternative. Condition (19) then becomes

$$\ell^* \to \infty, \qquad \ell^*/n \to 0, \qquad (1-\ell^*/n)\big|\tau^2-\sigma^2\big|n^{-1/2+\alpha}(\ell^*)^{1-\alpha} \to \infty.$$

Example 2. As another example, consider $g(t) = (1+\lambda_n t)^{1/2}$, $t \in [0,1]$, where the sequence λ_n may also converge to zero. Then

$$\left|\sum_{j=k^*+1}^{k^*+\ell^*}\left[g^2(j/n) - n^{-1}\sum_{i=1}^{n} g^2(i/n)\right]\right| \geqslant c|\lambda_n|\ell^*,$$

and (19) is satisfied with λ_n such that $n^{-1/2+\alpha}(\ell^*)^{1-\alpha}|\lambda_n| \to \infty$.

3. SIMULATION STUDY

In this section, we support theoretical results by simulations. We are interested in two types of alternatives: changed-segment and the type of heteroskedasticity where the variance of error term is proportional to the exogenous variable and increases according to time trend.

For the first type of alternative, we compare the performance of the three tests(depending on the value of α) from the introduced class of tests and the likelihood-ratio test. The likelihood-ratio test introduced by Bartlett is intended to test for grouped heteroskedasticity and can be found in many papers (see, for example, [1] for several variants of this type of test). A changed-segment alternative can be viewed as a special case of grouped heteroskedasticity (two groups of unequal variance); therefore, the likelihood-ratio test is a proper test for this kind of alternative. This test is based on comparison of sample variances of the squared residuals in the non intersecting subsamples. More precisely, assume that $s^2 = \hat{\boldsymbol{\varepsilon}}^{\mathrm{T}}\hat{\boldsymbol{\varepsilon}}/n$; then the general form of the likelihood-ratio test statistic to test for different variances in G groups of lengths n_g, $g = 1, \ldots, G$, is

$$\mathrm{LR} = n\log(s^2) - \sum_{g=1}^{G} n_g \log(s_g^2), \quad s_g^2 = n_g^{-1} \sum_{j \in \{I_g\}} \hat{\varepsilon}_j^2,$$

and I_g is the index set of the gth group. Under the null hypothesis of constant variance, this statistic has the $\chi^2(G-1)$ distribution (in our case, $\chi^2(1)$). For our experiment, we set $d = 1$ and $\boldsymbol{f}(t) = t$, i.e., $f_1(j/n) = j/n$, $j = 1, \ldots, n$. We analyze the case with $n = 128$ and $u_1 \sim \mathrm{N}(0, 1)$ and, under the null hypothesis, we fix $\sigma = 1.0$.

When it is suspected that the variance of error term in linear regression model is proportional to components of the design matrix, one can use the Breusch–Pagan (see [2]) or similar tests. These tests are based on the information from the auxiliary regression of the following or similar form:

$$\hat{\varepsilon}_j^2 - s^2 = \boldsymbol{z}_j^{\mathrm{T}}\boldsymbol{b} + w_j, \quad j = 1, \ldots, n,$$

where $\boldsymbol{z}_j = (1, z_{1j}, \ldots, z_{d_z j})^{\mathrm{T}}$ is a vector of explanatory variables, $\boldsymbol{b}$ is a d_z-length vector of model coefficients, and $(w_j,\ j = 1, \ldots, n)$ is a white noise component. Then the test statistic is obtained as an explained sum of squares from the regression

$$\mathrm{BP} = (\hat{\boldsymbol{\varepsilon}} - s^2)^{\mathrm{T}}\boldsymbol{Z}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{\varepsilon}} - s^2)/v^2, \quad v^2 = (\hat{\boldsymbol{\varepsilon}}^2 - s^2)^{\mathrm{T}}(\hat{\boldsymbol{\varepsilon}}^2 - s^2)/n,$$

where $\boldsymbol{Z} = (z_1, z_2, \ldots, z_n)^{\mathrm{T}}$ is an $N \times d_z$ design matrix for auxiliary regression. Under the null hypothesis of constant variance, this statistic has the $\chi^2(d_z - 1)$ distribution, which in our case is $\chi^2(1)$. We compare the performance of BP test and the three tests from the proposed class. For the second experiment, we set $d = 2$ and $\boldsymbol{f}(t) = 1 + t$, i.e., $f_1(j/n) \equiv 1$ and $f_2(j/n) = j/n$ for $j = 1, \ldots, n$. We again take $n = 128$, $u_1 \sim \mathrm{N}(0, 1)$, and $\sigma = 1.0$ under H_0. For the auxiliary regression, $\boldsymbol{z}_j = \boldsymbol{f}(j/n)$.

First, we investigate the behavior of the tests under the null hypothesis. We have computed $R = 10000$ replications of the test statistics and, for the significance level $c = 0.05$, measured the empirical size. For the LR and BP tests, we obtained 0.056 and 0.049, which shows a good performance under H_0. But for the class $T_{n,\alpha}$, due to the small sample size, we should expect discrepancy between c and the test size, especially with α values closer to 1/2. The explanation for this is the following. Recall that the Brownian bridge in limiting statistics T_α is approximated by the partial sum process $\zeta_m(0) = 0$,

$$\zeta_m(t) = \sum_{j=1}^{[mt]} Z_j + (mt - [mt])Z_{[mt]+1} - t\sum_{j=1}^{m} Z_j, \quad t \in [0, 1],$$

with normalization $m^{-1/2}$ and $Z_j \sim \mathrm{N}(0, 1)$, $j = 1, \ldots, m$. Taking as much as $m = 2^{17}$ (that we did in our calculations and that, it is believed, leads to a close approximation of T_α) does not necessarily provides us with the correct critical region when the sample size is small. As a consequence, we can obtain a discrepancy between the significance level and empirical size, especially for larger α. To fix this problem, we suggest the following:

- For a small sample size n, in order to obtain the right critical values (leading to the correct test size under the null hypothesis), approximate the Brownian bridge in T_α by taking $m = n$.

The cases where the problem could be not fully solved might be when we test for a changed segment in variance of observations (see [10] for this problem), which is also the case of heteroskedasticity of error terms in the linear model. In this case, we actually test for a change in the mean of variables that are $\chi^2(1)$-distributed if the observations are normal and close to $\chi^2(1)$ if the original observations are close to normal.

Table 2. The size of tests with different approximations to quantiles of distribution of T_α

α	$m = 2^{17}$	$m = n,\ Z$	$m = n,\ Z^2$	α	$m = 2^{17}$	$m = n,\ Z$	$m = n,\ Z^2$
0	0.017	0.040	0.046	$1/4$	0.068	0.144	0.047
$1/32$	0.018	0.041	0.046	$9/32$	0.116	0.228	0.051
$1/16$	0.019	0.043	0.045	$5/16$	0.198	0.349	0.051
$3/32$	0.019	0.045	0.045	$11/32$	0.302	0.498	0.051
$1/8$	0.021	0.048	0.045	$3/8$	0.423	0.645	0.053
$5/32$	0.025	0.054	0.046	$13/32$	0.521	0.773	0.054
$3/16$	0.030	0.067	0.045	$7/16$	0.564	0.866	0.053
$7/32$	0.040	0.093	0.044	$15/32$	0.510	0.919	0.053

Nevertheless, the Brownian bridge is approximated, using standard normal random variables. Therefore, taking $Z_j \sim \mathrm{N}(0,1)$ in $\zeta_m(t)$ is not sufficient to fix the problem of correct critical values. Thus, another adjustment is needed:

- For a small sample size n, when testing for a changed segment in variance, in addition, take Z_j^2 instead of Z_j, where $Z_j \sim \mathrm{N}(0,1)$. Note, however, that the normalization for $\zeta_m(t)$ also changes from $m^{-1/2}$ to $(2m)^{-1/2}$.

Table 2 presents test sizes for $T_{n,\alpha}$. The notations Z and Z^2 indicate that respectively $Z_j \sim \mathrm{N}(0,1)$ and Z_j^2 were taken.

The results show that, for the chosen significance level $c = 0.05$, the correct test size is obtained with $m = n$ and Z_j^2 instead of Z_j, even for α very close to $1/2$. Naturally, one should expect similar results for other values of c or small n. Both practical tips might not help if the observations are not normally distributed, and further inspection might be needed. However, in this paper, we restrict ourselves to classical assumptions of the linear regression model.

Next, in our simulation study, we assume that the alternative hypothesis is true. From the class of the test statistics we take those with $\alpha = 0$, $\alpha = 7/32$, and $7/16$. We repeat data generation and testing procedures $R = 10000$ times. For the first type of alternative, we generate a changed segment of length $\ell^* = 16$ and $\tau = 2.0$. We are interested in two cases of the location of the changed segment. First, we take the beginning of the changed segment fixed, i.e., $k^* = 56$, then every of R times we take a random point to begin the changed segment. In the case of the likelihood-ratio statistic, according to the configuration of simulations, $G = 2$, $n_1 = 112$, and $n_2 = 16$. For the second type of alternative, we fixed the function $g(t) = (1 + 3t)^{1/2}$, i.e., when H_1 is true, the variance of error increases trend-like from 1 to 4. For this case, we compare the proposed tests with the Breusch–Pagan test.

For test statistics $T_{n,\alpha}$, we set critical values following the suggested procedure (taking $m = n$ and Z_j^2). We saw that when H_0 holds, this allows us to obtain the right test size.

In Figure 1, the left graph displays the empirical power of tests with fixed k^*, and the middle graph with random k^* (denoted by E and RE in the left upper corners). The graph on the right displays the power of tests for the second type of alternative (T stands for trend-like-type alternative). As it was expected and confirmed in many studies, the LR test is powerful but, as results show, only when we "tell" the test where to look for the changed segment. When we have no a priori information about the location of possible change with the LR test, there is little chance to detect the change when it is actually present, in contrast to the proposed class of tests, for which random location of the changed segment has no impact on power.

Graphs E and RE in Figure 1 also show that $T_{n,\alpha}$ is able to detect very short changed segments with α away from 0, but not necessarily with the largest α. This suggests that there should be optimal (for example, in terms of the largest power) α for a changed segment of certain length. Moreover, this might lead to a data-driven choice of parameter α, in contrast to arbitrary choice, which we use now. So the question of choosing parameter α remains open.

Graph T shows that the BP test is more powerful to detect the second type ofalternative, but, again in advance, we suspect that the variance of the error term is proportional to a component of the design matrix

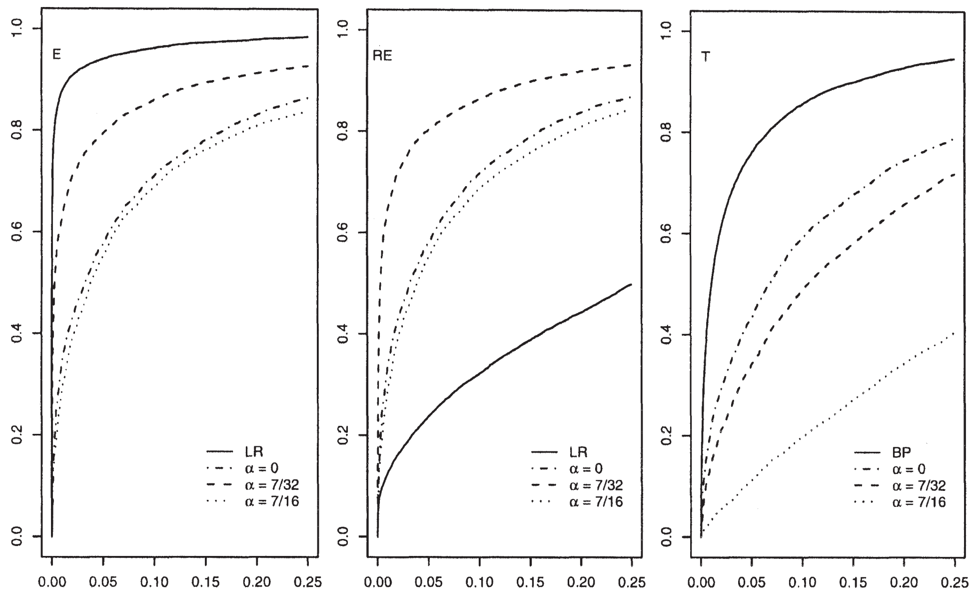

Figure 1. Empirical power of tests

(and this is the reason why we use the Breusch–Pagan test, which is a proper test in such situations). In contrast, the $T_{n,\alpha}$ class of tests needs no such a priori information to be able to detect the change. From this graph we also see that the $T_{n,\alpha}$ test has the largest power with $\alpha = 0$. This could be explained by the configuration of the alternative – the variance gradually increases. So, probably in the short segment, the variance does not differ much from that of the rest of the sample.

4. CONCLUSIONS

In this paper, we address the problem of testing for heteroskedasticity in linear regression models. The constant variance of the error term in the model is one of the assumptions in the classical linear regression framework. If this assumption is violated, further analysis may be incorrect. Therefore, a diagnostic check should be performed before taking further steps. There are a lot of classical tests like Glejser [4], Goldfeld–Quandt [6], Szroeter [11], White [12], and other for various types of heteroskedasticity. However, such tests usually require some prior information: the type of heteroskedasticity, the possible location of the change in the variance of errors, or other. We propose a class of tests $T_{n,\alpha}$ which does not need this type of information. It is stated in thetheorem and supported with the results of computer simulations that this class of tests has power to detect different types of heteroskedasticity, especially, the chan ged segment. A few computational tips are provided for implementing these tests in practice when having a small sample size.

Acknowledgments. Authors would like to thank the anonymous referee for valuable comments and precise suggestions that led to an improvement of the paper.

A PROOFS

A1. Auxiliary results

PROPOSITION 7. *For model* (1), *assume that* $\mathrm{E}\varepsilon_1^4 < \infty$ *and the function* $\boldsymbol{f}$ *is bounded. Then, for* $\delta^2 = \mathrm{E}\varepsilon_1^4 - \sigma^4$,

$$\hat{\delta}_n \xrightarrow[n\to\infty]{\mathrm{P}} \delta. \tag{21}$$

Proof. Let δ_n^2 denote the quantity defined as $\hat{\delta}_n^2$ but with disturbances instead of residuals. Then

$$|\hat{\delta}_n - \delta_n| = o_{\mathrm{P}}(1).$$

Indeed, by the triangular inequality one easily checks that

$$|\hat{\delta}_n - \delta_n| \leqslant \left[4\left(n^{-1} \sum_{j=1}^{n} \varepsilon_j^2 \big| \boldsymbol{f}(j/n) \big|^2 \right)^{1/2} + 2\left(n^{-1} \sum_{j=1}^{n} | \boldsymbol{f}(j/n) \big|^4 \right)^{1/2} \right] |\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|.$$

Noting that $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}| = o_{\mathrm{P}}(1)$ and $n^{-1} \sum_{j=1}^{n} \varepsilon_j^2 |\boldsymbol{f}(j/n)|^2 = O_{\mathrm{P}}(1)$, we conclude, since $\delta_n \xrightarrow[n\to\infty]{\mathrm{P}} \delta$ by the law of large numbers.

Proposition 8. *For model* (12)*, assume that* $\varepsilon_j^* = g(j/n)u_j$, $j = 1, \ldots, n$, *where* (u_j) *are i.i.d. mean zero and* $\mathrm{E}u_1^4 < \infty$ *and that both functions* $\boldsymbol{f}$ *and* g *are bounded. Then*

$$\hat{\delta}_n^2 \xrightarrow[n\to\infty]{\mathrm{P}} \mathrm{E}u_1^4 \int_0^1 g^4(s)\,\mathrm{d}s - (\mathrm{E}u_1^2)^2 \left[\int_0^1 g^2(s)\,\mathrm{d}s \right]^2. \tag{22}$$

Proof. Let δ_n^* be defined in the same way as $\hat{\delta}_n$ but with (ε_j^*) instead of residuals. Then

$$|\hat{\delta}_n - \delta_n^*| \leqslant \left[4\left(n^{-1} \sum_{j=1}^{n} \varepsilon_j^{*2} \big| \boldsymbol{f}(j/n) \big|^2 \right)^{1/2} + 2\left(n^{-1} \sum_{j=1}^{n} \big| \boldsymbol{f}(j/n) \big|^4 \right)^{1/2} \right] |\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|.$$

One easily checks that $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}| = o_{\mathrm{P}}(1)$ and $n^{-1} \sum_{j=1}^{n} \varepsilon_j^{*2} |\boldsymbol{f}(j/n)|^2 = O_{\mathrm{P}}(1)$. By the law of large numbers it follows that

$$\delta_n^{*2} \xrightarrow[n\to\infty]{\mathrm{P}} \mathrm{E}u_1^4 \int_0^1 g^4(s)\,\mathrm{d}s - \big(\mathrm{E}u_1^2\big)^2 \left[\int_0^1 g^2(s)\,\mathrm{d}s \right]^2.$$

The following proposition is a simple corollary from Young [13].

Proposition 9. *Let* $p, q \geqslant 1$ *and* $1/p + 1/q > 1$. *If* $\boldsymbol{g} \in V_p^d[0,1]$ *and* $x \in V_q[0,1]$, *then the integral* $\int_0^1 \boldsymbol{g}(u)\,\mathrm{d}x(u)$ *is well defined in the Riemann–Stieltjes sense and there exists a constant* $K_d > 0$ *depending on* d *only such that*

$$\left| \int_0^1 \boldsymbol{g}(u)\,\mathrm{d}x(u) \right| \leqslant K_d \|\boldsymbol{g}\|_{[p]} v_q^{1/q}(x).$$

Proposition 10. *Let* $0 < \alpha < 1/2$. *Assume that* $Z_1, \ldots, Z_n$ *are i.i.d. mean-zero random variables with variance* $\sigma^2 = \mathrm{E}Z_1^2$ *and consider the process*

$$\zeta_n(t) = \sum_{j=1}^{[nt]} Z_j + (nt - [nt]) Z_{[nt]+1}, \quad t \in [0,1].$$

Then

$$\|\zeta_n\|_\alpha = O_{\mathrm{P}}(n^{1/2})$$

if and only if

$$\sup_{t>0} t^{1/(1/2-\alpha)} \mathrm{P}\bigl(|Z_1| > t\bigr) < \infty. \tag{23}$$

Proof. First, we prove the necessity. For $n \geqslant 1$, we have

$$\begin{aligned}
n^{-1/2+\alpha} \max_{1\leqslant k\leqslant n} |Z_k| &\leqslant n^{-1/2+\alpha} \max_{0<|k-j|\leqslant 1} \frac{|\sum_{i=1}^{k} Z_i - \sum_{i=1}^{j} Z_i|}{|k-j|^\alpha} \\
&\leqslant n^{-1/2+\alpha} \max_{0<|k-j|<n} \frac{|\zeta_n(k/n) - \zeta_n(j/n)|}{|k-j|^\alpha} \\
&= \omega_\alpha\bigl(n^{-1/2}\zeta_n, 1\bigr) \leqslant \bigl\|n^{-1/2}\zeta_n\bigr\|_\alpha = \mathrm{O_P}(1),
\end{aligned}$$

and we easily conclude the necessity of condition.

Next, we prove sufficiency. It is easy to check that the Hölder norm of polygonal process is realized at vertices (see, e.g., [8] and references therein). Therefore,

$$\|\zeta_n\|_\alpha = \max_{1\leqslant k\leqslant n} |S_k| + T_n, \tag{24}$$

where

$$T_n = \max_{1\leqslant k<m\leqslant n} \frac{|S_m - S_k|}{(m/n - k/n)^\alpha} = \max_{1\leqslant \ell<n} \frac{1}{(\ell/n)^\alpha} \max_{0\leqslant k\leqslant n-\ell} |S_{k+\ell} - S_k|,$$

$S_0 = 0$, $S_k = \sum_{j=1}^{k} Z_j$, $k = 1, \dots, n$. Since $n^{-1/2} \max_{1\leqslant k\leqslant n} |S_k| = \mathrm{O_P}(1)$, we need only to prove that

$$\mathrm{P}_1(A) := \mathrm{P}\bigl(n^{-1/2} T_n > A\bigr) \to 0 \quad \text{as } A \to \infty. \tag{25}$$

Using the indicator function $\chi\{\cdot\}$, define the truncated random variables

$$Z'_j = Z_j \chi\bigl\{|Z_j| \leqslant A n^{1/p}\bigr\}, \qquad \widetilde{Z}_j = Z_j - Z'_j, \quad j = 1, \dots, n,$$

where we denote $p = (1/2 - \alpha)^{-1}$. Then

$$\mathrm{P}_1(A) \leqslant n\mathrm{P}\bigl(|Z_1| \geqslant A n^{1/p}\bigr) + \mathrm{P}'_1(A),$$

where

$$\mathrm{P}'_1(A) := \mathrm{P}\left(n^{-1/2} \max_{1\leqslant \ell<n} \frac{1}{(\ell/n)^\alpha} \max_{0\leqslant k\leqslant n-\ell} \left|\sum_{j=k+1}^{k+\ell} Z'_j\right| > A\right).$$

Condition (23) yields $n\mathrm{P}(|Z_1| \geqslant A n^{1/p}) \to 0$ as $A \to \infty$. Since

$$\begin{aligned}
\mathrm{E}Z'_j = -\mathrm{E}Z_j \chi\bigl\{|Z_j| \geqslant A n^{1/p}\bigr\} &= -\int_{A n^{1/p}}^{\infty} \mathrm{P}\bigl(|Z_1| \geqslant t\bigr)\,\mathrm{d}t \\
&\leqslant -\sup_t t^p \mathrm{P}\bigl(|Z_1| \geqslant t\bigr) \int_{A n^{1/p}}^{\infty} t^{-p}\,\mathrm{d}t \leqslant -c A^{-p+1} n^{-1+1/p},
\end{aligned}$$

we have

$$n^{-1/2}\max_{1\leqslant \ell<n}\frac{1}{(\ell/n)^{\alpha}}\max_{0\leqslant k\leqslant n-\ell}\left|\sum_{j=k+1}^{k+\ell}\mathrm{E}Z_j'\right|=n^{1/2}\left|\mathrm{E}Z_1'\right| \leqslant cA^{-p+1}n^{-1/2+1/p}.$$

Hence, (25) reduces to

$$\mathrm{P}_2'(A):=\mathrm{P}\bigl(n^{-1/2}T_n'>A\bigr)\to 0\quad \text{as } A\to\infty, \tag{26}$$

where

$$T_n'=\max_{1\leqslant \ell<n}\frac{1}{(\ell/n)^{\alpha}}\max_{0\leqslant k\leqslant n-\ell}\left|S_{k+\ell}'-S_k'\right|,$$

$S_0'=0, S_k'=\sum_{j=1}^{k}(Z_j'-\mathrm{E}Z_j'),\quad k=1,\dots,n.$

We shall use dyadic splitting of the ℓ's and k's indexation ranges. Defining the integer J_n by

$$2^{J_n}\leqslant n<2^{J_n+1},$$

we get

$$\begin{aligned}T_n'&=\max_{1\leqslant j\leqslant J_n+1}\max_{n2^{-j}<\ell\leqslant n2^{-j+1}}\frac{1}{(\ell/n)^{\alpha}}\max_{1\leqslant k\leqslant n-\ell}\left|S_{k+\ell}'-S_k'\right|\\&\leqslant \max_{1\leqslant j\leqslant J_n+1}\max_{n2^{-j}<\ell\leqslant n2^{-j+1}}2^{\alpha j}\max_{0\leqslant k<n-n2^{-j}}\left|S_{k+\ell}'-S_k'\right|\\&\leqslant \max_{1\leqslant j\leqslant J_n+1}\max_{n2^{-j}<\ell\leqslant n2^{-j+1}}2^{\alpha j}\max_{1\leqslant i<2^j}\max_{(i-1)n2^{-j}\leqslant k<in2^{-j}}\left|S_{k+\ell}'-S_k'\right|.\end{aligned}$$

For $n2^{-j}<\ell\leqslant n2^{-(j-1)}$ and $(i-1)n2^{-j}\leqslant k<in2^{-j}$, we have

$$\begin{aligned}\left|S_{k+\ell}'-S_k'\right|&\leqslant\left|S_{k+\ell}'-S_{[in2^{-j}]}'\right|+\left|S_{[in2^{-j}]}'-S_k'\right|\\&\leqslant\max_{in2^{-j}<u<(i+2)n2^{-j}}\left|S_u'-S_{[in2^{-j}]}'\right|+\max_{(i-1)n2^{-j}\leqslant k<in2^{-j}}\left|S_{[in2^{-j}]}'-S_k'\right|,\end{aligned}$$

where $[t]$ denotes the integer part of a real number t. Therefore,

$$T_n'\leqslant T_n''+T_n''',$$

where

$$\begin{aligned}T_n''&=n^{-1/2}\max_{1\leqslant j\leqslant J_n+1}2^{\alpha j}\max_{1\leqslant i<2^j}\max_{in2^{-j}<u<(i+2)n2^{-j}}\left|S_u'-S_{[in2^{-j}]}'\right|,\\T_n'''&=n^{-1/2}\max_{1\leqslant j\leqslant J_n+1}2^{\alpha j}\max_{1\leqslant i<2^j}\max_{(i-1)n2^{-j}\leqslant k<in2^{-j}}\left|S_{[in2^{-j}]}'-S_k'\right|.\end{aligned}$$

Consider the probability $\mathrm{P}_1''(A)=\mathrm{P}\{n^{-1/2}T_n''>A\}$. We have

$$\mathrm{P}_1''(A)\leqslant\sum_{j=1}^{J_n+1}\mathrm{P}\left\{\max_{1\leqslant i<2^j}\max_{in2^{-j}<u<(i+2)n2^{-j}}\left|S_u'-S_{[in2^{-j}]}'\right|>An^{1/2}2^{-\alpha j}\right\}$$

$$\leqslant \sum_{j=1}^{J_n+1} \sum_{1\leqslant i<2^j} \mathrm{P}\Big\{ \max_{in2^{-j}<u<(i+2)n2^{-j}} \big|S'_u - S'_{[in2^{-j}]}\big| > An^{1/2}2^{-\alpha j} \Big\}$$

$$\leqslant \sum_{j=1}^{J_n+1} 2^j \mathrm{P}\Big\{ \max_{u\leqslant 2n2^{-j}} \big|S'_u\big| > An^{1/2}2^{-\alpha j} \Big\}.$$

By the Chebyshev, Doob, and Rosenthal inequalities, for some q (accounting $\mathrm{E}(Z'_1-\mathrm{E}Z'_1)^2 < \mathrm{E}Z_1^2$), we have

$$\mathrm{P}\Big\{ \max_{u\leqslant 2n2^{-j}} \big|S'_u\big| > \lambda \Big\} \leqslant \lambda^{-q}\mathrm{E}\Big| \sum_{i=1}^{2n2^{-j}} \big(Z'_i - \mathrm{E}Z'_i\big)\Big|^q$$

$$\leqslant c\lambda^{-q}\big[\big(2n2^{-j}\big)^{q/2}\big(\mathrm{E}Z_1^2\big)^{q/2} + 2n2^{-j}\mathrm{E}\big|Z'_1-\mathrm{E}Z'_1\big|^q\big]$$

$$\leqslant c\lambda^{-q}\big[\big(n2^{-j}\big)^{q/2} + n2^{-j}A^{q-p}n^{(q-p)/p}\big],$$

since

$$\mathrm{E}\big|Z'_1-\mathrm{E}Z'_1\big|^q \leqslant c_q\mathrm{E}\big|Z'_1\big|^q \leqslant c_q \int_0^{An^{1/p}} u^{q-1}\mathrm{P}\big(|Z_1|>u\big)\,\mathrm{d}u$$

$$\leqslant n^{(q-p)/p}A^{q-p}\sup_u u^p P\big(|Z_1|>u\big).$$

Substituting this estimate, we obtain

$$\mathrm{P}''_1(A) \leqslant cA^{-q}\sum_{j=1}^{J_n+1} 2^{q\alpha j}n^{-q/2}2^j\big[\big(n2^{-j}\big)^{q/2} + n2^{-j}A^{q-p}n^{(q-p)/p}\big]$$

$$\leqslant cA^{-q}\sum_{j=1}^{J_n+1} 2^{q\alpha j}n^{-q/2}2^j\big(n2^{-j}\big)^{q/2}$$

$$+\,cA^{-p}\sum_{j=J}^{J_n+1} 2^{q\alpha j}n^{-q/2}2^j n2^{-j}n^{(q-p)/p}$$

$$\leqslant cA^{-q}\sum_{j=1}^{J_n+1} 2^{(q\alpha+1-q/2)j} + cA^{-p}n^{-q\alpha}\sum_{j=1}^{J_n+1} 2^{q\alpha j}.$$

Choosing $q > 1/(1/2-\alpha)$, we obtain

$$\mathrm{P}''_1(A) \leqslant c_0\big[A^{-q} + A^{-p}\big],$$

which converges to zero as $A\to\infty$.

Similarly, one proves that $\mathrm{P}\{n^{-1/2}T'''_n > A\} \to 0$.

It is convenient to introduce the following notation. For a vector $\boldsymbol{x} = (x_1,\dots,x_n)$, define $S_0(\boldsymbol{x}) = 0$ and

$$S_k(\boldsymbol{x}) = \sum_{j=1}^{k}\left[x_j - \frac{1}{n}\sum_{i=1}^{n} x_i\right], \quad k = 1,\dots,n.$$

Also define the function $\nu_n(\boldsymbol{x};\cdot)$ by

$$\nu_n(\boldsymbol{x};t)=S_{[nt]}(\boldsymbol{x})+(nt-[nt])\big(S_{[nt]+1}(\boldsymbol{x})-S_{[nt]}(\boldsymbol{x})\big),\quad t\in[0,1].$$

For two vectors $\boldsymbol{x}=(x_1,\dots,x_n)$ and $\boldsymbol{y}=(y_1,\dots,y_n)$, we denote $\boldsymbol{xy}=(x_1y_1,\dots,x_ny_n)$ and $\boldsymbol{x}^2=\boldsymbol{xx}$. For a function $g\colon[0,1]\to\mathbb{R}$ and $\boldsymbol{Z}=(Z_1,\dots,Z_n)$, let $\boldsymbol{Z}g=(Z_jg(j/n),\ \ j=1,\dots,n)$.

PROPOSITION 11. *If a continuous function $g\in V_p[0,1]$, then, assuming that $Z_1,\dots,Z_n$ are i.i.d. with zero mean and variance σ^2, we have*

$$n^{-1/2}\sigma^{-1}\nu_n(\boldsymbol{Z}g;\cdot)=\mathrm{O_P}(1),$$

provided that $\sup_{t>0}t^{1/(1/2-\alpha)}\mathrm{P}(|Z_1|>t)<\infty$.

Proof. For any $h\in V_p[0,1]$, consider the operator $T_h\colon H_\alpha^0\to H_\alpha^0$ defined by

$$T_h(x)(t)=\int_0^t h(u)\,\mathrm{d}x(u)-t\int_0^1 h(u)\,\mathrm{d}x(u).$$

First, we check that T_h is well defined. For a function $x\in H_\alpha^0[0,1]$ and $s,t\in[0,1]$ such that $|t-s|<\delta$, we have, with $q=1/\alpha$,

$$\begin{aligned}
v_q\big(x\mathbf{1}_{[s,t]}\big)&=\sup_{(t_k)\subset[s,t]}\sum_{k=1}^m\big|x(t_k)-x(t_{k-1})\big|^q\\
&=\sup_{(t_k)\subset[s,t]}\sum_{k=1}^m\left(\frac{|x(t_k)-x(t_{k-1})|}{|t_k-t_{k-1}|^{1/q}}\right)^q|t_k-t_{k-1}|\\
&\leqslant\left(\sup_{0<|t-s|<\delta}\frac{|x(t)-x(s)|}{|t-s|^{1/q}}\right)^q\sup_{(t_k)\subset[s,t]}\sum_{k=1}^m|t_k-t_{k-1}|\\
&=\omega_{1/q}^q(x,\delta)|t-s|.
\end{aligned}$$

By Proposition 9,

$$|T_hx(t)|\leqslant c\omega_\alpha(x,1)$$

for all $t\in[0,1]$, and, for $s\neq t$, we have

$$\begin{aligned}
\frac{|T_h(x)(t)-T_h(x)(s)|}{|t-s|^\alpha}&=\frac{1}{|t-s|^\alpha}\left|\int_0^1 h(u)\,\mathrm{d}x(u)\mathbf{1}_{[s,t]}(u)\right| \qquad (27)\\
&\quad+|t-s|^{1-\alpha}\left|\int_0^1 h(u)\,\mathrm{d}x(u)\right|\\
&\leqslant c\omega_\alpha(x,\delta)\|h\|_{[p]}. \qquad (28)
\end{aligned}$$

Hence, $T_h(x)\in H_\alpha^0[0,1]$ for any $x\in H_\alpha^0[0,1]$. Moreover,

$$\|T_h\|=\sup_{\|x\|_\alpha}\|T_hx\|_\alpha\leqslant c\|h\|_{[p]}. \qquad (29)$$

Noting that $\nu_n(\boldsymbol{Z}g; t) = T_{g_n}\zeta_n(t)$, $t \in [0, 1]$, where $g_n(t) = g((j-1)/n)$ for $t \in [(j-1)/n, j/n]$, $j = 1, \dots, n$, and $\zeta_n(t) = \sum_{j=1}^{[nt]} Z_j + (nt - [nt])Z_{[nt]+1}$, we have

$$\|T_{g_n}\zeta_n\|_\alpha \leqslant c\|g_n\|_{[p]} \cdot \|\zeta_n\|_\alpha.$$

Since $\|g_n\|_{[p]} < \infty$ and $\|\zeta_n\|_\alpha = \mathrm{O_P}(n^{1/2})$ by Proposition 10, we obtain the result.

A2. Proof of Theorem 1

Proof. Recall that $\hat{\varepsilon}_j = \gamma_j + \varepsilon_j$, $j = 1, \dots, n$; thus,

$$\nu_n(\hat{\boldsymbol{\varepsilon}}^2; t) = \nu_n(\boldsymbol{\varepsilon}^2; t) + 2\nu_n(\boldsymbol{\varepsilon\gamma}; t) + \nu_n(\boldsymbol{\gamma}^2; t), \quad t \in [0, 1].$$

So, it suffices to prove that

$$\left\|n^{-1/2}\nu_n(\boldsymbol{\varepsilon\gamma})\right\|_\alpha = \mathrm{o_P}(1), \tag{30}$$

$$\left\|n^{-1/2}\nu_n(\boldsymbol{\gamma}^2)\right\|_\alpha = \mathrm{o_P}(1). \tag{31}$$

Since $\nu_n(\boldsymbol{\varepsilon\gamma}, t) = \nu_n(\boldsymbol{\varepsilon f}, t)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}| = \mathrm{o_P}(1)$, the proof of (30) reduces to

$$\left\|n^{-1/2}\nu_n(\boldsymbol{\varepsilon} f_i)\right\|_\alpha = \mathrm{O_P}(1) \tag{32}$$

for each $i = 1, \dots, d$. But the latter follows from Proposition 11. Similarly, we prove (31). Since $\nu_n(\boldsymbol{\gamma}^2, t) = \nu_n(\boldsymbol{f}^2, t)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2$, we have only to check that

$$\sup_n n^{-1}\left\|\nu_n(\boldsymbol{f}^2)\right\|_\alpha < \infty.$$

This easily follows from the boundedness of $\boldsymbol{f}$:

$$\sup_n n^{-1}\left\|\nu_n(\boldsymbol{f}^2)\right\|_\alpha < c\|\boldsymbol{f}\|_\infty^2.$$

A3. Proof of Theorem 2

Proof follows from Theorem 1, Proposition 7, and Theorem 1 in Račkauskas and Suquet [9].

REFERENCES

1. J. K. Binkley, Finite sample behavior of tests for grouped heteroscedasticity, *The Review of Economics and Statistics*, **74**, 563–568 (1992).
2. T. S. Breusch and A. R. Pagan, A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, **47**, 1287–1294 (1979).
3. J. M. Dufour, L. Khalaf, J. T. Bernard, and I. Genest, Simulation-based finite-sample tests for heteroskedasticity and ARCH effects, *J. Econometrics*, **122**, 317–347 (2004).
4. H. Glejser, A new test for heteroscedasticity, *J. Amer. Statist. Assoc.*, **64**, 316–323 (1969).
5. L. G. Godfrey, C. D. Orme, and J. M. C. Santos Silva, Simulation-based tests for heteroskedasticity in linear regression models: Some further results, *Econometrics J.*, **9**, 76–97 (2006).
6. S. M. Goldfeld and R. Quandt, Some tests for heteroscedasticity, *J. Amer. Statist. Assoc.*, **60**, 539–547 (1965).

7. D. Hamadouche, Principe d'invariance dans les espaces hölderiens pour des variables α-mélangeantes ou associées, *Pub. IRMA Lille*, **37-IX** (1995).
8. D. Hamadouche, Weak convergence of smoothed empirical process in Hölder spaces, *Statist. Probab. Lett.*, **36**, 393–400 (1998).
9. A. Račkauskas and Ch. Suquet, Necessary and sufficient condition for the functional central limit theorem in Hölder spaces, *J. Theoret. Probab.*, **17**, 221–243 (2004).
10. A. Račkauskas and D. Zuokas, Off-line testing for a changed segment in the sample variance, *Lith. Math. J.*, **45**(2), 200–216 (2005).
11. J. Szroeter, A class of parametric tests for heteroscedasticity in linear econometric models, *Econometrica*, **46**, 1311–1327 (1978).
12. H. White, A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, *Econometrica*, **48**, 817–838 (1980).
13. L. C. Young, General inequalities for Stieltjes integrals and the convergence of Fourier series, *Math. Ann.*, **115**, 581–612 (1938).