Estimating a change-point in two-phases regression model based on the shift of parameter estimates

To Van Ban[1], **Nguyen Thi Quyen**[2]

Abstract

In this paper, we propose a method to estimate the change-point of a two phases regression based on the shift of parameter estimates. For the uniform design, we show that the change-point and the parameter estimates converge almost surely to the true values. Some simulations show that our proposed method is more effective than the others.

Mathematics Subject Classification: 62J05; 65C60; 68U20
Keywords: change-point; regression; estimate; convergence

1 Introduction

The change-point problem in regression model is increasingly attracted the attention in statistical journals as well as its applications. We can mention some researches using the information criteria BIC in [1], the CUSUM procedure in [2], [3]. Kim and Siegmund [4] proposed likelihood ratio tests for a change in the intercept with the same slope, and a change in both the intercept and slope. The power of tests were surveyed in [3], [4], [5]. Applying empirical likelihood in the change-point problem has been studied in [6], [7], [8]. The outstanding advantage of the empirical likelihood is that the test is little dependent on the distribution of errors.

Normally, after concluding that the change-point exists, we have to consider how to estimate it. Bai [9] studied the robustness, the rate of the convergence and the asymptotic distribution of the change-point estimate based on the least squares method for multiple linear regression model. The asymptotic of the maximum likelihood estimator and of M-estimator were considered in [10], [11]. Diniz and Brochi [12] studied the robustness of some likelihood ratios for the simple linear regression model. Comparison four methods for estimating as Bayesian, Julious', grid-search and the segmented methods was carried out in [13]. In [7], when calculating the residuals, Liu and Qian used the fitting y_i at x_i with swapped least square estimates of the regression parameters, that is the parameter estimates of the first (the last) phase used to calculate the residuals of the last (the first) phase. By this way, the residuals are exaggerated and then this makes the rate of the convergence of estimates increase. However, Liu and Qian [7], Zhao, Wu and Chen [14] used the empirical likehood method to

[1]Depar. of Math., Basic Faculty, University of Transport Technology, Hanoi, Vietnam
Email: tvban_hvkt@yahoo.com.vn
[2]Math. Faculty, Hai Phong University, Haiphong, Vietnam
Email:quyentthp@yahoo.com

study these residuals and they took care of the robustness of the change-point estimate as well as parameter estimates by simulation only.

In this paper, we will use the shift of parameter estimates to calculate the residuals proposed by Liu and Qian in [7]. However, another method is proposed to estimate the change-point, from that the almost sure convergence of the change-point estimate and the parameter ones to the true values are obtained. The present paper is organized in the following way. In Section 2, the main results are given including the almost sure convergence of the change-point estimate and the parameter estimates of the model. Some simulations are presented in Section 3 where we compare the effect of the estimate of the time of the change using the proposed method with using Liu and Qian's one and with some others. The detailed proofs are presented in Section 4.

2 Main results

Consider a two-phases regression model

$$y_i = \begin{cases} \alpha_0 + \alpha_1 x_i + \epsilon_i & \text{if } 1 \le i \le k^*, \\ \beta_0 + \beta_1 x_i + \epsilon_i & \text{if } k^* < i \le n \end{cases} \tag{1}$$

where $\{x_i, i = 1, \dots, n\}$ belong to a closed interval $I = [a, b]$ and without loss of generality we can assume $a \le x_1 < \dots < x_n \le b$, $\{\epsilon_i\}$ is a sequence of independent errors with mean zero and $E(\epsilon_i^2) = \sigma_1^2$ for $i = 1, \dots, k^*$, $E(\epsilon_i^2) = \sigma_2^2$ for $i = k^* + 1, \dots, n$, the parameters α_0, α_1, β_0, β_1, σ_1, σ_2, k^* are unknown.

If $(\alpha_0, \alpha_1) = (\beta_0, \beta_1)$ then the model (1) is called no change. Otherwise, the model is said to have a change and k^* is called the time of the change. In this case, if two lines $y = \alpha_0 + \alpha_1 x$ and $y = \beta_0 + \beta_1 x$ meet at τ in $[x_{k^*}, x_{k^*+1})$ that is $\alpha_0 + \alpha_1 \tau = \beta_0 + \beta_1 \tau$, $\tau \in [x_{k^*}, x_{k^*+1})$, then the model function is called the continuous segmented one, the model is called continuous and τ is called the change-point. Otherwise, the model is called the discontinuous one. In this paper, we only consider the continuous model. Let $h = \beta_1 - \alpha_1$, the model (1) is rewritten as $y_i = f(x_i) + \epsilon_i$, $i = 1, \dots, n$, where $f(x)$ is defined by

$$f(x) = \alpha_0 + \alpha_1 x + h(x - \tau) I(x > \tau) \tag{2}$$

and $I(.)$ is the indicator function.

There are a lot of methods have been proposed to test the null hypothesis $H_0 : (\alpha_0, \alpha_1) = (\beta_0, \beta_1)$ (see [3], [4], [9]). If the null hypothesis is rejected, we need to estimate the time of the change k^* and the change-point τ.

We divide the observations into two groups. The first one contains k observations (x_i, y_i), $i = 1, \dots, k$ and let $\hat{\alpha}_{0k}, \hat{\alpha}_{1k}$ denote the least squares estimators of the parameters of the simple linear regression through the set of them. The last one contains the rest $n - k$ observations and the least squares estimators are $\hat{\beta}_{0k}, \hat{\beta}_{1k}$. Each group has to contain enough observations. Following [7], [9] the restricted range of k is $k_0 \le k \le n - k_0$, where k_0 is large enough. Follows [7], we put

$$\widetilde{e}_{ik} = \begin{cases} y_i - (\hat{\beta}_{0k} + \hat{\beta}_{1k} x_i) & \text{if } 1 \le i \le k, \\ y_i - (\hat{\alpha}_{0k} + \hat{\alpha}_{1k} x_i) & \text{if } k < i \le n. \end{cases} \tag{3}$$

It should be noted that the residuals $\widetilde{e}_{ik}$ are not the ordinary least squares fitting residuals but the residuals of fitting y_i at x_i with swapped least squares estimates of the regression parameters. The advantage of these residuals is that they should be close to ordinary residuals under null hypothesis, but they are exaggerated under alternative hypothesis $H_1 : (\alpha_0, \alpha_1) \neq (\beta_0, \beta_1)$. The time of the change k^* is estimated by

$$\hat{k}_n = \underset{k_0 \leq k \leq n-k_0}{\arg\max} \sum_{i=1}^{n} \tilde{e}_{ik}^2. \tag{4}$$

We will need the following assumptions:

*A*1. $x_1 < x_2 < \ldots < x_n$.

*A*2. $x_{k^*} \leq \tau < x_{k^*+1}$ that is the model function (2) is continuous and has a break at τ: $\alpha_0 + \alpha_1\tau = \beta_0 + \beta_1\tau$, $(\alpha_0, \alpha_1) \neq (\beta_0, \beta_1)$.

*A*3. Exist $k_0 : 1 < k_0 < n/2$ such as $k_0 < k^* < n - k_0$ and $k_0 \to \infty$ as $n \to \infty$.

*A*4. $x_{k^*} \to \tau$ as $n \to \infty$.

*A*5. Errors ϵ_i are independent, zero mean and $E(\epsilon_i^2) = \sigma_1^2 < \infty$ for $1 \leq i \leq k^*$ and $E(\epsilon_i^2) = \sigma_2^2 < \infty$ for $k^* < i \leq n$.

From $A1$, the design points x_i are ordered. For random design, such as x_i are i.i.d., we work with a fixed realization of the x_i's then they can be assumed to be in $[a, b]$ and reordered. Assumption $A2$ shows that x_{k^*} is on the left of τ and as near it as possible. With $A3$, we only consider the times of the change from k_0 to $n - k_0$. Following Bai [9], k_0 is chosen such that $k_0/n \approx C_0 \in (0, 0.5)$. The assumption $A4$ will be need to guarantee for the convergence of change-point estimates.

The rest of this Section is concentrated on surveying the convergence of $\hat{k}_n/n$. To simplify presentation, we consider the case of $[a, b] = [0, 1]$, uniform designs, furthermore, τ and x_{k^*} are coincident. Hence, the following assumptions are made instead of $A1 - A3$:

$A'1$. The designs x_i is uniform in $[0, 1] : x_i = i/n$, $i = 1, \ldots, n$.

$A'2$. The model function can be written as

$$f(x) = \alpha_0 + \alpha_1 x + h(x - x_{k^*})I(x > x_{k^*}),\ h \neq 0.$$

$A'3$. There exists $\tau_0 \in (0, 1/2)$ such that for $k_0 = \lfloor n\tau_0 \rfloor + 1$, where $\lfloor a \rfloor$ is integer part of a, $k_0 < k^* < n - k_0$.
From $A'3$ it is clear that $k_0 \to \infty$ if and only if $n \to \infty$.

Theorem 2.1. *If $A'1 - A'3$, $A4, A5$ hold, then $\frac{\hat{k}_n}{n} \xrightarrow{a.s} \tau$, where $\hat{k}_n$ is defined by (4).*

We note that $h = \beta_1 - \alpha_1$ in Theorem 2.1 is fixed. In the case that h depends on n, we get the following theorem.

Theorem 2.2. *The assertion of Theorem 2.1 still holds if $h = h_n$, as long as exist $C, D > 0$ and $0 < p < \frac{1}{2}$ such as $\frac{C}{n^p} \leq h_n \leq D, \forall n$.*

Moreover, if $\lim_{n\to\infty} h_n = h_\infty \neq 0$ then estimates $\hat{\alpha}_{0\hat{k}_n}, \hat{\alpha}_{1\hat{k}_n}$ and $\hat{\beta}_{0\hat{k}_n}, \hat{\beta}_{1\hat{k}_n}$ of the parameters of the simple linear regression through the set of the first $\hat{k}_n$ observations and the last $(n - \hat{k}_n)$ ones converge almost surely to α_0, α_1 and $\alpha_0 - h_\infty\tau, \alpha_1 + h_\infty$, respectively.

In the case $h \neq 0$ is fixed then $\{\hat{\alpha}_{0\hat{k}_n}\}$, $\{\hat{\alpha}_{1\hat{k}_n}\}, \{\hat{\beta}_{0\hat{k}_n}\}, \{\hat{\beta}_{1\hat{k}_n}\}$ converge almost surely to α_0, α_1, $\alpha_0 - h\tau$, $\alpha_1 + h$, respectively.

3 Some simulations

Following to Liu and Qian [7] we consider the model $y_i = h(x_i - x_{k^*})I(x_i > x_{k^*}) + \epsilon_i$, where $h = 2, h = 4$ represent moderate and large slope changes and all are under two error settings $N(0, 0.5^2)$ and centered log $N(0, 0.1^2)$. In [7] the effectiveness of methods is evaluated through the relative frequency RF of the deviation $d = |\hat{k} - k^*|$ less than or equal the fine tuned acceptable deviation $D = [(U - L)/A]$, where $U = n - L$, $L = \log^2 n$, A is the range of x_i.

Table 1: Simulation results using the ERL method and the proposed one (PR) based on 1,000 replications and uniform design on $[-3, 3]$, random design $X_i \sim N(0, 1)$

d	$\epsilon_i \sim N(0, 0.5^2)$								$\epsilon_i \sim \log N(0, 0.1^2)$							
	h=2				h=4				h=2				h=4			
	Uniform		Normal		Uniform		Normal		Uniform		Normal		Uniform		Normal	
	ELR	PR	ELR	PR	ELR	PR	ELR	PR	ELR	PR	ELR	PR	ELR	PR	ELR	PR
0	88	208	85	87	167	429	99	161	322	525	171	316	330	494	233	362
1	174	211	124	151	295	391	186	260	361	472	249	380	396	506	330	489
2	132	103	105	101	145	95	126	137	83	3	164	156	63	0	136	95
3	121	80	86	72	83	33	97	95	59	0	94	60	40	0	91	34
4	76	51	91	72	59	16	73	65	33	0	68	38	42	0	51	14
5	73	51	82	84	53	6	87	52	32	0	60	17	34	0	39	3
6	67	41	78	93	47	3	69	42	24	0	44	14	19	0	27	2
≤ 7	269	255	349	340	151	27	263	188	86	0	150	19	76	0	93	1
RF(‰)	515	602	400	411	690	948	508	653	825	1000	678	912	829	1000	790	980

For the design $x_i \sim$ i.i.d. $N(0, 1)$ as in [7] as well as $x_i = -3 + 6i/n, i = 1, 2, \ldots, n$ spread evenly in $[-3, 3]$, we take $n = 50, A = 6, L = 16, U = 34$ and $D = 3$. Using R software, the results of estimator by ERL and by our proposed method based on 1,000 replications are displayed in Table 1. These show that our proposed method is more effective than ELR. If $\epsilon_i \sim N(0, 0.1^2)$ then the similar results are deduced.

Now we reconsider the model (2) in the continuous case that the parameters of the model (2) are constrained so that $\alpha_0 + \alpha_1\tau = \beta_0 + \beta_1\tau$, $\tau \in [x_{k^*}, x_{k^*+1})$ and need to estimate the change point τ. First, we estimate $\hat{k}_n$ for k^* by our proposed method, from this the parameter estimates $\hat{\alpha}_{0\hat{k}_n}, \hat{\alpha}_{1\hat{k}_n}, \hat{\beta}_{0\hat{k}_n}, \hat{\beta}_{1\hat{k}_n}$ are obtained. The abscissa of the intersect of two lines $y = \hat{\alpha}_{0\hat{k}_n} + \hat{\alpha}_{1\hat{k}_n}x$ and $y = \hat{\beta}_{0\hat{k}_n} + \hat{\beta}_{1\hat{k}_n}x$ is $\tau^* = (\hat{\beta}_{0\hat{k}_n} - \hat{\alpha}_{0\hat{k}_n})/(\hat{\alpha}_{1\hat{k}_n} - \hat{\beta}_{1\hat{k}_n})$. Now, the estimate of τ is chosen by

$$\hat{\tau} = x_{\hat{k}_n} I(\tau^* \leq x_{\hat{k}_n}) + \tau^* I(x_{\hat{k}_n} < \tau^* < x_{\hat{k}_n+1}) + x_{\hat{k}_n+1} I(\tau^* \geq x_{\hat{k}_n+1}).$$

We use the parameter estimates of the model (2) constrained by $\alpha_0 + \alpha_1\hat{\tau} = \beta_0 + \beta_1\hat{\tau}$ to be the final estimates for parameters.

For the model $y_i = 3.5 + 0.5x_i + 1.0(x_i - 10)I(x_i > 10) + \epsilon_i$ in [13], where ϵ_i are independent, $\epsilon_i \sim N(0, 1.0)$ if $x_i \leq 10$, $\epsilon_i \sim N(0, 0.25)$ if $x_i > 10$, the covariate x

is generated according to the autoregressive model $x_i = 2.0 + 0.8x_{i-1} + v_i$, $x_0 \sim N(0, 81)$, $v_i \sim N(0, 100)$. It is run for 120 iterations, after which the first 40 iterations are discarded to avoid the influence of the choice of x_0; 500 replicated data sets were generated.

By our proposed method, the mean of estimates of τ is $\overline{\tau} = 9.777$ which is nearer $\tau = 10$ than 9.640 obtained in [13] by grid-seach method that was highly evaluated.

4 Proofs

Proof for Theorem 2.1

Without loss of generality we only consider the case of $h > 0$. Moreover, we can assume that $k^* = \lfloor n\tau \rfloor$. In order to prove the theorem, we only need to show that for a given small enough $\epsilon > 0$ then

$$P\Big\{\inf_{n>N} \frac{\hat{k}_n}{n} > \tau - \epsilon\Big\} \to 1,\ P\Big\{\sup_{n>N} \frac{\hat{k}_n}{n} < \tau + \epsilon\Big\} \to 1 \text{ as } N \to \infty. \tag{5}$$

Let $k_1 = \lfloor (\tau - \epsilon)n \rfloor$, it is clear that

$$\Big\{\max_{k_0 \le k \le k_1} \sum_{i=1}^{n} \widetilde{e}_{ik}^2 < \sum_{i=1}^{n} \widetilde{e}_{ik^*}^2\Big\} \subset \Big\{\frac{\hat{k}_n}{n} > \tau - \epsilon\Big\}. \tag{6}$$

To evaluate the probability of the left-hand side of (6), we need some lemmas. For $k_0 \le k \le n - k_0$, put

$$\begin{aligned} \overline{\epsilon}_k &= \frac{1}{k}(\epsilon_1 + \ldots + \epsilon_k), & \overline{E}_k &= \frac{1}{k^2}(1\epsilon_1 + \ldots + k\epsilon_k), \\ \overline{\delta}_k &= \frac{1}{n-k}(\epsilon_{k+1} + \ldots + \epsilon_n), & \overline{\triangle}_k &= \frac{1}{(n-k)^2}((k+1)\epsilon_{k+1} + \ldots + n\epsilon_n). \end{aligned} \tag{7}$$

First, using formulas to find estimates for the parameters of the simple linear regression model (see [15]), by some algebraic calculations, we get

Lemma 4.1. *For $k_0 \le k \le n - k_0$, the estimates $\hat{\alpha}_{0k}, \hat{\alpha}_{1k}$ and $\hat{\beta}_{0k}, \hat{\beta}_{1k}$ of the parameters of the simple linear regression model (1) through the set of the first k observations and the last $(n-k)$ ones, respectively, are defined by*

$$\begin{aligned} \hat{\alpha}_{0k} &= \alpha_0 + A_{0k} + \hat{K}_{0k}, & \hat{\alpha}_{1k} &= \alpha_1 + A_{1k} + \hat{K}_{1k}, \\ \hat{\beta}_{0k} &= \alpha_0 + B_{0k} + \hat{L}_{0k}, & \hat{\beta}_{1k} &= \alpha_1 + B_{1k} + \hat{L}_{1k}, \end{aligned} \tag{8}$$

where

$$\begin{aligned} A_{0k} &= -\frac{hk^*(k-k^*)(k-k^*+1)}{nk(k-1)} I(k^* < k \le n - k_0), \\ A_{1k} &= \frac{h(k-k^*)(k-k^*+1)}{k(k^2-1)}(2k^* + k - 1) I(k^* < k \le n - k_0); \end{aligned} \tag{9}$$

$$
\begin{aligned}
B_{0k} &= h\frac{(n-k^*)(n-k^*+1)}{n(n-k)\big((n-k)^2-1\big)}\big(2k(k+1)-(n+k+1)k^*\big)I(k<k^*)\\
&\quad - \frac{hk^*}{n}I(k\ge k^*),\\
B_{1k} &= h\frac{(n-k^*)(n-k^*+1)}{(n-k)\big((n-k)^2-1\big)}(n+2k^*-3k-1)I(k<k^*)+hI(k\ge k^*);
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
\hat{K}_{0k} &= \frac{4k+2}{k-1}\bar{\epsilon}_k - \frac{6k}{k-1}\overline{E}_k, \quad \hat{K}_{1k} = -\frac{6n}{k-1}\bar{\epsilon}_k + \frac{12nk}{k^2-1}\overline{E}_k,\\
\hat{L}_{0k} &= \Big(1+3\frac{(n+k+1)^2}{(n-k)^2-1}\Big)\bar{\epsilon}_k - 6\frac{(n-k)(n+k+1)}{(n-k)^2-1}\overline{E}_k,\\
\hat{L}_{1k} &= -6\frac{n(n+k+1)}{(n-k)^2-1}\bar{\delta}_k + 12\frac{n(n-k)}{(n-k)^2-1}\overline{\triangle}_k.
\end{aligned}
\tag{11}
$$

Lemma 4.2. *The components of parameter estimates in Lemma 4.1 have following properties:*

i) $0 > A_{0k}\downarrow$, $0 < A_{1k}\uparrow$ *for* $k^* \le k \le n-k_0$,
$0 > B_{0k}\downarrow$, $0 < B_{1k}\uparrow$ *for* $k_0 \le k \le k^*$*;*

ii)

$$
\begin{aligned}
\sum_{i=1}^{k}(B_{0k^*}+B_{1k^*}x_i) &= \frac{hk}{2n}(-2k^*+k+1),\\
\sum_{i=1}^{k^*}(B_{0k^*}+B_{1k^*}x_i) &= \frac{hk^*}{2n}(1-k^*),\\
\sum_{i=k^*+1}^{n}(B_{0k^*}+B_{1k^*}x_i) &= \frac{h(n-k^*)(n-k^*+1)}{2n}.
\end{aligned}
\tag{12}
$$

Proof. For $k_0 \le k < k^*-1$ then

$$
\begin{aligned}
B_{0k}-B_{0k+1} &= \frac{2h(2n+k+2)(n-k^*)(n-k^*+1)}{n(n-k-2)(n-k-1)(n-k)(n-k+1)}(k^*-k-1) > 0,\\
B_{1k+1}-B_{1k} &= 6h\frac{(n-k^*)(n-k^*+1)}{(n-k-2)(n-k-1)(n-k)(n-k+1)}(k^*-k-1) > 0,
\end{aligned}
$$

and the monotony of B_{0k}, B_{1k} follows. The monotony of $A_{ik}, i = 0,1$ is obtained by a similar argument. By some calculations, we have (ii). □

Lemma 4.3. *For* $k_0 \le k < k_1$*, a given small enough* ϵ*, all large enough* n*'s,* B_{ij} *defined by (10) satisfy the following inequalities:*

$$
0 > B_{0k}+B_{1k}x_k \ge B_{0k}+B_{1k}x_i > B_{0k_1}+B_{1k_1}x_i \quad \textit{for} \quad 1 \le i \le k, \tag{13}
$$

$$
-(B_{0k_1}+B_{1k_1}x_i) \ge -\frac{7}{8}(B_{0k^*}+B_{1k^*}x_i) > 0 \quad \textit{for} \quad 1 \le i \le k_1, \tag{14}
$$

$$
\sum_{i=1}^{k^*}(B_{0k^*}+B_{1k^*}x_i)^2 - \sum_{i=1}^{k}(B_{0k}+B_{1k}x_i)^2 \ge \frac{3}{2}nh^2\epsilon^2\frac{\tau^2}{1-\tau}. \tag{15}
$$

Proof. Applying (i) in Lemma 4.2, (13) is easy to be deduced.
Clearly, $B_{0k_1}B_{1k^*} \geq B_{0k^*}B_{1k_1}$ then $\dfrac{-(B_{0k_1}+B_{1k_1}x_i)}{-(B_{0k^*}+B_{1k^*}x_i)} \geq \dfrac{-(B_{0k_1}+B_{1k_1}x_{k_1})}{-(B_{0k^*}+B_{1k^*}x_{k_1})}$.
Using (10), the right-hand side of this inequality becomes

$$\frac{n}{h(k^*-k_1)} \times \frac{h}{n}\frac{(n-k^*)(n-k^*+1)}{(n-k_1)(n-k_1+1)}(k^*-k_1) = \frac{(n-k^*)(n-k^*+1)}{(n-k_1)(n-k_1+1)} > \frac{7}{8}$$

for a given small enough ϵ and all large enough n's. Hence (14) follows.
From (13),

$$\begin{aligned}
&\sum_{i=1}^{k^*}(B_{0k^*}+B_{1k^*}x_i)^2 - \sum_{i=1}^{k}(B_{0k}+B_{1k}x_i)^2 \geq \sum_{i=1}^{k^*}(B_{0k^*}+B_{1k^*}x_i)^2 \\
&-\sum_{i=1}^{k}(B_{0k_1}+B_{1k_1}x_i)^2 > \sum_{i=1}^{k_1}\Big((B_{0k^*}+B_{1k^*}x_i)^2-(B_{0k_1}+B_{1k_1}x_i)^2\Big) \\
&\geq \Big((B_{0k_1}+B_{1k_1}x_{k_1})-(B_{0k^*}+B_{1k^*}x_{k_1})\Big)\times \\
&\qquad \times \sum_{i=1}^{k_1}\Big(-(B_{0k_1}+B_{1k_1}x_i)-(B_{0k^*}+B_{1k^*}x_i)\Big). \qquad (16)
\end{aligned}$$

To evaluate the right-hand side of (16), we first have

$$\begin{aligned}
B_{0k_1}-B_{0k^*}+(B_{1k_1}-B_{1k^*})x_{k_1} &= \frac{h}{n}(k^*-k_1)\Big[1-\frac{(n-k^*)(n-k^*+1)}{(n-k_1)(n-k_1+1)}\Big] \\
&\geq \frac{h}{n}(k^*-k_1)\Big[1-\frac{(n-k^*+1)^2}{(n-k_1+1)^2}\Big] \geq \frac{7h(k^*-k_1)^2}{4n(n-k_1+1)}. \qquad (17)
\end{aligned}$$

On the other hand, from (12), (14) yields

$$\begin{aligned}
&\sum_{i=1}^{k_1}\big(-(B_{0k_1}+B_{1k_1}x_i)-(B_{0k^*}+B_{1k^*}x_i)\big) \geq \frac{15}{8}\sum_{i=1}^{k_1}\big(-(B_{0k^*}+B_{1k^*}x_i)\big) \\
&\quad = \frac{15}{8}\frac{hk_1}{2n}(k^*+k^*-k_1-1) \geq \frac{15hk_1k^*}{16n}. \qquad (18)
\end{aligned}$$

From (16)-(18), we get

$$\begin{aligned}
&\sum_{i=1}^{k^*}(B_{0k^*}+B_{1k^*}x_i)^2 - \sum_{i=1}^{k}(B_{0k}+B_{1k}x_i)^2 \\
&\qquad > \frac{7h(k^*-k_1)^2}{4n(n-k_1+1)}\frac{15hk_1k^*}{16n} \geq \frac{3}{2}nh^2\epsilon^2\frac{\tau^2}{1-\tau}
\end{aligned}$$

for the given small enough ϵ, large enough n's, so (15) follows. □

Now, let $\{\phi_i\}$ and $\{\gamma_j\}$ be centered, independent, ϕ_i and γ_j be independent, $E(\phi_i^2)=\sigma_1^2$, $E(\gamma_j^2)=\sigma_2^2$ such as $\phi_i=\epsilon_i$, $1\leq i\leq k^*$, $\gamma_j=\epsilon_{n-j+1}$, $1\leq j\leq n-k^*$. Put

$$\begin{aligned}
\overline{\phi}_k &= \frac{1}{k}(\phi_1+\ldots+\phi_k), \qquad \overline{\Phi}_k = \frac{1}{k^2}(1\phi_1+\ldots+k\phi_k), \\
\overline{\gamma}_k &= \frac{1}{k}(\gamma_1+\ldots+\gamma_k), \qquad \overline{\Gamma}_k = \frac{1}{k^2}(1\gamma_1+\ldots+k\gamma_k).
\end{aligned} \qquad (19)$$

Lemma 4.4. *If A5 holds and $k_0 \to \infty$ iff $n \to \infty$ then*

$$T_n = \max_{k_0 \le k \le n-k_0} \max\left(|\overline{\phi}_k|, |\overline{\Phi}_k|, |\overline{\gamma}_k|, |\overline{\Gamma}_k|\right) \xrightarrow{a.s.} 0 \ \text{ as } n \to \infty.$$

Proof. By the strong law of large numbers, $\overline{\phi}_k$, $\overline{\Phi}_k$, $\overline{\gamma}_k$, $\overline{\Gamma}_k \xrightarrow{\text{a.s.}} 0$ as $k \to \infty$, hence $\max_{k_0 \le k} \max\left(|\overline{\phi}_k|, |\overline{\Phi}_k|, |\overline{\gamma}_k|, |\overline{\Gamma}_k|\right) \xrightarrow{\text{a.s.}} 0$ as $k_0 \to \infty$.
The assertion is clear from the fact that

$$0 \le T_n \le \max_{k_0 \le k} \max\left(|\overline{\phi}_k|, |\overline{\Phi}_k|, |\overline{\gamma}_k|, |\overline{\Gamma}_k|\right).$$

□

Lemma 4.5. *Let*

$$\overline{\epsilon}_{(n)} = \max_{k_0 \le k \le n-k_0} |\overline{\epsilon}_k|, \qquad \overline{E}_{(n)} = \max_{k_0 \le k \le n-k_0} |\overline{E}_k|,$$
$$\overline{\delta}_{(n)} = \max_{k_0 \le k \le n-k_0} |\overline{\delta}_k|, \qquad \overline{\triangle}_{(n)} = \max_{k_0 \le k \le n-k_0} |\overline{\triangle}_k|,$$

where $\overline{\epsilon}_k$, $\overline{E}_k$, $\overline{\delta}_k$, $\overline{\triangle}_k$ are defined by (7). If A5 holds and $k_0 \to \infty$ iff $n \to \infty$ then

$$G_n = \max\left(\overline{\epsilon}_{(n)}, \overline{E}_{(n)}, \overline{\delta}_{(n)}, \overline{\triangle}_{(n)}\right) \xrightarrow{a.s.} 0.$$

Proof. First, note that $\overline{\epsilon}_k = \overline{\phi}_k$ for $k_0 \le k \le k^*$, so $|\overline{\epsilon}_k| \le T_n$. For $k^* < k \le n - k_0$,

$$\overline{\epsilon}_k = \frac{k^*}{k}\overline{\phi}_{k^*} + \frac{n-k^*}{k}\overline{\gamma}_{n-k^*} - \frac{n-k}{k}\overline{\gamma}_{n-k}.$$

Hence $|\overline{\epsilon}_k| \le |\overline{\phi}_{k^*}| + \frac{n-k^*}{k^*}\left(|\overline{\gamma}_{n-k^*}| + |\overline{\gamma}_{n-k}|\right) \le \left(1 + 2\frac{n-k^*}{k^*}\right)T_n$.
It follows that the $\overline{\epsilon}_{(n)}$ converges almost surely.

Similarly, to prove the convergence of $\overline{E}_{(n)}$, we first have $\overline{E}_k = \overline{\Phi}_k$ for $k_0 \le k \le k^*$ then $|\overline{E}_k| \le T_n$. For $k^* < k \le n - k_0$ then $n - k \ge k_0$ and then

$$\begin{aligned}
\overline{E}_k &= \frac{1}{k^2}\left(1\epsilon_1 + \ldots + k^*\epsilon_{k^*} + (k^*+1)\epsilon_{k^*+1} + \ldots + k\epsilon_k\right) \\
&= \frac{k^{*2}}{k^2}\overline{E}_{k^*} + \frac{(n+1)(n-k^*)}{k^2}\overline{\gamma}_{n-k^*} - \frac{(n+1)(n-k)}{k^2}\overline{\gamma}_{n-k} \\
&\quad - \frac{(n-k^*)^2}{k^2}\overline{\Gamma}_{n-k^*} + \frac{(n-k)^2}{k^2}\overline{\Gamma}_{n-k}.
\end{aligned}$$

So $|\overline{E}_k| \le |\overline{E}_{k^*}| + \frac{n(n-k^*)}{k^{*2}}\left(|\overline{\gamma}_{n-k^*}| + |\overline{\gamma}_{n-k}|\right)$
$\qquad + \frac{(n-k^*)^2}{k^{*2}}\left(|\overline{\Gamma}_{n-k^*}| + |\overline{\Gamma}_{n-k}|\right)$.
Hence $|\overline{E}_k| \le \Big(1 + 2\frac{n(n-k^*)}{k^{*2}} + 2\frac{(n-k^*)^2}{k^{*2}}\Big)T_n$.
From that we obtain the convergence of $\overline{E}_{(n)}$.

To consider the convergence of $\overline{\delta}_{(n)}$, note that $\overline{\delta}_k = \overline{\gamma}_{n-k}$ for $k^* < k \le n - k_0$ then $|\overline{\delta}_k| \le T_n$. For $k_0 \le k \le k^*$,

$$\overline{\delta}_k = \frac{1}{n-k}(\epsilon_{k+1} + \ldots + \epsilon_n) = \frac{1}{n-k}(\epsilon_{k+1} + \ldots + \epsilon_{k^*} + \gamma_{n-k^*} + \ldots + \gamma_1)$$

$$= \frac{k^*}{n-k}\bar{\epsilon}_{k^*} - \frac{k}{n-k}\bar{\epsilon}_k + \frac{n-k^*}{n-k}\overline{\gamma}_{n-k^*}.$$

Therefore $|\overline{\delta}_k| \leq \frac{k^*}{n-k^*}\big(|\bar{\epsilon}_{k^*}| + |\bar{\epsilon}_k|\big) + |\overline{\gamma}_{n-k^*}| \leq \Big(2\frac{k^*}{n-k^*}+1\Big)T_n$. We get the convergence of $\overline{\delta}_{(n)}$.

For $\overline{\triangle}_{(k)}$, we find that for $k^* < k \leq n-k_0$ then

$$\overline{\triangle}_k = \frac{1}{(n-k)^2}\big((k+1)\epsilon_{(k+1)} + \ldots + n\epsilon_n)\big) = \frac{n+1}{n-k}\overline{\gamma}_{n-k} - \overline{\Gamma}_{n-k}.$$

So $|\overline{\triangle}_k| \leq \frac{n+1}{n-k_0}|\overline{\gamma}_{n-k}| + |\overline{\Gamma}_{n-k}| \leq \Big(\frac{n+1}{n-k_0}+1\Big)T_n$.

For $k_0 \leq k \leq k^*$ then

$$\begin{aligned}
\overline{\triangle}_k &= \frac{1}{(n-k)^2}\big((k+1)\epsilon_{k+1} + \ldots + n\epsilon_n\big) \\
&= \frac{1}{(n-k)^2}\big((k+1)\phi_{k+1} + \ldots + k^*\phi_{k^*} + (k^*+1)\epsilon_{k^*+1} + \ldots + n\epsilon_n\big) \\
&= \frac{k^{*2}}{(n-k)^2}\overline{\Phi}_{k^*} - \frac{k^2}{(n-k)^2}\overline{\Phi}_k + \frac{(n-k^*)(n+1)}{(n-k)^2}\overline{\gamma}_{n-k^*} - \frac{(n-k^*)^2}{(n-k)^2}\overline{\Gamma}_{n-k^*}.
\end{aligned}$$

From that

$$|\overline{\triangle}_k| \leq \frac{k^{*2}}{(n-k^*)^2}|\overline{\Phi}_{k^*}| + \frac{k^{*2}}{(n-k^*)^2}|\overline{\Phi}_k| + \frac{n(n-k^*)}{(n-k^*)^2}|\overline{\gamma}_{n-k^*}| + |\overline{\Gamma}_{n-k^*}|.$$

Hence $|\overline{\triangle}_k| \leq \Big(2\frac{k^{*2}}{(n-k^*)^2} + \frac{n}{n-k^*} + 1\Big)T_n$. The convergence of $\overline{\triangle}_{(n)}$ is obtained. Lemma 4.5 is deduced. □

Lemma 4.6. *If A5 holds and $k_0 \to \infty$ iff $n \to \infty$ then*

$$\begin{aligned}
U_n &= \max_{k_0 \leq k \leq n-k_0}\Big(|\hat{K}_{0k}| + |\hat{K}_{1k}|\Big) \leq \Big(11 + \frac{19}{\tau_0}\Big)G_n, \\
V_n &= q\max_{k_0 \leq k \leq n-k_0}\Big(|\hat{L}_{0k}| + |\hat{L}_{1k}|\Big) \leq \Big(1 + \frac{24}{\tau_0} + \frac{18}{\tau_0^2}\Big)G_n, \\
F_n &= \max(U_n, V_n) \xrightarrow{a.s.} 0.
\end{aligned}$$

Proof. Using (11) we get

$$\begin{aligned}
U_n &\leq \max_{k_0 \leq k \leq n-k_0}\Big(\frac{4k+2}{k-1}|\bar{\epsilon}_k| + \frac{6k}{k-1}|\overline{E}_k| + \frac{6n}{k-1}|\bar{\epsilon}_k| + \frac{12nk}{k^2-1}|\overline{E}_k|\Big) \\
&\leq \Big(\frac{10k+2}{k-1} + \frac{6n(k+1+2k)}{(k-1)(k+1)}\Big)G_n < \Big(\frac{10k+2}{k-1} + \frac{18n(k+1)}{(k-1)(k+1)}\Big)G_n \\
&\leq \Big(11 + \frac{19}{\tau_0}\Big)G_n.
\end{aligned}$$

$$V_n \leq \Big(1 + 3\frac{(n+k+1)^2}{(n-k)^2-1} + 6\frac{(n-k)(n+k+1)}{(n-k)^2-1} + 6\frac{n(n+k+1)}{(n-k)^2-1}$$

$$
+ 12\frac{n(n-k)}{(n-k)^2-1}\Big)G_n
$$

$$
\le \Big(1 + 3\frac{(2n)^2}{k_0^2} + 6\frac{2n}{k_0} + 6\frac{2n^2}{k_0^2} + 12\frac{n}{k_0}\Big)G_n \le \Big(1 + \frac{24}{\tau_0} + \frac{24}{\tau_0^2}\Big)G_n.
$$

The rest of the proof is deduced from the fact $G_n \xrightarrow{\text{a.s.}} 0$ proved in Lemma 4.5. Lemma 4.6 is proved. □

Lemma 4.7. *With the assumptions of Theorem 2.1, the inclusion*

$$
\Big\{\sup_{n>N}\Big(F_n^2 + 2F_nh\big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\big) + 4G_n(F_n + 2h\tau^2)\Big) \le \frac{3}{2}h^2\epsilon^2\frac{\tau^2}{1-\tau}\Big\}
$$

$$
\subset \Big\{\inf_{n>N}\frac{\hat{k}_n}{n} > \tau - \epsilon\Big\},
$$

where G_n is defined by Lemma 4.5, F_n is defined by Lemma 4.6, follows.

Proof. For $k_0 \le k \le k_1 = \lfloor(\tau-\epsilon)n\rfloor$,

$$
\sum_{i=1}^{n}\widetilde{e}_{ik^*}^2 = \sum_{i=1}^{k}\widetilde{e}_{ik^*}^2 + \sum_{i=k+1}^{k^*}\widetilde{e}_{ik^*}^2 + \sum_{i=k^*+1}^{n}\widetilde{e}_{ik^*}^2. \tag{20}
$$

Let's denote the terms in the right-hand side of (20) by I_1, I_2, I_3, respectively. The equality $(a-b)^2 = (a-c)^2 + b^2 - c^2 + 2a(c-b)$ is used to transform those terms. Put

$$
\breve{\alpha}_{0k} = \hat{\alpha}_{0k} - \alpha_0,\ \breve{\alpha}_{1k} = \hat{\alpha}_{1k} - \alpha_1,\ \breve{\beta}_{0k} = \hat{\beta}_{0k} - \alpha_0,\ \breve{\beta}_{1k} = \hat{\beta}_{1k} - \alpha_1.
$$

First, note that

$$
I_1 = \sum_{i=1}^{k}\widetilde{e}_{ik^*}^2 = \sum_{i=1}^{k}\Big(y_i - (\hat{\beta}_{0k^*} + \hat{\beta}_{1k^*}x_i)\Big)^2 = \sum_{i=1}^{k}\Big(\epsilon_i - (\breve{\beta}_{0k^*} + \breve{\beta}_{1k^*}x_i)\Big)^2
$$

$$
= \sum_{i=1}^{k}\Big(\epsilon_i - (\breve{\beta}_{0k} + \breve{\beta}_{1k}x_i)\Big)^2 + \sum_{i=1}^{k}\Big((\breve{\beta}_{0k^*} + \breve{\beta}_{1k^*}x_i)^2 - (\breve{\beta}_{0k} + \breve{\beta}_{1k}x_i)^2\Big)
$$

$$
+ \sum_{i=1}^{k}2\epsilon_i\Big((\breve{\beta}_{0k} + \breve{\beta}_{1k}x_i) - (\breve{\beta}_{0k^*} + \breve{\beta}_{1k^*}x_i)\Big) = \sum_{i=1}^{k}\widetilde{e}_{ik}^2 + R_1,
$$

where

$$
R_1 = \sum_{i=1}^{k}(B_{0k^*} + B_{1k^*}x_i)^2 - \sum_{i=1}^{k}(B_{0k} + B_{1k}x_i)^2
$$

$$
+ \sum_{i=1}^{k}\Big((\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i)^2 - (\hat{L}_{0k} + \hat{L}_{1k}x_i)^2\Big)
$$

$$
+ 2\sum_{i=1}^{k}(B_{0k^*} + B_{1k^*}x_i)(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i) - 2\sum_{i=1}^{k}(B_{0k} + B_{1k}x_i)(\hat{L}_{0k} + \hat{L}_{1k}x_i)
$$

$$
+ 2\sum_{i=1}^{k}\epsilon_i\Big((\hat{L}_{0k} + \hat{L}_{1k}x_i) - (\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i)\Big) + 2\sum_{i=1}^{k}\epsilon_i(B_{0k} + B_{1k}x_i)
$$

$$- 2\sum_{i=1}^{k}\epsilon_i(B_{0k^*} + B_{1k^*}x_i).$$

Using similar reductions, we get

$$I_2 = \sum_{i=k+1}^{k^*}\widetilde{e}_{ik^*}^2 = \sum_{i=k+1}^{k^*}\widetilde{e}_{ik}^2 + R_2,$$
$$I_3 = \sum_{i=k^*+1}^{n}\widetilde{e}_{ik^*}^2 = \sum_{i=k^*+1}^{n}\widetilde{e}_{ik}^2 + R_3,$$

where

$$\begin{aligned}
R_2 = &\sum_{i=k+1}^{k^*}(B_{0k^*} + B_{1k^*}x_i)^2 + \sum_{i=k+1}^{k^*}(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i)^2 \\
&+ 2\sum_{k+1}^{k^*}(B_{0k^*} + B_{1k^*}x_i)(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i) - \sum_{i=k+1}^{k^*}(\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i)^2 \\
&+ 2\sum_{i=k+1}^{k^*}\epsilon_i(\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i) - 2\sum_{i=k+1}^{k^*}\epsilon_i(B_{0k^*} + B_{1k^*}x_i) - 2\sum_{i=k+1}^{k^*}\epsilon_i(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i), \\
R_3 = &\sum_{i=k^*+1}^{n}(\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*}x_i)^2 - \sum_{i=k^*+1}^{n}(\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i)^2 \\
&+ 2\sum_{i=k^*+1}^{n}(B_{0k^*} + B_{1k^*}x_i)\Big((\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i) - (\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*}x_i)\Big) \\
&+ 2\sum_{i=k^*+1}^{n}\epsilon_i\Big((\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i) - (\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*}x_i)\Big).
\end{aligned}$$

Let $R = \sum_{i=1}^{n}\widetilde{e}_{ik^*}^2 - \sum_{i=1}^{n}\widetilde{e}_{ik}^2$ then $R = R_1 + R_2 + R_3$. After reduction, we get

$$R = J_1 + J_2 + J_3 \tag{21}$$

where

$$\begin{aligned}
J_1 = &\sum_{i=1}^{k^*}(B_{0k^*} + B_{1k^*}x_i)^2 - \sum_{i=1}^{k}(B_{0k} + B_{1k}x_i)^2, \\
J_2 = &\Big(\sum_{i=1}^{k^*}(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i)^2 + \sum_{i=k^*+1}^{n}(\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*}x_i)^2\Big) \\
&- \Big(\sum_{i=1}^{k}(\hat{L}_{0k} + \hat{L}_{1k}x_i)^2 + \sum_{i=k+1}^{n}(\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i)^2\Big) \\
&- 2\sum_{i=1}^{k}(\hat{L}_{0k} + \hat{L}_{1k}x_i)(B_{0k} + B_{1k}x_i) - 2\sum_{i=k^*+1}^{n}(\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*}x_i)(B_{0k^*} + B_{1k^*}x_i) \\
&+ 2\sum_{i=1}^{k^*}(\hat{L}_{0k^*} + \hat{L}_{1k^*}x_i)(B_{0k^*} + B_{1k^*}x_i) + 2\sum_{i=k^*+1}^{n}(\breve{\alpha}_{0k} + \breve{\alpha}_{1k}x_i)(B_{0k^*} + B_{1k^*}x_i),
\end{aligned}$$

$$J_3 = \sum_{i=1}^{k} \epsilon_i (B_{0k} + B_{1k} x_i) - \sum_{i=1}^{k^*} \epsilon_i (B_{0k^*} + B_{1k^*} x_i) + \sum_{i=1}^{k} \epsilon_i (\hat{L}_{0k} + \hat{L}_{1k} x_i)$$
$$+ \sum_{i=k+1}^{n} \epsilon_i (\breve{\alpha}_{0k} + \breve{\alpha}_{1k} x_i) - \sum_{i=1}^{k^*} \epsilon_i (\hat{L}_{0k^*} + \hat{L}_{1k^*} x_i) - \sum_{i=k^*+1}^{n} \epsilon_i (\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*} x_i).$$

According to (15)

$$J_1 = \sum_{i=1}^{k^*} (B_{0k^*} + B_{1k^*} x_i)^2 - \sum_{i=1}^{k} (B_{0k} + B_{1k} x_i)^2 \geq \frac{3}{2} n h^2 \epsilon^2 \frac{\tau^2}{1-\tau}.$$

Because $|\hat{L}_{0k} + \hat{L}_{1k} x_i| \leq |\hat{L}_{0k}| + |\hat{L}_{1k}| \leq F_n$, $|\breve{\alpha}_{0k} + \breve{\alpha}_{1k} x_i| \leq |\hat{K}_{0k}| + |\hat{K}_{1k}| \leq F_n$ and using Lemma 4.2, 4.5 and 4.6 we obtain

$$J_2 \geq -|J_2| \geq 0 - \Big(\sum_{i=1}^{k} (\hat{L}_{0k} + \hat{L}_{1k} x_i)^2 + \sum_{i=k+1}^{n} (\breve{\alpha}_{0k} + \breve{\alpha}_{1k} x_i)^2 \Big)$$
$$+ 2 \sum_{i=1}^{k} |(\hat{L}_{0k} + \hat{L}_{1k} x_i)| (B_{0k} + B_{1k} x_i) - 2 \sum_{i=k^*+1}^{k} |(\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*} x_i)| (B_{0k^*} + B_{1k^*} x_i)$$
$$+ 2 \sum_{i=1}^{k^*} |(\hat{L}_{0k^*} + \hat{L}_{1k^*} x_i)| (B_{0k^*} + B_{1k^*} x_i) - 2 \sum_{i=k^*+1}^{k} |(\breve{\alpha}_{0k} + \breve{\alpha}_{1k} x_i)| (B_{0k^*} + B_{1k^*} x_i)$$
$$\geq - \Big(\sum_{i=1}^{k} F_n^2 + \sum_{i=k+1}^{n} F_n^2 \Big) + 2 \sum_{i=1}^{k} F_n (B_{0k} + B_{1k} x_i) - 2 \sum_{i=k^*+1}^{n} F_n (B_{0k^*} + B_{1k^*} x_i)$$
$$+ 2 \sum_{i=1}^{k^*} F_n (B_{0k^*} + B_{1k^*} x_i) - 2 \sum_{i=k^*+1}^{n} F_n (B_{0k^*} + B_{1k^*} x_i)$$
$$\geq -n F_n^2 + 2 \sum_{i=1}^{k} F_n (B_{0k} + B_{1k} x_i) - 2 \sum_{i=k^*+1}^{n} F_n (B_{0k^*} + B_{1k^*} x_i)$$
$$+ 2 \sum_{i=1}^{k^*} F_n (B_{0k^*} + B_{1k^*} x_i) - 2 \sum_{i=k^*+1}^{n} F_n (B_{0k^*} + B_{1k^*} x_i)$$
$$\geq -n F_n^2 + 4 F_n \Big(\sum_{i=1}^{k^*} (B_{0k^*} + B_{1k^*} x_i) - \sum_{i=k^*+1}^{n} (B_{0k^*} + B_{1k^*} x_i) \Big)$$
$$= -n F_n^2 - 2 n h F_n \Big(\frac{k^*(k^*-1)}{n^2} + \frac{(n-k^*)(n-k^*+1)}{n^2} \Big)$$
$$\geq -n F_n^2 - 2 n h F_n \Big(\tau^2 + (1-\tau)^2 + \frac{3}{n} \Big),$$

$$J_3 = \sum_{i=1}^{k} \epsilon_i \big((\hat{L}_{0k} + B_{0k}) + (\hat{L}_{1k} + B_{1k}) x_i \big) - \sum_{i=1}^{k^*} \epsilon_i \big((\hat{L}_{0k^*} + B_{0k^*}) + (\hat{L}_{1k^*} + B_{1k^*} x_i) \big)$$
$$+ \sum_{i=k+1}^{n} \epsilon_i (\breve{\alpha}_{0k} + \breve{\alpha}_{1k} x_i) - \sum_{i=k^*+1}^{n} \epsilon_i (\breve{\alpha}_{0k^*} + \breve{\alpha}_{1k^*} x_i)$$

$$
\begin{aligned}
&= (\hat{L}_{0k} + B_{0k})k\overline{\epsilon}_k + (\hat{L}_{1k} + B_{1k})\frac{k^2}{n}\overline{E}_k - (\hat{L}_{0k^*} + B_{0k^*})k^*\overline{\epsilon}_{k^*} - (\hat{L}_{1k^*} + B_{1k^*})\frac{k^{*2}}{n}\overline{E}_{k^*} \\
&\quad + \breve{\alpha}_{0k}(n-k)\overline{\delta}_k + \breve{\alpha}_{1k}\frac{(n-k)^2}{n}\overline{\triangle}_k - \breve{\alpha}_{0k^*}(n-k^*)\overline{\delta}_{k^*} + \breve{\alpha}_{1k^*}\frac{(n-k^*)^2}{n}\overline{\triangle}_{k^*} \\
&\geq -(|\hat{L}_{0k}| - B_{0k^*})k|\overline{\epsilon}_k| - (|\hat{L}_{1k}| + B_{1k})\frac{k^2}{n}|\overline{E}_k| - (|\hat{L}_{0k^*}| - B_{0k^*})k^*|\overline{\epsilon}_{k^*}| \\
&\quad - (|\hat{L}_{1k^*}| + B_{1k^*})\frac{k^{*2}}{n}|\overline{E}_{k^*}| - |\hat{K}_{0k}|(n-k)|\overline{\delta}_k| \\
&\quad - |\hat{K}_{1k}|\frac{(n-k)^2}{n}|\overline{\triangle}_k| - |\hat{K}_{0k^*}|(n-k^*)|\overline{\delta}_{k^*}| - |\hat{K}_{1k^*}|\frac{(n-k^*)^2}{n}|\overline{\triangle}_{k^*}| \\
&= -G_n\Big((|\hat{L}_{0k}| + \frac{hk^*}{n})k + (|\hat{L}_{1k}| + h)\frac{k^2}{n} + (|\hat{L}_{0k^*}| + \frac{hk^*}{n})k^* + (|\hat{L}_{1k^*}| + h)\frac{k^{*2}}{n} \\
&\quad + |\hat{K}_{0k}|(n-k) + |\hat{K}_{1k}|\frac{(n-k)^2}{n} + |\hat{K}_{0k^*}|(n-k^*) + |\hat{K}_{1k^*}|\frac{(n-k^*)^2}{n}\Big) \\
&\geq -G_n\Big(k(|\hat{L}_{0k}| + |\hat{L}_{1k}|) + \frac{hkk^*}{n} + h\frac{k^2}{n} + k^*(|\hat{L}_{0k^*}| + |\hat{L}_{1k^*}|) + 2h\frac{k^{*2}}{n} \\
&\quad + (|\hat{K}_{0k}| + |\hat{K}_{1k}|)(n-k) + (|\hat{K}_{0k^*}| + |\hat{K}_{1k^*}|)(n-k^*)\Big) \\
&\geq -G_n\Big(kF_n + \frac{hkk^*}{n} + h\frac{k^2}{n} + k^*F_n + 2h\frac{k^{*2}}{n} + F_n(n-k) + F_n(n-k^*)\Big) \\
&\geq -2nG_n(F_n + 2h\tau^2).
\end{aligned}
$$

Substituting the above assessments into (21) yields

$$
\begin{aligned}
R &\geq \frac{3}{2}nh^2\epsilon^2\frac{\tau^2}{1-\tau} + n\Big(-F_n^2 - 2F_nh\big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\big)\Big) - 4nG_n(F_n + 2h\tau^2) \\
&= n\Big(\frac{3}{2}h^2\epsilon^2\frac{\tau^2}{1-\tau} - F_n^2 - 2F_nh\big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\big) - 4G_n(F_n + 2h\tau^2)\Big).
\end{aligned}
$$

The right-hand side of the last inequality is nonnegative if and only if

$$
F_n^2 + 2F_nh\Big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\Big) + 4G_n(F_n + 2h\tau^2) \leq \frac{3}{2}h^2\epsilon^2\frac{\tau^2}{1-\tau}. \qquad (22)
$$

From (6), yields

$$
\begin{aligned}
&\Big\{F_n^2 + 2F_nh\Big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\Big) + 4G_n(F_n + 2h\tau^2) \leq \frac{3}{2}h^2\epsilon^2\frac{\tau^2}{1-\tau}\Big\} \\
&\subset \Big\{\max_{k_0 \leq k \leq k_1}\Big(\sum_{i=1}^{n}\widetilde{e}_{ik^*}^2 - \sum_{i=1}^{n}\widetilde{e}_{ik}^2\Big) \geq 0\Big\} \subset \Big\{\frac{\hat{k}_n}{n} \geq \tau - \epsilon\Big\}.
\end{aligned}
$$

Lemma 4.7 is proved. □

Proof of Theorem 2.1 *(continuous).* From Lemma 4.7, we deduce

$$
\begin{aligned}
&P\Big\{\inf_{n>N}\frac{\hat{k}_n}{n} \geq \tau - \epsilon\Big\} \\
&\geq P\Big\{\sup_{n>N}\Big(F_n^2 + 2F_nh\big(\tau^2 + (1-\tau)^2 + \frac{3}{n}\big) + 4G_n(F_n + 2h\tau^2)\Big) \qquad (23) \\
&\qquad \leq \frac{3}{2}h^2\epsilon^2\frac{\tau^2}{1-\tau}\Big\} \longrightarrow 1 \text{ as } N \to \infty.
\end{aligned}
$$

The last convergence follows from the fact $F_n, G_n \xrightarrow{\text{a.s.}} 0$ by Lemma 4.5, 4.6.

By a similar argument, we thus get

$$\begin{aligned} &P\Big\{\sup_{n>N}\frac{\hat{k}_n}{n} < \tau+\epsilon\Big\} \\ &\geq P\Big\{\sup_{n>N}\Big(F_n^2 + 2F_n h\big(\tau^2+(1-\tau)^2+\frac{3}{n}\big) + 4G_n(F_n+2h\tau^2)\Big) \\ &\qquad\quad \leq \frac{3}{2}h^2\epsilon^2\frac{(1-\tau)^2}{\tau}\Big\} \longrightarrow 1 \text{ as } N\to\infty. \end{aligned} \tag{24}$$

So (5) holds and the Theorem 2.1 is completely proved. □

Proof of Theorem 2.2. By the strong law of large numbers,

$$k^p\overline{\phi}_k,\ k^p\overline{\Phi}_k,\ k^p\overline{\gamma}_k,\ k^p\overline{\Gamma}_k \xrightarrow{\text{a.s.}} 0 \text{ as } k\to\infty.$$

Note that $n \leq (k_0+1)/\tau_0$ we get $n^pT_n \xrightarrow{\text{a.s.}} 0$, hence $n^pG_n \xrightarrow{a.s.} 0$ so $n^pF_n \xrightarrow{\text{a.s.}} 0$.

The inequality (22) still holds if h is replaced by h_n. Multiplying both sides of (22) by n^{2p} we get

$$\begin{aligned} &(n^pF_n)^2 + 2(n^pF_n)(n^ph_n)\Big(\tau^2+(1-\tau)^2+\frac{3}{n}\Big) + 4(n^pG_n)(n^pF_n+2n^ph_n\tau^2) \\ &\qquad \leq \frac{3}{2}(n^ph_n)^2\epsilon^2\frac{\tau^2}{1-\tau} \\ &\text{or } (n^pF_n)^2 + 4(n^pG_n)(n^pF_n) \\ &\qquad \leq (n^ph_n)\Big(\frac{3}{2}(n^ph_n)\epsilon^2\frac{\tau^2}{1-\tau} - 2(n^pF_n)\big(\tau^2+(1-\tau)^2+\frac{3}{n}\big) - 4n^pG_n2\tau^2\Big). \end{aligned}$$

This inequality holds for all large enough n's. Hence, (23) follows, so does (24). From that $\hat{k}_n/n \xrightarrow{\text{a.s.}} \tau$.

Now, consider the case $\lim_{n\to\infty} h_n = h_\infty \neq 0$. Without loss of generality, we can assume that $h_\infty > 0$. By the above proved assertion, $\hat{k}_n/n \xrightarrow{\text{a.s.}} \tau$. Because $k_0 < \hat{k}_n < n-k_0$, according to Lemma 4.6, $\hat{K}_{i\hat{k}_n}, \hat{L}_{i\hat{k}_n} \xrightarrow{\text{a.s.}} 0\,(i=0,1)$. From that and (9), (10) we obtain

$$\lim_{n\to\infty} A_{i\hat{k}_n} = 0,\ \lim_{n\to\infty} B_{0\hat{k}_n} = -h_\infty\tau,\ \lim_{n\to\infty} B_{1\hat{k}_n} = h_\infty.$$

Now, applying (8) we thus proved the almost sure limits of the parameter estimates, which completes the proof of Theorem 2.2. □

Remark. For $p = 1/2$ the conclusion of Theorem 2.2 may not hold. Indeed, with the same assumptions as in the Theorem 2.2, let's assume further that ϵ_i are i.i.d. By the law of iterated logarithm, $\lim_{n\to\infty}\sup \frac{\phi_1+\ldots+\phi_n}{\sqrt{n\log\log n}} = \sqrt{2}$, thence $\lim_{n\to\infty}\sup \sqrt{n}\ \overline{\phi}_n = \infty$ and in general we do not obtain $\sqrt{n}G_n \xrightarrow{\text{a.s.}} 0$, $\sqrt{n}F_n \xrightarrow{\text{a.s.}} 0$ and neither do (23) and (24).

References

[1] G. Shen and J. K. Ghosh, "Developing a new BIC for detecting change-points," Journal of Statistical Planning and Inference, vol. 141, issue 3, 2011, pp. 1436–1447.

[2] I. Berkes, L. Horvath and J. Schauer, "Asymptotics of trimmed CUSUM statistics," Bernoulli, vol. 17, 2011, pp. 1344–1367.

[3] L. Horvath, Z. Horvath and M. Huskova, "Ratio tests for change point detection," IMS Colections, vol. 1, 2008, pp. 293–304.

[4] H. J. Kim and D. Siegmund, "The likelihood ratio test for a change point in simple linear regression," Biometrika, vol. 76, no. 3, 1989, pp. 409–423.

[5] T. V. Ban and N. T. Quyen, "The power of change-point test for two-phase regression", Applied Mathematics, vol. 5, 2014, pp. 2994–3000.

[6] R. Baragona, F. Battaglia and D. Cucina, "Empirical likelihood for break detection in time series," Electronic Journal of Statistics, vol. 7, 2013, pp. 3089–3123.

[7] Z. Liu and L. F. Qian, "Changepoint estimation in a segmented linear regression via empirical likelihood," Communications in Statistics - Simulation and Computation, vol. 39, no. 1, 2009, pp. 85–100.

[8] Y. Liu, C. Zou and R. Zhang, "Empirical likelihood ratio test for a change-point in linear regression model," Communications in Statistics - Theory and Methods, vol. 37, issue 16, 2008, pp. 2551–2563.

[9] J. Bai, "Estimation of a change point in multiple regression models," Review of Economic and Statistics, vol. 79, no. 4, 1997, pp. 551–563.

[10] H. L. Koul and L. F. Qian, "Asymptotics of maximum likelihood estimator in a two-phase linear regression model,' J. of Statistical Planning and Inference, vol. 108, 2002, pp. 99–119.

[11] H. L. Koul, L. F. Qian and D. Surgailis, "Asymptotics of M-estimators in two-phase linear regression models," Stochastic Proc. & their App., vol. 103, no. 1, 2003, pp. 123–154.

[12] C. A. R. Diniz and L. C. Brochi, "Robustness of two-phase regression tests," REVSTAT - Statistical Journal, vol. 3, no. 1, 2005, pp. 1-18.

[13] C. W. S. Chen, J. S. K. Chan, R. Gerlach and W. Y. L. Hsieh, "A comparison of estimators for regression models with change points," Stat Comput., vol. 21, issue 3, 2011, pp. 395–414.

[14] H. Zhao, X. Wu and H. Chen, "A new algorithm in detecting changepoint in linear regression models," 3rd Int. Conf. on Biomedical Engineering and Informatics (BMEI), vol. 6, Yantai, October 2010, pp. 2261-2264.

[15] J. Neter, W. Wasserman and M. H. Kutner, "Applied linear regression models," Irwin, 1989.