S©MR ISSN 1813-3304

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports
http://semr.math.nsc.ru

Том 19, №1, стр. 292–308 (2022) УДК 519.233
DOI 10.33048/semi.2022.19.024 MSC 62F03

ON DETECTING ALTERNATIVES BY ONE-PARAMETRIC RECURSIVE RESIDUALS

A. I. SAKHANENKO

Abstract. We consider a linear regression model with one unknown parameter which is estimated by the least squares method. We suppose that, in reality, the given observations satisfy a close alternative to the linear regression model. We investigate the limiting behaviour of the normalized process of sums of recursive residuals. Such residuals were introduced by Brown, Durbin and Evans (1975) and their sums are a convenient tool for detecting discrepancy between observations and the studied model. In particular, under less restrictive assumptions we generalize a key result from Bischoff (2016).

Keywords: linear regression, recursive residuals, weak convergence, Wiener process, close alternative.

1. Introduction

1.1. **Statement of the problem.** Suppose that we observe random variables $Y_{n,1},\dots,Y_{n,N(n)}$ for which the following representation

$$(1)\qquad Y_{n,i}=g_{n,i}+\varepsilon_{n,i},\qquad i=1,2,\dots,N(n),\quad N(n)>1,$$

takes place, where (non-random) numbers $\{g_{n,i}\}$ may be unknown and random variables $\{\varepsilon_{n,i}\}$ are unobservable. We have introduced parameter n in (1) because below we are additionally going to consider the case of a triangular array of random variables. Our aim is to test the (null) hypothesis $\mathcal{H}_0(\mathrm{n})$ that

$$(2)\qquad g_{n,i}=\beta x_{n,i}\qquad\text{for all}\quad i=1,2,\dots,N(n),$$

Sakhanenko, A.I., On detecting alternatives by one-parametric recursive residuals.

The work is supported by Mathematical Center in Akademgorodok under agreement No. 075-15-2022-282 with the Ministry of Science and Higher Education of the Russian Federation.

Received October, 4, 2021, published May, 30, 2022.

with the unknown β and with given regressors $x_{n,1}, \dots, x_{n,N(n)}$.

Thus, when Hypothesis $\mathcal{H}_0$(n) is true, we have the simplest linear regression model with one unknown parameter β. As an alternative hypothesis we consider the case, when (1) takes place with numbers $g_{n,1}, \dots, g_{n,N(n)}$ such that

$$g_{n,k(\beta)} \neq \beta x_{n,k(\beta)} \qquad \text{with some} \quad k(\beta) \in \{1, 2, \dots, N(n)\}$$

for all possible values of the unknown parameter β. Of course, we suppose (as it is usual in applications) that only the pairs

$$(3) \qquad (Y_{n,i}, x_{n,i}), \qquad i = 1, 2, \dots, N(n),$$

are observed for the given n.

Our investigation is motivated by the paper of Wolfgang Bischoff (2016) where the simplest partial case of the stated problem was considered, when $x_{n,i} = 1$ for all $i, n \geq 1$. We present more general results under less restrictive assumptions in Section 2. It is worth mentioning that Bischoff (2016) used his results to discuss problems in experimental design. As in Bischoff (2016), we are going to construct tests using normalized sums $W_n(\cdot)$ of recursive regression residuals (see Subsection 1.4 for the definition). Such residuals were introduced in a well known paper of Brown, Durbin and Evans (1975). We cite this paper below as BDE. In Subsections 1.3 and 1.5 we remind some ideas and results introduced in BDE.

We emphasize that in the paper all unspecified limits are taken with respect to $n \to \infty$. Also we often use the following notation:

$$(4) \qquad v_{n,k}^2 := x_{n,1}^2 + \dots + x_{n,k}^2, \qquad k = 1, 2, \dots, N(n).$$

1.2. **Main assumptions on the model.** To consider the case when $n \to \infty$ we suppose below that we are given in (3) a triangular array of pairs. To clarify our initial assumptions we gathered them into the following two groups.

($\mathcal{A}_x$) we are given (non-random) integers n_0, $N(n)$ and $r(n)$ together with (non-random) real numbers $(x_{n,1}, \dots, x_{n,N(n)}), n \geq n_0$, such that

$$v_{n,r(n)}^2 > 0 \quad \text{with} \quad 0 < r(n) < N(n) \quad \text{for all} \quad n \geq n_0;$$

and, in addition, $m(n) := N(n) - r(n) \to \infty$.

($\mathcal{A}_Y$) for each $n \geq n_0$ we are given random variables $Y_{n,1}, \dots, Y_{n,N(n)}$ for which representation (1) takes place, where the joint distribution of random variables $\varepsilon_{n,1}, \dots, \varepsilon_{n,N(n)}$ does not depend on the choice of the (non-random) real numbers $g_{n,1}, \dots, g_{n,N(n)}$.

We say below that Hypothesis $\mathcal{H}_0$ is true, when Assumption ($\mathcal{A}_x$) is fulfilled and for all $n \geq n_0$ conditions (2) take place with some unknown $\beta = \beta_n$.

Thus, Assumption ($\mathcal{A}_x$) contains additional conditions on numbers $\{x_{n,i}\}$ from (2), by which we define the Hypothesis $\mathcal{H}_0$. We stress that both Assumptions ($\mathcal{A}_x$) and ($\mathcal{A}_Y$) may take place also in cases when the Hypothesis $\mathcal{H}_0$ is not true.

Note that Assumptions ($\mathcal{A}_x$) and ($\mathcal{A}_Y$) hold everywhere in the paper, but sometimes we use their partial cases. For example, instead of general Assumption ($\mathcal{A}_Y$) we often use the following more classical condition.

($\mathcal{B}$) for every $n \geq n_0$ random variables $\varepsilon_{n,1}, \dots, \varepsilon_{n,N(n)}$ are independent and identically distributed with a random variable ε such that

$$\mathbb{E}\varepsilon = 0 \qquad \text{and} \qquad 0 < \sigma^2 = \mathbb{E}\varepsilon^2 < \infty.$$

1.3. **Ideas from BDE.** For $n \ge n_0$ denote by $\widehat{\beta}_{n,k}$ the least-squares estimator of β, based on the first k observations from (3). It is known that

$$(5)\qquad \widehat{\beta}_{n,k} := \sum_{j=1}^{k} x_{n,j}Y_{n,j} \Big/ \sum_{j=1}^{k} x_{n,j}^2 = \sum_{j=1}^{k} x_{n,j}Y_{n,j}/v_k^2, \qquad \forall\, k \in [r(n), N(n)].$$

Introduce the following regression residuals

$$(6)\qquad \widehat{\varepsilon}_{n,k} := Y_{n,k} - x_{n,k}\widehat{\beta}_{n,k-1} = Y_{n,k} - x_{n,k}\sum\nolimits_{j=1}^{k-1} x_{n,j}Y_{n,j}/v_{n,k-1}^2,$$

for $k = r(n)+1, \dots, N(n)$, where the second equality in (6) follows from (5). Following BDE, we define the recursive regression residuals:

$$(7)\quad w_{n,i} := \frac{v_{n,r(n)+i-1}}{v_{n,r(n)+i}}\widehat{\varepsilon}_{n,r(n)+i} = \frac{v_{n,r(n)+i-1}}{v_{n,r(n)+i}}(Y_{n,r(n)+i} - x_{n,r(n)+i}\widehat{\beta}_{n,r(n)+i-1}),$$

for all possible $i = 1, 2, \dots, m(n) = N(n) - r(n)$.

On the other hand, introduce random variables

$$(8)\quad \widetilde{\eta}_{n,k} := \sum_{j=1}^{k} x_{n,j}\varepsilon_{n,j}, \quad \eta_{n,i} := \frac{v_{n,r(n)+i-1}}{v_{n,r(n)+i}}\Big(\varepsilon_{n,r(n)+i} - \frac{x_{n,r(n)+i}}{v_{n,r(n)+i-1}^2}\widetilde{\eta}_{n,r(n)+i-1}\Big),$$

for all possible values $k = r(n)+i$ and $i = 1, \dots, m(n)$. It is not difficult to verify (see details in Lemma 1 below) that

$$(9)\qquad w_{n,i} = \eta_{n,i}, \quad i = 1, 2, \dots, m(n), \quad \text{when Hypothesis } \mathcal{H}_0 \text{ is true.}$$

In BDE these equalities were used as evident.

Property A. *Suppose that Assumptions* $(\mathcal{A}_x)$ *and* $(\mathcal{B})$ *hold. Then for all* $n \ge n_0$

$$\mathbb{E}\eta_{n,i} = 0 \quad \textit{and} \quad \mathbb{E}\eta_{n,i}^2 = \sigma^2, \quad i = 1, 2, \dots, m(n).$$

Moreover, random variables $\eta_{n,1}, \dots, \eta_{n,m(n)}$ *are uncorrelated, i.e.*

$$\mathbb{E}\eta_{n,k}\eta_{n,j} = 0, \quad \textit{for all possible} \quad k > j \ge 1.$$

In particular, if, in addition, random variable ε *has a normal distribution, then* $\eta_{n,1}, \dots, \eta_{n,m(n)}$ *are independent and identically distributed with* ε.

This property follows from Lemma 1 in BDE. It is the main idea of the cited work. Property A together with equalities (9) shows that the approach of Brown, Durbin and Evans (1975) has a number of advantages in comparison with an alternative one developed by MacNeill (1978) and Bishoff (1998).

1.4. **On Invariance Principle for recursive residuals.** For any sequence of (one-dimensional) random or non-random values, say $\xi_{n,1}, \dots, \xi_{n,m}$, we will denote by $S_m(t; \xi_{n,\bullet})$ a continuous function of $t \in [0,1]$ with the following properties:

$$(10)\qquad \forall t \in [k/m, (k+1)/m] \quad S_m(t; \xi_{n,\bullet}) := \sum\nolimits_{i=1}^{k} \xi_{n,i} + (mt - k)\xi_{n,k+1}$$

for all $k = 0, 1, 2, \dots, m-1$, where $\sum_{i=1}^{0} \dots = 0$.

When $\{\xi_{n,\bullet}\}$ are random, such processes $S_m(\cdot; \xi_{n,\bullet})$ are often called the random broken lines. Recall that Donsker (1951) proved that, under Assumption $(\mathcal{B})$

$$\frac{1}{\sqrt{m}}S_m(\cdot; \varepsilon_{n,\bullet}) \Longrightarrow \sigma B(\cdot) \quad \text{in } C[0,1], \quad \text{when} \quad N(n) \ge m \to \infty,$$

where $B(\cdot)$ is a Brownian motion process called also a standard Wiener process. The natural question arises whether we have the similar convergence for the recursive residuals:

$$(11)\qquad W_n(\cdot) := \frac{1}{\sqrt{m(n)}} S_{m(n)}(\cdot; w_{n,\bullet}) \Longrightarrow \sigma B(\cdot) \quad \text{in } C[0,1], \quad \text{when } \mathcal{H}_0 \text{ is true,}$$

where $m(n) = N(n) - r(n) \to \infty$.

On the other hand, substituting equalities from (9) into (11) we obtain that for all $n \geq n_0$

$$(12)\qquad W_n(\cdot) = \widetilde{W}_n(\cdot) := \frac{1}{\sqrt{m(n)}} S_{m(n)}(\cdot; \eta_{n,\bullet}), \quad \text{when Hypothesis } \mathcal{H}_0 \text{ is true.}$$

But this fact implies that the following convergence

$$(13)\qquad \widetilde{W}_n(\cdot) = \frac{1}{\sqrt{m(n)}} S_{m(n)}(\cdot; \eta_{n,\bullet}) \Longrightarrow \sigma B(\cdot) \quad \text{in } C[0,1],$$

is a sufficient condition for the property (11).

Convergence (13) plays a crucial role in the present paper. It has several advantages over the condition (11) which plays an important role in statistical applications (see below Subsection 2.5 for details). For example, Hypothesis $\mathcal{H}_0$ is not involved in (13).

1.5. **On sufficient conditions for convergence** (13)**.** In this subsection we suppose that Assumptions $(\mathcal{A}_x)$ and $(\mathcal{B})$ hold.

Property B. *Assume that the random variable ε has a normal distribution. Then convergence* (13) *takes place.*

In BDE this fact was correctly used as an obvious corollary of Property A.

However, when the $\{\varepsilon_{n,i}\}$ are not normally distributed, the variables $\{\eta_{n,i}\}$ from (8) are not necessarily independent nor identically distributed, and hence, invariance principles for these weighted random variables are not obvious. Moreover, the problem to obtain the invariance principle (13) appeared to be sufficiently difficult. Here we agree with Bischoff (2016) that the paper of Sen (1982) until now remains the only known investigation of this problem.

Property C. *Convergence* (13) *also holds under certain additional conditions which may be found in Sen (1982), p. 311. In particular, convergence* (13) *takes place under Assumption* $(\mathcal{B})$ *with $r(n) = 1$, when $x_{n,i} = 1$ for all $i, n \geq 1$.*

We stress that in Sen (1982) the first assertion of Property C was proved in a more general case, when the unknown parameter β is multidimensional.

2. Results

2.1. **On main assumptions.** In the present paper, we consider the space $C[0,1]$ equipped with the supremum norm

$$\|x\| = \|x(\cdot)\| := \sup_{t\in[0,1]} |x(t)| = \max_{t\in[0,1]} |x(t)| < \infty \quad \forall\, x = x(\cdot) \in C[0,1].$$

Using ideas from strong approximations, due to Skorohod and Strassen, we may rewrite property (13) in the following way.

Proposition 1. *Suppose that convergence* (13) *takes place with some processes* $\widetilde{W}_n(\cdot)$ *defined for all* $n \geq n_0$. *Then for each* $n \geq n_0$ *there exist a number* $\pi_n > 0$ *and a standard Wiener process* $B_n(\cdot)$ *such that*

$$\mathbb{P}\big(\|\widetilde{W}_n(\cdot) - \sigma B_n(\cdot)\| > \pi_n\big) \leq \pi_n \to 0. \tag{14}$$

We say below that Assumption $(\mathcal{A})$ is fulfilled if Assumptions $(\mathcal{A}_x)$ and $(\mathcal{A}_Y)$ are satisfied and property (13) takes place.

So, Properties B and C contain sufficient conditions under which Assumption $(\mathcal{A})$ holds and, in particular, relations (13) and (14) take place together with (11).

2.2. **Main result.** Under Assumption $(\mathcal{A}_x)$ for any $\{g_{n,i}\}$ from (1) we may introduce numbers

$$\widetilde{h}_{n,k} := \sum_{j=1}^{k} x_{n,j} g_{n,j}, \quad h_{n,i} := \frac{v_{n,r(n)+i-1}}{v_{n,r(n)+i}} \Big(g_{n,r(n)+i} - \frac{x_{n,r(n)+i}}{v^2_{n,r(n)+i-1}} \widetilde{h}_{n,r(n)+i-1} \Big), \tag{15}$$

for all possible values $k = r(n), \ldots, N(n)$ and $i = 1, \ldots, m(n) = N(n) - r(n)$. Using general formula (10) define non-random functions

$$H_n(t) := \frac{1}{\sqrt{m(n)}} S_{m(n)}(t; h_{n,\bullet}) \qquad \text{for} \quad t \in [0, 1]. \tag{16}$$

Theorem 1. *Consider observations* $\{Y_{n,i}\}$ *from representation* (1) *and suppose that Assumption* $(\mathcal{A})$ *holds. Then for any numbers* $\{g_{n,i}\}$ *the following inequality*

$$\mathbb{P}\big(\|W_n(\cdot) - H_n(\cdot) - \sigma B_n(\cdot)\| > \pi_n\big) \leq \pi_n \to 0 \tag{17}$$

takes place with the standard Wiener processes $B_n(\cdot)$ *and numbers* π_n *introduced in Proposition 1.*

In particular, we have from (17) that for all possible functions $H_n(\cdot)$

$$W_n(\cdot) - H_n(\cdot) \Longrightarrow \sigma B(\cdot) \quad \text{in } C[0, 1]. \tag{18}$$

Remark 1. *We emphasize that from* (17) *we have that convergence* (18) *is, in some sense, uniform with respect to all possible values of functions* $H_n(\cdot)$ *and of numbers* $\{g_{n,i}\}$ *from representation* (1) *by which these functions* $H_n(\cdot)$ *are constructed.*

2.3. **A case of close alternatives.** Consider now functions $H_n(\cdot)$ which converge to some function $H(\cdot)$ so that

$$\Delta_n := \max \Big\{ \Big| H_n\big(k/m(n)\big) - H\big(k/m(n)\big) \Big| : 1 \leq k \leq m(n) \Big\} \to 0. \tag{19}$$

Recall that

$$H_n\big(k/m(n)\big) = \frac{1}{\sqrt{m(n)}} \sum\nolimits_{j=1}^{k} h_{n,j} \qquad \text{for} \quad k = 1, \ldots, m(n). \tag{20}$$

Corollary 1. *Suppose that a continuous function* $H(\cdot)$ *is such that* (19) *holds. Then, under Assumption* $(\mathcal{A})$

$$W_n(\cdot) \Longrightarrow \sigma B(\cdot) + H(\cdot) \quad \textit{in} \quad C[0, 1]. \tag{21}$$

Moreover, in this case

$$\mathbb{P}\big(\|W_n(\cdot) - H(\cdot) - \sigma B_n(\cdot)\| > \Delta_n + \widetilde{\Delta}_n + \pi_n\big) \leq \pi_n \to 0 \tag{22}$$

with the standard Wiener process $B_n(\cdot)$ *and numbers* $\pi_n > 0$ *from Theorem 1, where*

$$(23)\quad \widetilde{\Delta}_n := \widetilde{\Delta}_n(H) := \max_{1\le k<m(n)} \max_{|t|\le 1/m(n)} \Big| H\big(k/m(n)+t\big) - H\big(k/m(n)\big)\Big| \to 0.$$

The presented corollary immediately follows from Theorem 1, because

$$\|H_n(\cdot) - H(\cdot)\| \le \Delta_n + \widetilde{\Delta}_n \to 0.$$

We stress that convergence in (23) holds for any continuous (on $[0,1]$) function H.

2.4. **Partial Cases.** As in Bischoff (2016) consider in more detail the case when

$$(24)\qquad \forall\ i,n\ge 1 \quad x_{n,i}=1 \quad \text{with} \quad r(n)=1, \quad N(n)=n+1, \quad m(n)=n.$$

By Property C, for such $\{x_{n,i}\}$ we obtain (13) directly from Assumption $(\mathcal{B})$ without using complicated additional conditions from Sen (1982).

To approximate $\{g_{n,i}\}$ introduce notations

$$(25)\quad G_{n,k} := \sum\nolimits_{j=1}^{k} g_{n,j} \quad \text{and} \quad \delta_{n,k} := G_{n,k+1}/\sqrt{n} - G(k/n), \quad k=0,1,\ldots,n,$$

for some function G. We are going to essentially use the following conditions

$$(26)\qquad \delta_n := \sum\nolimits_{k=1}^{n} \frac{|\delta_{n,k}|}{k} \to 0 \qquad \text{and} \qquad \overline{\delta}_n := \max_{0\le k\le n} |\delta_{n,k}| \to 0.$$

Our aim is to obtain (21) with

$$(27)\qquad H(t) := G(t) - \int_0^t \frac{G(u)}{u}du \quad \text{for} \quad t\in(0,1] \qquad \text{and} \quad H(0)=0,$$

when the integral in (27) is well defined.

Theorem 2. *Let continuous function* G *have bounded variation on* $[0,1]$ *with*

$$(28)\qquad G(0)=0 \qquad \textit{and} \qquad \int_0^1 \log\frac{3}{u}\ |dG(u)| < \infty.$$

And suppose that conditions from (26) *hold together with assumptions* (24) *and* $(\mathcal{B})$. *Then convergence* (21) *and estimate* (22) *take place, where function* H *is defined in* (27) *and it is continuous on* $[0,1]$.

For example, all conditions in (26) are fulfilled when

$$g_{n,k+1} := \sqrt{n}G_n^*(k/n) = \sqrt{n}\big(G(k/n) - G((k-1)/n)\big), \quad k=0,1,\ldots,n.$$

Remark 2. *Note that, under condition* (28) *by Fubini's Theorem,*

$$\int_0^1 \frac{|G(u)|}{u}du \le \int_0^1\int_0^u |dG(v)|\frac{du}{u} = \int_0^v |dG(v)| \int_v^t \frac{du}{u} = \int_0^1 \log\frac{1}{u}\ |dG(u)| < \infty.$$

Hence, the integral in (27) *converges absolutely and function* H *is (uniformly) continuous on* $[0,1]$ *under assumption* (28) *of Theorem 2. Similarly, for all* $t\in(0,1]$ *we may rewrite the integral in* (27) *in the following form*

$$(29)\qquad \int_0^t \frac{G(u)}{u}du = \int_0^t\int_0^u dG(v)\frac{du}{u} = \int_0^v dG(v)\int_v^t \frac{du}{u} = \int_0^t \log\frac{t}{v}\ dG(v).$$

Now we are going to obtain a partial case of Theorem 2 when

$$G(t) = \int_0^t g(u)du, \qquad t \in [0,1], \tag{30}$$

for some measurable function g such that

$$\int_0^1 v \log \frac{3}{v}\ |dg(v)| < \infty. \tag{31}$$

Note that condition (31) implies that for each $n > 1$ function g has a bounded variation on $[1/n, 1]$. We will also use the following assumption:

$$\widetilde{\rho}_n := |\widetilde{\rho}_{n,0}| \log 3n + \sum_{k=1}^{n} |\widetilde{\rho}_{n,k}| \log \frac{3n}{k} \to 0 \quad \text{with} \quad \widetilde{\rho}_{n,k} := \frac{g_{n,k+1}}{\sqrt{n}} - \frac{g(k/n)}{n}. \tag{32}$$

Corollary 2. *Suppose that conditions* (30) – (32) *are fulfilled together with* (24) *and* ($\mathcal{B}$). *Then all assumptions and assertions of Theorem 2 take place.*

It is easy to see that for any function g assumption (32) is satisfied when

$$g_{n,k+1} = g(k/n)/\sqrt{n} \quad \text{for all possible values of} \quad n \geq k \geq 0,\ n > 1. \tag{33}$$

Under condition (33), the partial case of Corollary 2, when g is a function of bounded variation on $[0,1]$, was proved in Bischoff (2016); and it is the key result there.

Remark 3. *Note that both assumptions in* (26) *hold when*

$$\rho_n := |g_{n,1}| \frac{\log 3n}{\sqrt{n}} + \sum_{k=1}^{n} |\rho_{n,k}| \log \frac{3n}{k} \to 0 \quad \textit{with } \rho_{n,k} := \frac{g_{n,k+1}}{\sqrt{n}} - G_n^*(k/n), \tag{34}$$

where $G_n^*(t) := G(t) - G(t-1/n)$ *for* $t \in [1/n, 1]$. *This fact follows from inequality* $\delta_n + \overline{\delta}_n \leq 2\rho_n$ *which will be proved in Lemma 6.*

When G is defined in (30) *consider assumption:*

$$\rho_n^* := \sum_{k=1}^{n} |\rho_{n,k}^*| \log \frac{3n}{k} \to 0 \qquad \textit{with} \quad \rho_{n,k}^* := \frac{g(k/n)}{n} - G_n^*(k/n). \tag{35}$$

Note that $g_{n,1}/\sqrt{n} = \widetilde{\rho}_{n,0} + g(0)/n$ *and* $\rho_{n,k} = \widetilde{\rho}_{n,k} + \rho_{n,k}^*$ *for* $k = 1, \ldots, n$. *Hence,*

$$\rho_n \leq |g(0)| \frac{\log 3n}{n} + \widetilde{\rho}_n + \rho_n^*,$$

and condition (34) *follows from* (32) *and* (35).

2.5. **On estimation of σ.** It is well known that under assumptions ($\mathcal{B}$) and ($\mathcal{A}_x$), when Hypothesis $\mathcal{H}_0$ is true, the following statistic

$$s_n^2 := \sum\nolimits_{i=1}^{N(n)} \big(Y_{n,i} - x_{n,i}\widehat{\beta}_{n,n}\big)^2 \Big/ \big(N(n) - 1\big), \qquad N(n) > 1,$$

is the classical unbiased estimator of the unknown σ^2. It is based on the classical residuals $\{Y_{n,i} - x_{n,i}\widehat{\beta}_{n,n}\}$. On the other hand, by Property A, under the same assumptions, the next statistic

$$\widehat{\sigma}_n^2 := \sum\nolimits_{i=1}^{m(n)} w_{n,i}^2 \Big/ m(n), \qquad m(n) \geq 1, \tag{36}$$

is also an unbiased estimator of the unknown σ^2. Since main results of the present paper are based on recursive residuals $\{w_{n,i}\}$, it is natural to consider properties of the estimator from (36) which is also defined by $\{w_{n,i}\}$.

First of all, we are interested in the conditions under which estimator $\widehat{\sigma}_n$ is consistent, i.e.

$$\widehat{\sigma}_n \xrightarrow{P} \sigma, \qquad \text{when Hypothesis } \mathcal{H}_0 \text{ is true.} \tag{37}$$

But by equalities in (9) we immediately obtain that the next condition

$$\widetilde{\sigma}_n^2 := \sum_{i=1}^{m(n)} \eta_{n,i}^2 \Big/ m(n) \xrightarrow{P} \sigma^2, \qquad \text{as} \quad m(n) \to \infty, \tag{38}$$

is sufficient for the property (37). Note, that under assumptions of Property B convergence (38) is evident. Our conjecture is that convergence (38) also holds under assumptions from Property C.

A more difficult task is to find conditions under which estimator $\widehat{\sigma}_n$ is consistent also in cases when Hypothesis $\mathcal{H}_0$ is not true. That is when

$$\widehat{\sigma}_n \xrightarrow{P} \sigma, \qquad \text{for the observations from (1).} \tag{39}$$

In this case the following property will be useful

Proposition 2. *Suppose that Assumption* $(\mathcal{A}_x)$ *holds together with condition* (38) *and, in addition,*

$$\check{h}_n^2 := \sum_{i=1}^{m(n)} h_{n,i}^2 \Big/ m(n) \to 0, \qquad \textit{as} \quad m(n) \to \infty. \tag{40}$$

Then convergence (39) *takes place.*

In particular, condition (40) *holds under assumptions* (28) *and* (34) *with continuous function* G.

Hence, (40) is also satisfied under conditions of Corollary 2.

2.6. **On applications of main results.** Thus, if (37) holds (for some estimator $\widehat{\sigma}_n$) together with convergence (11) (or (13)), then

$$f(W_n/\widehat{\sigma}_n) \Rightarrow f(B) \qquad \text{(when Hypothesis } \mathcal{H}_0 \text{ is true)}$$

for all continuous functionals $f : C[0,1] \to \mathbb{R}$. In the important partial case, when

$$\text{distribution function} \quad \mathbb{P}(f(B) \le z) \quad \text{is continuous,} \tag{41}$$

we have the following uniform convergence:

$$\sup_z \big| \mathbb{P}(f(W_n/\widehat{\sigma}_n) \le z) - \mathbb{P}(f(B) \le z) \big| \to 0 \quad (\text{under } \mathcal{H}_0).$$

Recall that the standard statistical criteria for testing Hypothesis $\mathcal{H}_0$ have the following form:

$$\text{reject } \mathcal{H}_0 \quad \Longleftrightarrow \quad f(W_n/\widehat{\sigma}_n) > z, \tag{42}$$

for appropriate functionals f with property (41). This test has a desired asymptotic size $\alpha \in (0,1)$ when z is found from equation:

$$\lim_{n\to\infty} \mathbb{P}(f(W_n/\widehat{\sigma}_n) > z \,|\mathcal{H}_0) = \mathbb{P}(f(B) > z) = \alpha.$$

For example, we may use

$$f(W_n/\widehat{\sigma}_n) = \sup_{t\in[0,1]} \frac{|W_n(t)|}{\widehat{\sigma}_n} = \max_{1\le k\le m(n)} \frac{|w_{n,1} + \cdots + w_{n,k}|}{\widehat{\sigma}_n \sqrt{m(n)}},$$

or $f(\cdot) = f_0(\cdot)$ in one-sided case, where

$$f_0(W_n/\widehat{\sigma}_n) = \sup_{t\in[0,1]} \frac{W_n(t)}{\widehat{\sigma}_n} = \max_{1\le k\le m(n)} \frac{w_{n,1} + \cdots + w_{n,k}}{\widehat{\sigma}_n \sqrt{m(n)}}.$$

See Theorem 3 in Bischoff (2016) for an example.

For the criterion from (42) the probability of the type II error is equal to $\mathbb{P}(f(W_n/\widehat{\sigma}_n) \le z)$, where we do not assume that Hypothesis $\mathcal{H}_0$ is true. To deal with this probability suppose that convergence (21) is proved (by Corollary 1 or Theorem 2) and that property (39) holds (for some estimator $\widehat{\sigma}_n$). Then

$$W_n(\cdot)/\widehat{\sigma}_n \Longrightarrow B(\cdot) + H(\cdot)/\sigma \quad \text{in} \quad C[0,1].$$

Hence, for all continuous functionals $f : C[0,1] \to \mathbb{R}$ and each real z

$$\limsup_{n\to\infty} \mathbb{P}(f(W_n/\widehat{\sigma}_n) \le z) \le \mathbb{P}(f(B + H/\sigma) \le z),$$
$$\liminf_{n\to\infty} \mathbb{P}(f(W_n/\widehat{\sigma}_n) \le z) \ge \mathbb{P}(f(B + H/\sigma) < z).$$

Thus, we have shown how to approximate the probability $\mathbb{P}(f(W_n/\widehat{\sigma}_n) \le z)$ of the type II error using the results of the present paper.

For an example of such approach see Theorem 4 in Bischoff (2016), where functional $f(\cdot) = f_0(\cdot)$ was considered in the case when $\widehat{\sigma}_n = \sigma = 1$ is a known parameter. As a future research direction, the author is going to present statistical examples based on the more accurate formula (17).

3. Proofs of Proposition 1 and Theorem 1

3.1. **Proof of Proposition 1.** Denote by π_n the Prokhorov distance between the distributions of the processes $\widetilde{W}_n(\cdot)$ and $B(\cdot)$ in the space $C[0,1]$. Definition and properties of the Prokhorov distance may be found, for example, in Prokhorov (1956) or Billingsley (1999). In particular, $\pi_n \to 0$ when (13) takes place.

Next, by the powerful Theorem of Strassen (1965), on some rich probability space we may construct a standard Wiener process $B_n^*(\cdot)$ together with a process W_n^* such that W_n^* is identically distributed with $\widetilde{W}_n$ and, in addition, the next inequality holds:

$$\mathbb{P}\big(\|W_n^*(\cdot) - \sigma B_n^*(\cdot)\| > \pi_n\big) \le \pi_n. \tag{43}$$

But as it follows from the work of Skorohod (1976), on this rich probability space we may also construct a random variable ν_n^* such that (a^*) it is uniformly distributed over $[0,1]$; (b^*) it is independent of $W_n^*(\cdot)$; (c^*) there exists a function Ψ_n with the following property:

$$B_n^*(\cdot) = \Psi_n\big(W_n^*(\cdot), \nu_n^*\big).$$

Now introduce a random variable $\widetilde{\nu}$ such that $(\tilde{a})$ it is uniformly distributed over $[0,1]$ and $(\tilde{b})$ it is independent of the investigated process $\widetilde{W}_n(\cdot)$. Define

$$B_n(\cdot) = \Psi_n\big(\widetilde{W}_n(\cdot), \widetilde{\nu}\big).$$

It is clear that the triplets $\big(B_n(\cdot), \widetilde{W}_n(\cdot), \widetilde{\nu}\big)$ and $\big(B_n^*(\cdot), W_n^*(\cdot), \nu_n^*\big)$ are identically distributed. Hence,

$$\mathbb{P}\big(\|\widetilde{W}_n(\cdot) - \sigma B_n(\cdot)\| > \pi_n\big) = \mathbb{P}\big(\|W_n^*(\cdot) - \sigma B_n^*(\cdot)\| > \pi_n\big). \tag{44}$$

Since $\pi_n \to 0$ when convergence (13) takes place, the desired assertion (14) of Proposition 1 follows immediately from (43) and (44).

3.2. **Key Representations.** Remind that random variables $\{\eta_{n,i}\}$ and numbers $\{h_{n,i}\}$ were defined in (8) and (15).

Lemma 1. *Under Assumption* $(\mathcal{A}_x)$ *the following representations hold*

$$w_{n,i} = \eta_{n,i} + h_{n,i}, \qquad i = 1, 2, \dots, m(n), \quad n \geq n_0; \tag{45}$$

$$W_n(\cdot) = \widetilde{W}_n(\cdot) + H_n(\cdot), \qquad n \geq n_0. \tag{46}$$

In particular, for all $n \geq n_0$ *properties* (9) *and* (12) *take place.*

Proof. Substituting (1) into (5) we obtain for $k \geq r(n)$ that

$$\widehat{\beta}_{n,k} := \sum_{j=1}^{k} \frac{x_{n,j} Y_{n,j}}{v_k^2} = \sum_{j=1}^{k} \frac{x_{n,j} g_{n,j}}{v_k^2} + \sum_{j=1}^{k} \frac{x_{n,j} \varepsilon_{n,j}}{v_k^2} = \frac{\widetilde{h}_{n,k}}{v_k^2} + \frac{\widetilde{\eta}_{n,k}}{v_k^2},$$

where numbers $\widetilde{h}_{n,k}$ and $\widetilde{\eta}_{n,k}$ were defined in (15) and (8). After that, from (6) and (7) we have for $k > r(n)$ that

$$\widehat{\varepsilon}_{n,k} := Y_{n,k} - x_{n,k}\widehat{\beta}_{n,k-1} = g_{n,k} + \varepsilon_{n,k} - \frac{x_{n,k}\widetilde{h}_{n,k-1}}{v_{n,k-1}^2} - \frac{x_{n,k}\widetilde{\eta}_{n,k-1}}{v_{n,k-1}^2},$$

$$w_{n,k-r(n)} := \frac{v_{n,k-1}}{v_{n,k}}\widehat{\varepsilon}_{n,k} = \frac{v_{n,k-1}}{v_{n,k}}\Big(g_{n,k} + \varepsilon_{n,k} - \frac{x_{n,k}\widetilde{h}_{n,k-1}}{v_{n,k-1}^2} - \frac{x_{n,k}\widetilde{\eta}_{n,k-1}}{v_{n,k-1}^2}\Big)$$

$$= \frac{v_{n,k-1}}{v_{n,k}}\Big(g_{n,k} - \frac{x_{n,k}\widetilde{h}_{n,k-1}}{v_{n,k-1}^2} + \varepsilon_{n,k} - \frac{x_{n,k}\widetilde{\eta}_{n,k-1}}{v_{n,k-1}^2}\Big) = h_{n,k-r(n)} + \eta_{n,k-r(n)}.$$

The latter equality follows from (15) and (8) with $k = i + r(n)$, and it implies (45).

Substituting now equalities from (45) into the general definition (10) we immediately obtain (46), if only take into account corresponding definitions (11), (12) and (16) of $W_n(\cdot)$, $\widetilde{W}_n(\cdot)$ and $H_n(\cdot)$.

Next, when $\mathcal{H}_0$ is true, we may substitute (2) into (15) to obtain

$$\frac{x_{n,r(n)+i}}{v_{n,r(n)+i-1}^2}\widetilde{h}_{n,r(n)+i-1} = \frac{x_{n,r(n)+i}}{v_{n,r(n)+i-1}^2}\sum_{j=1}^{r(n)+i-1}\beta x_{n,j}^2$$

$$= \frac{x_{n,r(n)+i}}{v_{n,r(n)+i-1}^2}\beta v_{n,r(n)+i-1}^2 = x_{n,r(n)+i}\beta = g_{n,r(n)+i}.$$

Thus, we have from (15) that $h_{n,i} = 0$ under $\mathcal{H}_0$, and (9) follows from (45). In addition, $H_n(\cdot) = 0$ by (16) and (12) is a partial case of (46). □

Remark 4. *From definition* (8) *it is easy to see that each* $\eta_{n,i}$ *is a function only of random variables* $\varepsilon_{n,1}, \dots, \varepsilon_{n,N(n)}$ *and of numbers, used in Assumption* $(\mathcal{A}_x)$*. Hence, the joint distribution of random variables* $\eta_{n,1}, \dots, \eta_{n,m(n)}$ *is a function of the joint distribution of* $\varepsilon_{n,1}, \dots, \varepsilon_{n,N(n)}$*. Thus, under Assumption* $(\mathcal{A}_Y)$*, these joint distributions do not depend on the choice of numbers* $g_{n,1}, \dots, g_{n,N(n)}$*; in particular, they also do not depend on the choice of non-random* $h_{n,1}, \dots, h_{n,N(n)}$ *which are constructed in* (15) *as functions of* $g_{n,1}, \dots, g_{n,N(n)}$*.*

Similarly, under Assumption $(\mathcal{A})$, *for each* $n \geq n_0$ *the distribution in* $C[0,1]$ *of the process* $\widetilde{W}_n(\cdot)$, *as a function of variables* $\eta_{n,1}, \ldots, \eta_{n,m(n)}$, *does not depend on the choice of the non-random* $H_n(\cdot)$, *as a function of* $h_{n,1}, \ldots, h_{n,m(n)}$.

3.3. **Proof of Theorem 1.** Let Assumption $(\mathcal{A})$ be satisfied and fix $n \geq n_0$. Then, by Proposition 1, relationship (14) holds for some $B_n(\cdot)$ and π_n. By Lemma 1 in this case identity (46) takes place. In particular, $\widetilde{W}_n(\cdot) = W_n(\cdot) - H_n(\cdot)$. Hence,

$$\big\|\widetilde{W}_n(\cdot) - \sigma B_n(\cdot)\big\| = \big\|W_n(\cdot) - H_n(\cdot) - \sigma B_n(\cdot)\big\|. \tag{47}$$

Next, by Remark 4, the distribution of the random process $\widetilde{W}_n(\cdot)$ does not depend on the choice of the function H_n. Thus, we have from (47) that for any H_n

$$\mathbb{P}\big(\|W_n(\cdot) - H_n(\cdot) - \sigma B_n(\cdot)\| > \pi_n \big| H_n\big) = \mathbb{P}\big(\|\widetilde{W}_n(\cdot) - \sigma B_n(\cdot)\| > \pi_n\big).$$

And the assertion of Theorem 1 follows.

4. Proofs of Theorem 2, Corollary 2 and Proposition 2

Everywhere below we suppose that conditions (24) and (28) take place for a continuous function G with bounded variation on $[0,1]$.

4.1. **Useful Representations.** First of all, we are going to obtain convenient representations for the numbers $H_n(k/n)$ introduced in (20), where $m(n) = n$.

Lemma 2. *For all* $n \geq k \geq 1$

$$\sqrt{n} H_n(k/n) = (1 + c_k^*) G_{n,k+1} - \sum\nolimits_{i=0}^{k} c_i G_{n,i+1}, \tag{48}$$

where numbers $\{c_i\}$ *and* $\{c_i^*\}$ *are such that* $c_0 = \sqrt{2}$ *and*

$$\forall\ i \geq 1 \qquad 0 < c_i < 1/i, \quad |c_i - 1/i| < 2/i^2 \quad \textit{and} \quad 0 < c_i^* < 1/i \leq 1. \tag{49}$$

Proof. To prove (48) consider an arbitrary positive integer $k \leq n$. Since $v_{n,k}^2 = k$ as it follows from (24) and (4), it is easy to see from (15) and (25) that

(50)

$$\widetilde{h}_{n,r(n)+i-1} = G_{n,i}, \quad h_{n,i} = \sqrt{\frac{i}{i+1}}\Big(g_{n,i+1} - \frac{G_{n,i}}{i}\Big) = \sqrt{\frac{i+1}{i}} g_{n,i+1} - \frac{G_{n,i+1}}{\sqrt{i(i+1)}},$$

for $i = 1, \ldots, n$. Hence, we have from (20) that

$$\sqrt{n} H_n(k/n) = \sum_{i=1}^{k} h_{n,i} = \sum_{i=1}^{k} \sqrt{\frac{i+1}{i}} g_{n,i+1} - \sum_{i=1}^{k} \frac{G_{n,i+1}}{\sqrt{i(i+1)}}. \tag{51}$$

On the other hand, $g_{n,i+1} = G_{n,i+1} - G_{n,i}$ for $i = 1, \ldots, n$. Hence,

$$\sum_{i=1}^{k} \sqrt{\frac{i+1}{i}} g_{n,i+1} = \sqrt{\frac{k+2}{k+1}} G_{n,k+1} + \sum_{i=1}^{k} \widetilde{c}_i G_{n,i+1} - \sqrt{2} G_{n,1}, \tag{52}$$

where for all $i \geq 1$

$$\widetilde{c}_i := \sqrt{\frac{i+1}{i}} - \sqrt{\frac{i+2}{i+1}} = \frac{i+1-\sqrt{i^2+2i}}{\sqrt{(i+1)i}} = \frac{1}{\sqrt{(i+1)i}(i+1+\sqrt{i^2+2i})}.$$

So, $\widetilde{c}_i < \frac{1}{i(i+1)}$. Substituting (52) into (51) we obtain (48) with $\sqrt{\frac{k+2}{k+1}} = 1 + c_k^*$ and

$$\frac{1}{i} > c_i := \frac{1}{\sqrt{(i+1)i}} - \widetilde{c}_i > \frac{1}{i+1} - \frac{1}{i(i+1)} = \frac{1}{i} - \frac{2}{i(i+1)}, \quad i, k \geq 1. \tag{53}$$

To prove (49) consider arbitrary $k = i \geq 1$. We have:

$$0 < c_i^* := \sqrt{\frac{i+2}{i+1}} - 1 = \frac{\sqrt{i+2} - \sqrt{i+1}}{\sqrt{i+1}} = \frac{1}{(\sqrt{i+2} + \sqrt{i+1})\sqrt{i+1}} < \frac{1}{i} \leq 1.$$

Thus, we have proved the last group of inequalities in (49). But other inequalities in (49) immediately follows from (53). □

For any $n \geq k \geq 1$ introduce notation:

$$J_{n,k} := \sum\nolimits_{i=1}^{k} \frac{G(i/n)}{i} \quad \text{and} \quad J_{n,k}^* := \sum\nolimits_{i=1}^{k} c_i G(i/n) - c_k^* G(k/n). \tag{54}$$

Lemma 3. *For all* $n \geq 1$

$$\max_{1 \leq k \leq n} |\delta_{n,k}^*| \leq 4\overline{\delta}_n + \delta_n, \quad \textit{where} \quad \delta_{n,k}^* := H_n(k/n) - G(k/n) + J_{n,k}^*; \tag{55}$$

$$\max_{1 \leq k \leq n} |J_{n,k}^* - J_{n,k}| \leq \widetilde{\Delta}_n(G) + 2\Sigma_n^*, \quad \textit{where} \quad \Sigma_n^* := \sum\nolimits_{i=1}^{n} \frac{|G(i/n)|}{i^2}. \tag{56}$$

Proof. First, using notation $\delta_{n,k} := G_{n,k+1}/\sqrt{n} - G(k/n)$ introduced in (25) we obtain from (48) and (54) that for $n \geq k \geq 1$

$$J_{n,k}^* - J_{n,k} = \sum\nolimits_{i=1}^{k} (c_i - 1/i) G(i/n) - c_k^* G(k/n),$$
$$\delta_{n,k}^* = (1 + c_k^*)\delta_{n,k} - \sum\nolimits_{i=0}^{k} c_i \delta_{n,i}, \quad \text{where} \quad G(0) = 0.$$

Second, by estimates from (49) we have that for all possible $k \leq n$

$$|J_{n,k}^* - J_{n,k}| \leq \sum\nolimits_{i=1}^{k} \frac{2}{i^2} |G(i/n)| + \frac{1}{k} |G(k/n)| \leq \Sigma_n^* + \frac{|G(k/n)|}{k}, \tag{57}$$

$$|\delta_{n,k}^*| \leq 2|\delta_{n,k}| + c_0|\delta_{n,0}| + \sum\nolimits_{i=1}^{k} \frac{|\delta_{n,i}|}{i} \leq 4\overline{\delta}_n + \sum\nolimits_{i=1}^{n} \frac{|\delta_{n,i}|}{i} = 4\overline{\delta}_n + \delta_n,$$

where we also use notations from (26). Now inequality (55) follows.

Next, using notation $\widetilde{\Delta}_n(\cdot)$ introduced in (22) we get:

$$|G(k/n)| \leq \sum\nolimits_{i=1}^{k} |G\big(\tfrac{i}{n}\big) - G\big(\tfrac{i-1}{n}\big)| \leq k\widetilde{\Delta}_n(G), \quad k = 1, \ldots, n.$$

Substituting this estimate into (57), we arrive to (56). □

4.2. **Auxiliary lemmas.** Below we need several elementary facts.

Lemma 4. *Let* $a(\cdot)$ *and* $F(\cdot)$ *be functions on* $[0,1]$. *Then*

$$\int_U^T \big|a(t)\big(F(T) - F(t)\big)\big| dt \leq \int_{U+0}^T A(v)|dF(v)| \leq A(T) \int_{U+0}^T |dF(v)| \tag{58}$$

for any real numbers $0 \leq U < T \leq 1$, *where* $A(v) := \int_U^v |a(t)| dt$.

Indeed, by Fubini's Theorem,

$$\int_U^T \big|a(t)\big(F(T) - F(t)\big)\big| dt \leq \int_{U+0}^T |a(t)| \int_t^T |dF(v)| dt = \int_{U+0}^T |dF(v)| \int_U^v |a(t)| dt.$$

Lemma 5. *Let $F(\cdot)$ be non-decreasing function on $[0,1]$ such that $F(1) < \infty$ and $F(+0) = F(0) = 0$. And suppose that for all $n \ge 1$ we are given numbers $a_{n,1}, \dots, a_{n,n}$ with the next property:*

$$\forall\ n \ge k \ge 1 \quad \sum\nolimits_{i=1}^{k} |a_{n,i}| \le F(k/n). \tag{59}$$

Then

$$\sum\nolimits_{i=1}^{n} \frac{|a_{n,i}|}{i^2} \le F(1/\sqrt{n}) + F(1)/\sqrt{n} \to 0. \tag{60}$$

Indeed, choose integers $l = l(n)$ such that $n \ge l = l(n) \ge \sqrt{n} \ge l-1 \ge 0$. Then

$$\sum_{i=1}^{n} \frac{|a_{n,i}|}{i} \le \sum_{i=1}^{l-1} |a_{n,i}| + \frac{1}{l}\sum_{i=l}^{n} |a_{n,i}| \le F\big(\tfrac{l-1}{n}\big) + \tfrac{F(1)}{l} \le F(1/\sqrt{n}) + \tfrac{F(1)}{\sqrt{n}} \to 0.$$

Lemma 6. *For any numbers $a_1, \dots, a_n$ and each $k \ge 1$*

$$\sum_{i=1}^{k} \frac{1}{i}\Big|\sum_{j=1}^{i} a_j\Big| \le \sum_{i=1}^{k} \frac{1}{i}\sum_{j=1}^{i} |a_j| \le \sum_{i=1}^{k} |a_i| \log\frac{3k}{i}. \tag{61}$$

In particular, for notations introduced in (26) *and* (34) *the following relations hold:*

$$\forall\ n \ge 1 \quad \delta_n \le \rho_n \qquad \text{and} \qquad \overline{\delta}_n \le |g_{n,1}|/\sqrt{n} + \sum_{k=1}^{n} |\rho_{n,k}| \le \rho_n. \tag{62}$$

Proof. For $k \ge j \ge 1$ introduce numbers:

$$K_{k,j} := \sum_{i=j}^{k} \frac{1}{i} \le \sum_{i=j}^{k} \int_{i-1/2}^{i+1/2} \frac{dx}{x} = \int_{j-1/2}^{k+1/2} \frac{dx}{x} = \log\frac{2k+1}{2j-1} \le \log\frac{3k}{j}.$$

After that for any numbers $a_1, \dots, a_k$ we obtain (61) because:

$$\sum_{i=1}^{k} \frac{1}{i}\Big|\sum_{j=1}^{i} a_j\Big| \le \sum_{i=1}^{k} \frac{1}{i}\sum_{j=1}^{i} |a_j| = \sum_{j=1}^{k} |a_j| K_{k,j} \le \sum_{j=1}^{k} |a_j| \log\frac{3k}{j}.$$

Next, it is easy to see from (25) and (34) that $\delta_{n,k} = g_{n,1}/\sqrt{n} + \sum_{i=1}^{k} \rho_{n,i}$ for $1 \le k \le n$. Hence, the last estimate in (62) is evident since $\log(3n/i) \ge \log 3 > 1$ for $n \ge i \ge 1$. Applying after that (61) with $a_1 := g_{n,1}/\sqrt{n} + \rho_{n,1}$ and $a_i := \rho_{n,i}$ when $i = 2, \dots, n$, we obtain, using notations from (26), that $\delta_n \le \rho_n$ and $\overline{\delta}_n \le \rho_n$. □

4.3. Approximations of integral sums.

For $t \in (0,1]$ introduce functions:

$$J(t) := \int_0^t \frac{G(u)}{u} du, \quad \overline{G}(t) := \int_0^t |dG(u)|, \quad Q(t) := \int_0^t \log\frac{3}{u} d\overline{G}(u). \tag{63}$$

And let $\overline{G}_n^*(t) := \overline{G}(t) - \overline{G}(t - 1/n)$ for $t \in [1/n, 1]$.

Lemma 7. *For all $n \ge 1$*

$$\max_{1 \le k \le n} \big|J_{n,k} - J(k/n)\big| \le \Sigma_n := \Sigma_n^* + |J(1/n)| + 2\sum\nolimits_{i=1}^{n} \frac{\overline{G}_n^*(i/n)}{i}. \tag{64}$$

Proof. For any $n \ge k \ge 1$ we have from (54) and (63) that

$$J_{n,k} - J(k/n) = \sum_{i=1}^{k} \left(\frac{G(i/n)}{i} - \int_{(i-1)/n}^{i/n} \frac{G(v)}{v} dv \right) = \sum_{i=1}^{k} (a_{n,i} + b_{n,i}), \tag{65}$$

where $b_{n,1} := G(1/n)$, $a_{n,1} := -J(1/n)$ and for $i \ge 2$

$$a_{n,i} := \int_{(i-1)/n}^{i/n} \frac{G(i/n) - G(v)}{v} dv, \quad b_{n,i} := \frac{G(i/n)}{i} - \int_{(i-1)/n}^{i/n} \frac{G(i/n)}{v} dv.$$

It is easy to see, that $b_{n,i} = \check{c}_i G(i/n)$ for $i \ge 2$ with $\check{c}_i := 1/i + \log(1 - 1/i)$. Hence,

$$\sum_{i=1}^{n} |b_{n,i}| \le \sum_{i=1}^{n} \frac{|G(i/n)|}{i^2} = \Sigma_n^*, \text{ since } |\check{c}_i| = \sum_{m=2}^{\infty} \frac{1}{m i^m} < \sum_{m=2}^{\infty} \frac{1}{2 i^2 2^{m-2}} = \frac{1}{i^2}. \tag{66}$$

To estimate $a_{n,i}$ for $i \ge 2$ we are going to apply Lemma 4 with $F = G$ and $a(v) = 1/v$. As a result, we obtain for $T = i/n = U + 1/n$ that

$$|a_{n,i}| \le \int_{(i-1)/n}^{i/n} |dG(v)| A(T) \text{ with } A(T) = \int_U^T \frac{dv}{v} \le \int_U^T \frac{dv}{U} = \frac{1}{i-1} \le \frac{2}{i}.$$

Thus, $|a_{n,i}| \le 2\overline{G}_n^*(i/n)/i$ for $i \ge 2$ with $|a_{n,1}| = |J(1/n)|$. Substituting this bounds and (66) into representation (65), we obtain (64). □

Lemma 8. *For all* $n \ge k \ge 1$

$$\sum\nolimits_{i=1}^{k} \frac{|G(i/n)|}{i} \le Q(k/n), \quad \textit{so that} \quad \sum\nolimits_{i=1}^{n} \frac{|G(i/n)|}{i} \le Q(1) < \infty. \tag{67}$$

In addition, $0 \le \Sigma_n^* \le \Sigma_n \to 0$.

Proof. Since $|G(i/n)| \le \overline{G}(i/n) = \sum_{j=1}^{i} \overline{G}_n^*(j/n)$, we have from Lemma 6 with $a_{n,i} = \overline{G}_n^*(i/n)$ that (67) is true because

$$\sum_{i=1}^{k} \frac{|G(i/n)|}{i} \le \sum_{i=1}^{k} \overline{G}_n^*(i/n) \log \frac{3k}{i} \le \sum_{i=1}^{k} \int_{(i-1)/n}^{i/n} \log \frac{3}{v} d\overline{G}(v) = \int_0^{k/n} \log \frac{3}{v} d\overline{G}(v).$$

Now note, that functions $J(\cdot)$, $\overline{G}(\cdot)$ and $Q(\cdot)$ are continuous because $G(\cdot)$ is continuous. By (67), we may apply Lemma 5 with numbers $a_{n,i} = |G(i/n)|/i$ and function $F(\cdot) = Q(\cdot)$; as a result we obtain that $\Sigma_n^* \to 0$. Next, using again Lemma 6, but with $F(\cdot) = \overline{G}(\cdot)$ and $a_{n,i} = \overline{G}_n^*(i/n) = \overline{G}\big(\frac{i}{n}\big) - \overline{G}\big(\frac{i-1}{n}\big)$, we find that $\sum_{i=1}^{n} \overline{G}_n^*(i/n)/i \to 0$. Thus, $\Sigma_n \to 0$. □

4.4. **Proof of Theorem 2.** Note that $H(\cdot) = G(\cdot) - J(\cdot)$ as it follows from (27) and (63). Hence, for all integers $n \ge k \ge 1$,

$$\begin{aligned}\Delta_{n,k} := H_n(k/n) - H(k/n) &= H_n(k/n) - G(k/n) + J(k/n) \\ &= \delta_{n,k}^* + (G_{n,k}^* - G_{n,k}) + (G_{n,k} - G(k/n)),\end{aligned}$$

where we use notations introduced in (55) and (54). So,

$$\max_{1\le k\le n} |\Delta_{n,k}| \le \max_{1\le k\le n} |\delta_{n,k}^*| + \max_{1\le k\le n} |G_{n,k}^* - G_{n,k}| + \max_{1\le k\le n} |G_{n,k} - G(k/n)|.$$

Applying now statements of Lemmas 3 and 7, we obtain:

$$\max_{1\le k\le n} |\Delta_{n,k}| \le 4\overline{\delta}_n + \delta_n + \widetilde{\Delta}_n(G) + 2\Sigma_n^* + \Sigma_n \to 0.$$

Here convergence follows from condition (26) and the second assertion of Lemma 8.

Thus, $\Delta_n = \max_{1\le k\le n}|\Delta_{n,k}| \to 0$ and, as a result, the main assertions (21) and (22) of Theorem 2 follow from Corollary 1 when function H is defined by formula (27).

4.5. **Proof of Proposition 2.** For any vector of the form $(a_{n,1},\dots,a_{n,m(n))}$ introduce the simplified notations:

$$a_{n,\bullet} := (a_{n,1},\dots,a_{n,m(n)}) \quad \text{and} \quad ||a_{n,\bullet}||_*^2 := \sum\nolimits_{i=1}^{m(n)} a_{n,i}^2.$$

Since $||a_{n,\bullet}||_*$ is a norm of the vector $a_{n,\bullet}$, we have from equalities in (45) that

$$\big|||w_{n,\bullet}||_* - ||\eta_{n,\bullet}||_*\big| = \big|||\eta_{n,\bullet} + h_{n,\bullet}||_* - ||\eta_{n,\bullet}||_*\big| \le ||h_{n,\bullet}||_*.$$

Now we may rewrite the last relationship in the following form:

$$\big|\widehat{\sigma}_n - \widetilde{\sigma}_n\big| = \big|||w_{n,\bullet}/\sqrt{n}||_* - ||\eta_{n,\bullet}/\sqrt{n}||_*\big| \le ||h_{n,\bullet}/\sqrt{n}||_* = \check{h}_n \to 0,$$

where we use condition (40) and notations from (36) and (38). So, convergence $\widetilde{\sigma}_n \xrightarrow{P} \sigma$ implies that $\widehat{\sigma}_n \xrightarrow{P} \sigma$ and the first statement of Proposition 2 is proved.

Now we need the following auxiliary assertion.

Lemma 9. *Under assumptions* (28) *and* (34) *with continuous functions* G

$$\forall\ n\ge 1 \quad \max_{1\le k\le n} |h_{n,k}|/\sqrt{n} \le 2\widetilde{\Delta}_n(G) + 2\rho_n \to 0, \tag{68}$$

$$\sum\nolimits_{i=1}^{n} |h_{n,i}|/\sqrt{n} \le 2\overline{G}(1) + Q(1) + 3\rho_n \to 2\overline{G}(1) + Q(1) < \infty. \tag{69}$$

Proof. We see from (50) that for $i = 1,\dots,n$

$$|h_{n,i}| \le \big|g_{n,i+1} - G_{n,i}/i\big| \le 2\max_{0\le i\le n}|g_{n,i+1}|, \quad |h_{n,i}| \le 2|g_{n,i+1}| + \frac{|G_{n,i+1}|}{i}. \tag{70}$$

Recall that we have from (25) and (34) that for $i = 1,\dots,n$

$$\frac{|G_{n,i+1}|}{\sqrt{n}} \le |G(i/n)| + |\delta_{n,i}| \quad \text{and} \quad \frac{|g_{n,i+1}|}{\sqrt{n}} \le |G_n^*(i/n)| + |\rho_{n,i}|, \tag{71}$$

and we may put $\rho_{n,0} := g_{n,1}|/\sqrt{n}$. Substituting these inequalities into (70) and taking into account (62) we obtain that

$$\sum\nolimits_{i=1}^{n} \frac{|h_{n,i}|}{\sqrt{n}} \le 2\sum\nolimits_{i=1}^{n} |G_n^*(i/n)| + \sum\nolimits_{i=1}^{n} \frac{|G(i/n)|}{i} + 3\rho_n, \tag{72}$$

$$\max_{1\le i\le n}\frac{|h_{n,i}|}{\sqrt{n}} \le 2\max_{0\le i\le n}\frac{|g_{n,i+1}|}{\sqrt{n}} \le 2\max_{1\le i\le n}|G_n^*(i/n)| + 2\max_{0\le i\le n}|\rho_{n,i}|$$
$$\le 2\max_{1\le i\le n}\big|G\big(\tfrac{i}{n}\big) - G\big(\tfrac{i-1}{n}\big)\big| + 2\rho_n \le 2\widetilde{\Delta}_n(G) + 2\rho_n \to 0,$$

where notation $\widetilde{\Delta}_n(\cdot)$ was introduced in (22), and $\widetilde{\Delta}_n(G) \to 0$ for a continuous function G. So, relation (68) is proved.

Recall that numbers $G_n^*(\cdot)$ were introduced in Remark 3, so that

$$|G_n^*(i/n)| = \big|G\big(\tfrac{i}{n}\big) - G\big(\tfrac{i-1}{n}\big)\big| \le \textstyle\int_{(i-1)/n}^{i/n} |dG(t)|.$$

Hence,

$$\sum\nolimits_{i=1}^{n} |G_n^*(i/n)| \le \int_0^1 |dG(t)| = \overline{G}(1) < \infty.$$

Substituting this estimate, together with (67), into (72) we arrive at (69). □

At last, by using (68) and (69) we may verify assumption (40):

$$\check{h}_n^2 = \sum\nolimits_{i=1}^n \frac{h_{n,i}^2}{n} \le \max_{1\le i\le n} \frac{|h_{n,i}|}{\sqrt{n}} \cdot \sum\nolimits_{i=1}^n \frac{|h_{n,i}|}{\sqrt{n}} = o(1)\cdot O(1) \to 0.$$

Thus, for continuous functions G assumption (40) follows from (28) and (34).

4.6. **Proof of Corollary 2.** We first prove the following assertion.

Lemma 10. *Suppose that functions g and G satisfy conditions* (30) *and* (31)*. Then assumptions* (28) *and* (35) *also take place.*

Proof. We first apply Lemma 4 to function $a(t) = \log\frac{3}{t}$ and $F = g$. Then

$$A(v) = \int_0^v \log\frac{3}{u}du = v\log 3 - v\log v + v < 2v\log\frac{3}{v}, \qquad v>0.$$

$$A_* := \int_0^1 \log\frac{3}{t}\ |g(1)-g(t)|dt \le \int_0^1 A(v)|dg(v)| \le 2\int_0^1 v\log\frac{3}{v}|dg(v)| < \infty,$$

where we use assumption (31). Since $|g(t)| \le |g(1)| + |g(1)-g(t)|$, we have:

$$\int_0^1 \log\frac{3}{u}\ |dG(u)| = \int_0^1 \log\frac{3}{u}\ |g(u)|du \le \int_0^1 \log\frac{3}{u}\ |g(1)|du + A_* < \infty.$$

Thus, condition (28) follows from (31).

To prove (35) introduce notations:

$$\text{(73)}\qquad \Psi(t) := \int_{+0}^t v\log\frac{3}{v}\ |dg(v)|,\ \ t\in(0,1], \quad\text{and}\quad \Psi_n^*(t) := \Psi(t) - \Psi(t-1/n)$$

when $t\in[1/n,1]$. For $T = k/n = U + 1/n$ we may rewrite definition (35) in the next form:

$$\rho_{n,k}^* = \frac{g(T)}{n} - \int_U^T g(t)dt = \int_U^T (g(T)-g(t))dt.$$

Now we are going to apply Lemma 4 with function $a(\cdot) = 1$ and $A(v) = v - U \ge 0$ when $F = g$. First, $A(v) = v$ when $T = 1/n$; hence

$$\log(3n)|\rho_{n,1}^*| \le \int_0^{1/n} v\log\frac{3}{v}\ |dg(v)| = \Psi(1/n) = \Psi_n^*(1/n).$$

Second, if $k\ge 2$ then $nU = k-1 \ge 1$ and $A(T) = \frac{1}{n} \le \frac{v}{nU} = \frac{v}{k-1}$. Hence, by (58)

$$\log\frac{3n}{k}\ |\rho_{n,k}^*| \le \log\frac{3n}{k}\int_{(k-1)/n}^{k/n} \frac{v|dg(v)|}{k-1} \le \int_{(k-1)/n}^{k/n} v\log\frac{3}{v}\ \frac{|dg(v)|}{k-1} = \frac{\Psi_n^*(k/n)}{k-1}.$$

By substituting the last two estimates into (35) we obtain:

$$\rho_n^* \le \Psi_n^*(1/n) + \sum_{k=2}^n \frac{\Psi_n^*(k/n)}{k-1} \le 2\sum_{k=1}^n \frac{\Psi_n^*(k/n)}{k} \to 0.$$

Here convergence to 0 follows from Lemma 6, with $F = g$ and $a_{n,i} = \Psi_n^*(i/n) = \Psi\big(\frac{i}{n}\big) - \Psi\big(\frac{i-1}{n}\big)$. We have also used the fact that function Ψ is continuous on $[0,1]$ with $\Psi(+0) = 0$ due to (73).

So, condition (35) also holds under assumptions of the lemma. □

It follows from Remark 3 and Lemma 10 that Corollary 2 is a partial case of Theorem 2. Thus, all results of the paper are proved.

Acknowledgment. The author is sincerely grateful to Artyom Kovalevskii, who drew my attention to the paper of Brown, Durbin and Evans (1975). I am also very appreciative of the anonymous referees for their deep remarks which led to the improvement of the presentation of my results in the paper.

References

[1] P. Billingsley. *Convergence of probability measures*, Wiley, Chichester, 1999. Zbl 0944.60003
[2] W. Bischoff, *A functional central limit theorem for regression models*, Ann. Stat., **26**:4 (1998), 1398–1410. Zbl 0936.62072
[3] W. Bischoff, *On Designs for recursive least squares residuals to detect alternatives*, in Joachim Kunert et al (eds.), *mODa 11 - Advances in Model-Oriented Design and Analysis*, Proceedings of the 11th International Workshop in Model-Oriented Design and Analysis held in Hamminkeln, Germany, June 12-17, 2016, Springer, Cham, 2016, 37–45.
[4] R.L. Brown, J. Durbin, J.M. Evans, *Techniques for testing the constancy of regression relationships over time*, J. R. Stat. Soc. Ser. B, **37**:2 (1975), 149–192. Zbl 0321.62063
[5] M.D. Donsker, *An invariance principle for certain probability limit theorems*, Mem. Am. Math. Soc., **6**:6 (1951), 885–900. Zbl 0042.37602
[6] I.B. MacNeill, *Limit processes for sequences of partial sums of regression residuals*, Ann. Probab., **6**:4 (1978), 695–698. Zbl 0377.60028
[7] P.K. Sen, *Invariance principles for recursive residuals*, Ann. Stat., **10**:1 (1982), 307–312. Zbl 0484.62098
[8] Yu.V. Prokhorov, *Convergence of random processes and limit theorems in probability theory*, Teor. Veroyatn. Primen., **1**:2 (1956), 177–238. Zbl 0075.29001
[9] A.V. Skorohod, *On a representation of random variables*, Theory Probab. Appl., **21** (1977), 628–632. Zbl 0362.60004
[10] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Stat., **36**:2 (1965), 423–439. Zbl 0135.18701

Alexander Sakhanenko
Novosibirsk State University,
2, Pirogova str.,
Novosibirsk, 630090, Russia
Sobolev Institute of Mathematics,
4, Acad. Koptyug ave.,
630090, Novosibirsk, Russia.
Email address: aisakh@mail.ru