



**HAL**  
open science

# Convergence rate of a new non-parametric change-point estimator within a linear model

Jean-Marc Bardet, Najmieh Maksaei

## ► To cite this version:

Jean-Marc Bardet, Najmieh Maksaei. Convergence rate of a new non-parametric change-point estimator within a linear model. 2026. ⟨hal-05550979⟩

**HAL Id: hal-05550979**

**<https://hal.science/hal-05550979v1>**

Preprint submitted on 13 Mar 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Convergence rate of a new non-parametric change-point estimator within a linear model

Jean-Marc Bardet<sup>1</sup>, Najmieh Maksaei<sup>1</sup>

bardet@univ-paris1.fr, Najmieh.Maksaei@univ-paris1.fr

<sup>1</sup> SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris, FRANCE

March 13, 2026

## Abstract

We consider data initially derived from a homoscedastic linear model that undergoes a change at an unknown point in time, such as a change in model (not necessarily linear) or heteroscedasticity. Based on an idea of the sum of the squares of the residuals penalized introduced in Bardet (2025), we propose a new version of this criterion for estimating the change point. In the very general nonparametric framework we have adopted, we demonstrate that the convergence rate of this estimator is the same as in the parametric case to within a logarithmic term. To do this, we use concentration inequalities for sub-exponential variables.

*Keywords:* Change detection; Linear model; Non-parametric estimator

## 1 Introduction

We consider the following regression problem: we assume that  $((Y_1, Z_1), \dots, (Y_N, Z_N))$  is observed with  $Y_i$  real random variables and  $Z_i$  vectors of  $\mathbb{R}^p$ ,  $p \geq 1$ . Furthermore, we know that for  $1 \leq t \leq n < N$ , the relationship between  $Y_t$  and  $Z_t$  is linear, i.e., there exists  $\theta \in \mathbb{R}^p$  such that  $Y_t = Z_t' \theta + \varepsilon_t$ , where  $M'$  is the transpose of a matrix  $M$  and assuming further that the  $(\varepsilon_i)$  are independent identically distributed and centered random variables. But we do not know  $n$  and we do not know the relationship between  $Y_t$  and  $Z_t$  for  $n + 1 \leq t \leq N$ , except that it is no longer the previous linear model (more details are provided in Assumption A defined below). Our goal is to detect a break with the estimation of  $n$  and, incidentally, the estimation of  $\theta$ .

This framework of change-point in a linear model can take different forms in literature. The first and undoubtedly most studied is that of a change in structure (e.g., economic) which implies that for  $t > n$ , we still have a linear model, but this time satisfying  $Y_t = Z_t' \bar{\theta} + \varepsilon_t$  with  $\bar{\theta} \neq \theta$ . We can refer to the work of Chow (1960) and his famous test, and more recently the work of Bai and Perron (1998) or Lavielle (1999) based on least squares regression, or Bai (1998) based on least absolute values regression. We are then able to give very precise results on the law of the estimator of the breakpoint  $n$  and the convergence rate, which is in  $\mathcal{O}_{\mathbb{P}}(1)$ , even under assumptions of dependence

of the  $(\varepsilon_i)$ . However, we are here in a very parametric setting that is much more restrictive than the general non-parametric framework described above.

Non-parametric approaches also exist in the literature. We can mention, in particular, the work based on a difference in empirical cumulative distribution functions by Carlstein (1988) and Ben Hariz *et al.* (2007), or statistical learning approaches such as that of Arlot *et al.* (2019). However, this work cannot be applied to data that is not identically distributed according to a linear model for  $1 \leq t \leq n$ , and the nonparametric assumption in Assumption A for  $t \geq n + 1$  is much more general than the one presented in this work. The only statistical learning approach we found, based on the difference in likelihoods of regularized Gaussian processes and use of bootstrap, and which can be likened to the general framework we considered, is the one developed in Avanesov (2019). However the framework of a pure linear model for  $t \leq n$  is not taken into account, and the modeling for  $t \geq n + 1$  is too restrictive compared to that allowed by Assumption A.

We therefore opted for an approach already studied in Bardet (2025) based on the residual sum of squares penalized appropriately, i.e., an approach similar to the  $R_{adjusted}^2$  criterion used for the selection of exogenous variables. However, we have chosen a new penalty function for this criterion, which is written as follows:

$$\hat{n} = \arg \min_{t \leq N} \{\widehat{Cr}(t)\} \quad \text{where} \quad \widehat{Cr}(t) = \left(1 + M \frac{N}{t \log^\beta(t)}\right) \widehat{\sigma}^2(t),$$

where  $\widehat{\sigma}^2(t)$  is the usual unbiased estimator of the error variance calculated from  $((Y_i, Z_i))_{1 \leq i \leq t}$ ,  $M$  and  $\beta$  are positive real constants.

This new criterion has been shown to significantly improve the results obtained in Bardet (2025). Firstly, convergence results can now be obtained in the case where errors follow sub-exponential distributions, whereas previous results and techniques were purely Gaussian. Secondly, we improve the convergence rate of the estimator: it is possible to estimate  $n$  at a rate of  $\mathcal{O}_{\mathbb{P}}(\log^\kappa(N))$  with  $\kappa$  being arbitrarily close to 0, which is close to  $\mathcal{O}_{\mathbb{P}}(1)$  obtained in the context of linear model switching (see above). Furthermore, we also show that an estimator of the parameter  $\theta$  satisfies asymptotic normality at optimal speed under fairly weak assumptions.

Subsequently, we will present the framework and assumptions in detail in Section 2, and we will obtain the convergence results in Section 3. Finally, all of the proofs will be presented in Section 4.

## 2 Framework and assumptions

Let  $(\varepsilon_i)_{i \in \mathbb{N}}$  be a sequence of random variables such that  $\mathbb{E}[\varepsilon_i] = 0$  and  $\mathbb{E}[\varepsilon_i^2] = \sigma^2 < \infty$  for all  $i \in \mathbb{N}$ . Let  $p \in \mathbb{N}^*$ , known, and  $(Z_i^{(j)})_{1 \leq j \leq p, i \in \mathbb{N}}$  a family of known real numbers (deterministic design) such that  $\text{rank}((Z_i^{(j)})_{1 \leq j \leq p, 1 \leq i \leq p}) = p$ .

Let  $N \geq p$  such that a sample  $(Y_1^{(N)}, \dots, Y_N^{(N)})$  of a triangular array  $(Y_i^{(N)})_{1 \leq i \leq N, p \leq N}$  is observed.

We will use the following notations:

- for any  $u \in \mathbb{R}^k$ ,  $\|u\|^2 := u' u$  where  $u'$  denotes the transpose of  $u$ .
- $[Z]_k := \{Z(k) \theta, \theta \in \mathbb{R}^p\}$  which is a linear subspace of  $\mathbb{R}^k$  for  $k \in \{p, \dots, N\}$ ;
- $H(k) := Z(k)(Z'(k)Z(k))^{-1}Z'(k)$  which is the orthogonal projection matrix on  $[Z]_k$  for  $k \in \{p, \dots, N\}$ ;
- $I(t)$  is the identity matrix of size  $t$ ;
- $Z(t) := (Z_i^{(j)})_{1 \leq i \leq t, 1 \leq j \leq p}$  for  $t \in \{1, \dots, N\}$ ;
- $Y^{(N)}(t) := (Y_i^{(N)})_{1 \leq i \leq t}$  for  $t \in \{1, \dots, N\}$ ;
- $\varepsilon(t) := (\varepsilon_i)_{1 \leq i \leq t}$  for  $t \in \{1, \dots, N\}$ .

In the sequel we will also assume:

**Assumption A:** *There exist  $\theta = (\theta_1, \dots, \theta_p)'$  an unknown vector of  $\mathbb{R}^p$  and  $p+1 \leq n \leq N-p$  an unknown integer number and a sequence of random variables  $(\delta_t)_{t \in \mathbb{N}^*}$ , independent to  $(\varepsilon_t)_{t \in \mathbb{N}}$ , such that*

$$\begin{cases} Y_i^{(N)} &= \sum_{j=1}^p \theta_j Z_i^{(j)} + \varepsilon_i & \text{for all } i \in \{1, \dots, n\} \\ Y_{n+k}^{(N)} &= \sum_{j=1}^p \theta_j Z_{n+k}^{(j)} + \delta_k + \varepsilon_{n+k} & \text{for all } k \in \{1, \dots, N-n\} \end{cases} \quad (2.1)$$

**Assumption B:** *There exists  $C > 0$  and a positive integer random variable  $k_0$ , which do not depend on  $N$ , such that for any  $k \geq k_0$*

$$\Delta_k := \sum_{i=1}^k \delta_i^2 \geq Ck \quad \text{and} \quad \max_{1 \leq i \leq k} \delta_i^2 \leq \frac{1}{2p} \Delta_k \quad \text{almost surely.} \quad (2.2)$$

**Assumption C:** *there exists  $0 < \tau < 1$ , an unknown real numbers such that for any  $N \in \mathbb{N}^*$ ,*

$$n = \lceil \tau N \rceil.$$

This last assumption, Assumption C, is fairly standard but could be made less strict, for example by requiring that  $N^\gamma \leq n \leq N - N^\gamma$ ; however, this would make the exponential bounds that we present below much more complex.

The framework of (2.1) under Assumption B is very general and may correspond to:

- A non-parametric change of the trend of  $(Y_i^{(N)})_{1 \leq i \leq N, N \in \mathbb{N}^*}$ . A typical non-parametric example is as follows:

$$\begin{cases} Y_i^{(N)} = \theta_0 + \theta_1 x_i + \varepsilon_i & \text{for } i = 1, \dots, n \\ Y_{n+k}^{(N)} = \theta_0 + \theta_1 x_{n+k} + \varepsilon_{n+k} + g(k) & \text{for } k = 1, \dots, N-n \end{cases}$$

where the function  $g$  is such that there exists  $C > 0$  satisfying  $g^2(k) \geq C$  for any  $k \in \mathbb{N}^*$  and with  $\max_{1 \leq i \leq k} g^2(i) \leq \frac{1}{2p} \sum_{i=1}^k g^2(i)$  for  $k$  large enough.

- A non-parametric change of variance of  $(Y_i^{(N)})_{1 \leq i \leq N, N \in \mathbb{N}^*}$ , which corresponds for instance to an increasing of the volatility. A typical non-parametric example is the following:

$$\begin{cases} Y_i^{(N)} = \theta_0 + \theta_1 x_i + \varepsilon_i & \text{for } i = 1, \dots, n \\ Y_{n+k}^{(N)} = \theta_0 + \theta_1 x_{n+k} + \varepsilon_{n+k} + \xi_k & \text{for } k = 1, \dots, N-n \end{cases}$$

where  $(\xi_i)_{i \in \mathbb{N}}$  is a sequence of independent Gaussian centered random variables such that  $\text{Var}(\xi_i) = \gamma^2$  with  $\gamma^2 > 0$ .

The aim of your study is the estimation of  $n$  and  $\theta$ . For estimating  $n$  under the previous assumptions, we consider a new criterion  $\widehat{Cr}(\cdot)$  such that for any  $t \in \{p+1, \dots, N\}$ ,

$$\widehat{Cr}(t) = \frac{a(t)}{t-p} \|Y^{(N)}(t) - H(t)Y^{(N)}(t)\|^2 \quad \text{with} \quad a(t) := 1 + M \frac{N}{t \log^\beta(t)}, \quad (2.3)$$

with  $M > 0$  and  $\beta > 0$ .

**Remark 1.** *We have used the same approach here for appropriately penalizing the residual sum of squares as that already introduced in Bardet (2025). However, here we have considered the penalty function  $a(t) = 1 + M \frac{N}{t \log^\beta(t)}$  with  $M > 0$  and  $\beta > 0$ , whereas in that paper we considered  $a(t) = 1 + M \frac{N^\alpha}{t^\beta}$  with  $M > 0$  and  $\beta > \alpha > 0$ . This new choice will allow us to improve the convergence rate of the break estimator (see Theorem 3.1).*

Finally, from these definitions and choosing  $0 < \alpha$  which can be arbitrary small, we define the following estimators:

$$\widehat{n} := \arg \min_{[\log^\alpha(N)] \leq t \leq N} \widehat{Cr}(t) \quad \text{and} \quad \widehat{\tau} := \frac{\widehat{n}}{N}. \quad (2.4)$$

**Remark 2.** *Since  $\frac{1}{t-p} \|Y^{(N)}(t) - H(t)Y^{(N)}(t)\|^2$  represents the usual unbiased estimator of variance in a linear model obtained from the residual sum of squares, we see that  $\widehat{Cr}(t)$  is defined as a penalty for this estimator. In Bardet (2025), a detailed heuristic for writing the penalty  $a(t)$  is proposed. To summarize, we can say that a decreasing function  $a(\cdot)$  in  $t$  is needed so that the "largest" linear area is selected, but not too much so as not to search too far for the break in the area where  $t > n$ . Finally, the term  $N$  allows the probability of choosing a "very small" break-point to tend towards 0.*

**Remark 3.** *In Bardet (2025), the convergence of  $\widehat{\tau}$  was demonstrated by defining  $\widehat{n}$  as  $\arg \min_{p \leq t \leq N} \widehat{Cr}(t)$ . Since we based our work on exponential concentration inequalities verified by  $\mathbb{P}(\widehat{Cr}(t) < \widehat{Cr}(n))$ , we were unable to obtain convergence for  $p \leq t < [\log^\alpha(N)]$ . This is not surprising because in Bardet (2025), which gives sharper rates but only in the Gaussian framework, the decay towards 0 of  $\mathbb{P}(\widehat{Cr}(t) < \widehat{Cr}(n))$  is also power-law and not only exponential when  $p \leq t < [\log^\alpha(N)]$ .*

### 3 Convergence rate of the estimators

Now, the consistency of the estimators obtain from the criterion can be established:

**Theorem 3.1.** *Suppose that the distribution of  $\varepsilon_1$  is a sub-exponential distribution, i.e., that:*

$$\|\varepsilon_1\|_{\psi_1} = \inf \left\{ K > 0 : \mathbb{E} \left[ \exp(|\varepsilon_1|/K) \right] \leq 2 \right\} < \infty \quad \text{with} \quad \psi_1(x) = e^x - 1.$$

*Then, under Assumptions A, B and C, for any  $\beta' > 2\beta$ , the estimators  $\widehat{n}$  and  $\widehat{\tau}$  defined in (2.4) satisfy*

$$\mathbb{P} \left( |\widehat{n} - n| \geq \log^{\beta'}(N) \right) \simeq \mathbb{P} \left( |\widehat{\tau} - \tau| \geq \frac{\log^{\beta'}(N)}{N} \right) \xrightarrow[N \rightarrow \infty]{} 0.$$

From this theorem, we deduce that the smaller  $\beta$  the smaller  $\beta'$  and the more accurate the estimation  $\hat{n}$ . Hence, the convergence rate of  $\hat{n}$  is close to the convergence rates  $\mathcal{O}_{\mathbb{P}}(1)$  obtained in the parametric framework of Bai and Perron (1998) or Bai (1998), whereas here we have placed ourselves within the framework of a non-parametric break. It can also be observed that this convergence rate given by Theorem 3.1 is faster than that obtained in Bardet (2025), which was in  $\mathcal{O}_{\mathbb{P}}(N^\gamma)$  for all  $\gamma > 0$ , thanks to the introduction of the new penalty.

It is also possible to show the asymptotic normality of an estimator of  $\theta \in \mathbb{R}^p$  defined in Assumption A. It is clear that it is not possible to use the least squares estimator from  $(Y_1^{(N)}, \dots, Y_N^{(N)})$ , and in the proof of Corollary 1 below, it is also not possible to do so from  $(Y_1^{(N)}, \dots, Y_{\hat{n}}^{(N)})$ . Instead, let us consider the least squares estimator obtained from  $(Y_1^{(N)}, \dots, Y_{\hat{n} - \log^{3\beta}(N)}^{(N)})$ , i.e.

$$\hat{\theta} = (Z'(\tilde{n}) Z(\tilde{n}))^{-1} Z'(\tilde{n}) Y^{(N)}(\tilde{n}) \quad \text{with} \quad \tilde{n} = \hat{n} - \log^{3\beta}(N).$$

However, this requires adding the following assumption:

$$\max_{1 \leq i \leq N} (H(N))_{i,i} \xrightarrow{N \rightarrow \infty} 0, \quad (3.1)$$

which means that the maximum diagonal term of the orthogonal projection matrix  $H(N)$  must tend toward 0 as  $N$  tends toward infinity. We then show that:

**Corollary 1.** *Under the assumptions of Theorem 3.1 and assumption (3.1), then:*

$$(Z'(\tilde{n}) Z(\tilde{n}))^{1/2} (\hat{\theta} - \theta) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_p(0, \sigma^2 I(p)). \quad (3.2)$$

From this corollary, we can obtain confidence intervals on  $\theta$  or Wald tests on linear combinations of  $\theta$ . Regarding , if we consider a fairly classic case where  $\frac{1}{N^s} Z'(N) Z(N) \xrightarrow[N \rightarrow \infty]{\mathcal{P}} Q$  with  $s > 0$  and  $Q$  a definite positive square matrix of size  $p$ , then by Slutsky's Lemma we can show that the convergence rate of (3.2) is asymptotically  $(Z'(n) Z(n))^{1/2}$ , i.e., the same as we would have obtained if we had known the breakpoint.

## 4 Proofs

**Lemma 4.1.** *For  $p + 1 \leq t \leq n - 1$ ,*

$$\mathbb{P}(\widehat{Cr}(t) < \widehat{Cr}(n)) \leq \exp\left(-C_1 t \mathbf{1}_{p < t < [N \log^{-\beta}(N)]} - \frac{C_2 N}{\log^{2\beta}(N)} \mathbf{1}_{[N \log^{-\beta}(N)] \leq t < [n/2]} - \frac{C_3 (n-t)}{\log^{2\beta}(N)} \mathbf{1}_{[n/2] \leq t < n}\right) \quad (4.1)$$

$$\text{where} \begin{cases} C_1 = (1 - \kappa_1) C_{HW} \left( \frac{\sigma_\varepsilon^2 M}{\|\varepsilon_1\|_{\psi_1}^2 (M+1)} \right) \min\left(1, \frac{\sigma_\varepsilon^2 M}{\|\varepsilon_1\|_{\psi_1}^2 (M+1)}\right) \\ C_2 = (2 - \kappa_2) C_{HW} \left( \frac{\sigma_\varepsilon^2 M}{(1+M) \|\varepsilon_1\|_{\psi_1}^2} \right)^2 \\ C_3 = C_{HW} \left( \frac{\sigma_\varepsilon^4 M^2}{4 \tau^2 \|\varepsilon_1\|_{\psi_1}^4} \right) \end{cases}$$

with  $C_{HW}$  a universal constant defined in (4.3), and for any  $\kappa_1, \kappa_2 > 0$  and for  $N$  large enough.

*Proof of Lemma 4.1.* In such a case, i.e. for  $p \leq t < n$  but also for  $t = n$ , we first notice from Assumption A that

$$\widehat{C}r(t) = \frac{a(t)}{t-p} \|(I(t) - H(t))\varepsilon(t)\|^2.$$

But  $\|(I(t) - H(t))\varepsilon(t)\|^2 = \|V_t\|^2$  with  $V_t = P_{[Z]_n^\perp} P_{J_{t,n}} \varepsilon(n)$  where  $P_{[Z]_n^\perp}$  and  $P_{J_{t,n}}$  are respectively the matrices of the orthogonal projections in  $\mathbb{R}^n$  on  $[Z]_n^\perp$  and  $J_{t,n}$ , where  $J_{t,n}$  is the subspace of  $\mathbb{R}^n$  such that  $J_{t,n} = \{(x_1, \dots, x_t, 0, \dots, 0), (x_1, \dots, x_t) \in \mathbb{R}^t\}$ .

Moreover, since  $\|(I(n) - H(n))\varepsilon(n)\|^2 = \|P_{[Z]_n^\perp} \varepsilon(n)\|^2$  and  $P_{[Z]_n^\perp} \varepsilon(n) = P_{[Z]_n^\perp} P_{J_{t,n}} \varepsilon(n) + P_{[Z]_n^\perp} P_{J_{t,n}^\perp} \varepsilon(n)$ , from the Pythagorean theorem, we obtain that  $\|(I(n) - H(n))\varepsilon(n)\|^2 = \|V_t\|^2 + \|W_{n-t}\|^2$  where  $W_{n-t} = P_{[Z]_n^\perp} P_{J_{t,n}^\perp} \varepsilon(n) = P_{[Z]_n^\perp \cap J_{t,n}^\perp} \varepsilon(n)$ .

Then, for any  $t \leq n$ :

$$\begin{aligned} \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) &= \mathbb{P}\left(\frac{a(t)}{t-p} \|(I(t) - H(t))\varepsilon(t)\|^2 < \frac{a(n)}{n-p} \|(I(n) - H(n))\varepsilon(n)\|^2\right) \\ &= \mathbb{P}\left(\frac{a(t)}{t-p} \|V_t\|^2 < \frac{a(n)}{n-p} (\|V_t\|^2 + \|W_{n-t}\|^2)\right) \\ &= \mathbb{P}\left(\|W_{n-t}\|^2 - A(t) \|V_t\|^2 > 0\right) \quad \text{with} \quad A(t) := \frac{(n-p)a(t)}{(t-p)a(n)} - 1. \end{aligned}$$

It is clear that  $\|W_{n-t}\|^2 - A(t) \|V_t\|^2$  is a quadratic form of the vector  $\varepsilon(n)$ . Then we can write that:

$$\mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) = \mathbb{P}\left(\varepsilon(n)' Q_{t,n} \varepsilon(n) - \mathbb{E}[\varepsilon(n)' Q_{t,n} \varepsilon(n)] > -\mathbb{E}[\varepsilon(n)' Q_{t,n} \varepsilon(n)]\right) \quad (4.2)$$

with  $\varepsilon(n)' Q_{t,n} \varepsilon(n) = \|W_{n-t}\|^2 - A(t) \|V_t\|^2$  which is a quadratic form of the vector  $\varepsilon(n)$ .

Now we apply the sub-exponential version of the Hanson-Wright inequality (concentration inequality for quadratic forms of i.i.d.r.v.), see for instance Hanson and Wright (1971) or Rudelson and Vershynin (2013). Hence, from (4.2), such inequality can be written as follows: there exists  $C_{HW} > 0$  such that for all  $\delta \geq 0$ ,

$$\mathbb{P}(\varepsilon(n)' Q_{t,n} \varepsilon(n) - \mathbb{E}[\varepsilon(n)' Q_{t,n} \varepsilon(n)] \geq \delta) \leq \exp\left(-C_{HW} \min\left(\frac{\delta^2}{\|\varepsilon_1\|_{\psi_1}^4 \|Q_{t,n}\|_F^2}, \frac{\delta}{\|\varepsilon_1\|_{\psi_1}^2 \|Q_{t,n}\|_2}\right)\right) \quad (4.3)$$

with  $\mathbb{E}[\varepsilon(n)' Q_{t,n} \varepsilon(n)] = \mathbb{E}[\varepsilon_1^2] \text{tr}(Q_{t,n})$ , and  $\|Q_{t,n}\|_F = \sqrt{\text{tr}(Q_{t,n}' Q_{t,n})}$  is the Frobenius norm of  $Q_{t,n}$ , while  $\|Q_{t,n}\|_2$  is its operator (spectral) norm and  $\|\varepsilon_1\|_{\psi_1} = \inf\{C > 0, \mathbb{E}[\exp(|\varepsilon_1|/C)] \leq 2\}$ .

In the specific context of our work, we have:

$$\begin{cases} \delta &= -\mathbb{E}[\varepsilon(n)' Q_{t,n} \varepsilon(n)] = -\sigma_\varepsilon^2((n-t) - (t-p)A(t)) \\ \|Q_{t,n}\|_2 &\leq \max\{1, |A(t)|, |1 - A(t)|\} \\ \|Q_{t,n}\|_F &= \sqrt{(n-t) + tA^2(t)} \end{cases} \quad (4.4)$$

We will try to give bounds for each of these terms in order to simplify the exponential part of (4.3) a little. A simple computation allows us to write that we also have:

$$\delta = \sigma_\varepsilon^2(n-p) \left(\frac{a(t)}{a(n)} - 1\right). \quad (4.5)$$

- Firstly, we consider the case  $t \in \{[n/2], [n/2] + 1, \dots, n\}$ .

From (4.5) and usual Taylor-Lagrange expansion, with  $h(t) = t^{-1} \log^{-\beta}(t)$ , we have:

$$\begin{aligned} \frac{a(t)}{a(n)} - 1 &= \frac{a(t) - a(n)}{a(n)} \\ &= \frac{MN}{a(n)} |h'(c_{t,n})| |n - t| \quad \text{with } t \leq c_{t,n} \leq n \\ &\geq (1 - \kappa) MN n^{-2} \log^{-\beta}(N) (n - t) \end{aligned}$$

for any  $\kappa > 0$  when  $N$  is large enough and therefore from (4.5),

$$\delta \geq ((1 - \kappa) \sigma_\varepsilon^2 M \tau^{-1}) \log^{-\beta}(N) (n - t). \quad (4.6)$$

Moreover, for any  $\kappa > 0$ ,

$$\begin{aligned} A(t) &= \frac{(n-p)a(t) - (t-p)a(n)}{(t-p)a(n)} \\ &\leq (1 + \kappa) \frac{(n-t)}{t} + MN \left( \frac{g(n) - g(t)}{n \log^\beta(n) t \log^\beta(t)} \right) \quad \text{with } g(t) = (t-p)t \log^\beta(t) \\ &\leq (1 + \kappa) \frac{(n-t)}{t} + MN \left( \frac{\sup_{t \leq c \leq n} |g'(c)|}{n \log^\beta(n) t \log^\beta(t)} \right) (n-t) \\ &\leq (1 + \kappa) \frac{(n-t)}{t} + \frac{5MN}{n \log^\beta(n)} (n-t) \quad \text{since } t \geq [n/2] \\ &\leq (1 + 2\kappa) \frac{(n-t)}{t}, \end{aligned}$$

for  $N$  large enough. Therefore,  $\|Q_{t,n}\|_2 \leq 2$  and for any  $\kappa' > 0$  and when  $N$  is large enough,  $\|Q_{t,n}\|_F^2 = (n-t) + tA^2(t) \leq (3 + \kappa')(n-t)$ . As a consequence:

$$\begin{aligned} \min \left( \frac{\delta^2}{\|\varepsilon_1\|_{\psi_1}^4 \|Q_{t,n}\|_F^2}, \frac{\delta}{\|\varepsilon_1\|_{\psi_1}^2 \|Q_{t,n}\|_2} \right) \\ \geq \left( \frac{(1 - \kappa) \sigma_\varepsilon^2 M \tau^{-1}}{2 \|\varepsilon_1\|_{\psi_1}^2} \right) \frac{(n-t)}{\log^\beta(n)} \min \left( 1, \left( \frac{2(1 - \kappa) \sigma_\varepsilon^2 M \tau^{-1}}{(3 + \kappa') \|\varepsilon_1\|_{\psi_1}^2} \right) \frac{1}{\log^\beta(n)} \right) \\ \geq \left( \frac{\sigma_\varepsilon^4 M^2}{4 \tau^2 \|\varepsilon_1\|_{\psi_1}^4} \right) \frac{(n-t)}{\log^{2\beta}(N)}, \quad (4.7) \end{aligned}$$

for  $N$  large enough.

- Secondly, we consider the case  $t \in \{p+1, \dots, [n/2] - 1\}$ .

In such a case, for any  $\kappa > 0$ ,  $\frac{a(t)}{a(n)} - 1 = \frac{a(t) - a(n)}{a(n)} \geq (1 - \kappa) MN t^{-1} \log^{-\beta}(t)$  for  $N$  large enough, and therefore from (4.5), for any  $\kappa' > 0$ , for  $N$  large enough

$$\delta \geq (1 - \kappa') \sigma_\varepsilon^2 MN t^{-1} \log^{-\beta}(t) n. \quad (4.8)$$

We also have  $A(t) = \frac{(n-p)a(t) - (t-p)a(n)}{(t-p)a(n)} \leq \frac{n}{t} \left( 1 + M \frac{N}{t \log^\beta(t)} \right)$ , which induces two different behaviors:

1. If  $t \in \{[N \log^{-\beta}(N)], \dots, [n/2]-1\}$ , then  $A(t) \leq (1+\kappa+M) \frac{n}{t}$  and this implies when  $N$  is large enough, that  $\|Q_{t,n}\|_2 \leq (1+\kappa+M) \frac{n}{t}$  and  $\|Q_{t,n}\|_F^2 = (n-t) + t A^2(t) \leq (2+(1+\kappa+M)^2) \frac{n^2}{t}$ . Then,

$$\begin{cases} \frac{\delta}{\|Q_{t,n}\|_2} \geq \frac{(1-\kappa') \sigma_\varepsilon^2 M}{(1+\kappa+M)} N \log^{-\beta}(t) \geq (1-\kappa'') \frac{\sigma_\varepsilon^2 M}{(1+M)} N \log^{-\beta}(N) \\ \frac{\delta^2}{\|Q_{t,n}\|_F^2} \geq (1-\kappa')^2 \left( \frac{\sigma_\varepsilon^4 M^2}{2+(1+\kappa+M)^2} \right) \frac{N^2}{t \log^{2\beta}(t)} \geq (2-\kappa'') \left( \frac{\sigma_\varepsilon^2 M}{1+M} \right)^2 \frac{N}{\log^{2\beta}(N)} \end{cases} .$$

As a consequence, for any  $\kappa'' > 0$ , for  $N$  large enough,

$$\min \left( \frac{\delta^2}{\|\varepsilon_1\|_{\psi_1}^4 \|Q_{t,n}\|_F^2}, \frac{\delta}{\|\varepsilon_1\|_{\psi_1}^2 \|Q_{t,n}\|_2} \right) \geq (2-\kappa'') \left( \frac{\sigma_\varepsilon^2 M}{(1+M) \|\varepsilon_1\|_{\psi_1}^2} \right)^2 \frac{N}{\log^{2\beta}(N)}. \quad (4.9)$$

2. If  $t \in \{p+1, \dots, [N \log^{-\beta}(N)] - 1\}$ , then  $A(t) \leq (1+M) \frac{n}{t} \frac{N}{t \log^\beta(t)}$  and this implies, that  $\|Q_{t,n}\|_2 \leq (1+M) \frac{n}{t} \frac{N}{t \log^\beta(t)}$  and  $\|Q_{t,n}\|_F^2 = (n-t) + t A^2(t) \leq (1+\kappa+M)^2 \frac{n^2 N^2}{t^3 \log^{2\beta}(t)}$  when  $N$  is large enough. Then

$$\begin{cases} \frac{\delta}{\|Q_{t,n}\|_2} \geq \frac{(1-\kappa') \sigma_\varepsilon^2 M N t^{-1} \log^{-\beta}(t) n}{(1+M) \frac{n}{t} \frac{N}{t \log^\beta(t)}} \geq (1-\kappa') \sigma_\varepsilon^2 \left( \frac{M}{1+M} \right) t \\ \frac{\delta^2}{\|Q_{t,n}\|_F^2} \geq \frac{(1-\kappa')^2 \sigma_\varepsilon^4 M^2 N^2 t^{-2} \log^{-2\beta}(t) n^2}{(1+\kappa+M)^2 \frac{n^2 N^2}{t^3 \log^{2\beta}(t)}} \geq (1-\kappa'') \sigma_\varepsilon^4 \left( \frac{M}{1+M} \right)^2 t. \end{cases} .$$

As a consequence, for any  $\kappa' > 0$ , for  $N$  large enough,

$$\begin{aligned} \min \left( \frac{\delta^2}{\|\varepsilon_1\|_{\psi_1}^4 \|Q_{t,n}\|_F^2}, \frac{\delta}{\|\varepsilon_1\|_{\psi_1}^2 \|Q_{t,n}\|_2} \right) \\ \geq (1-\kappa') \left( \frac{\sigma_\varepsilon^2 M}{\|\varepsilon_1\|_{\psi_1}^2 (M+1)} \right) \min \left( 1, \frac{\sigma_\varepsilon^2 M}{\|\varepsilon_1\|_{\psi_1}^2 (M+1)} \right) t. \end{aligned} \quad (4.10)$$

From (4.7), (4.9) and (4.10), we deduce (4.1).  $\square$

**Lemma 4.2.** *Let  $n+1 \leq t \leq N$ . Then, with  $C_{HW}$  and  $C_B$  two universal constants respectively defined in (4.3) and (4.19), with  $\tilde{A} \xi C$  defined in Assumption B,*

$$\mathbb{P}(\widehat{Cr}(t) < \widehat{Cr}(n)) \leq \exp(-C_{HW} C_4 (t-n)) + \exp(-C_B \min(C_5 (t-n), C_6 \sqrt{t-n})), \quad (4.11)$$

$$\text{with } \begin{cases} C_4 = (1-\kappa_4) C \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2} \frac{\tau}{4(1-\tau)} \min \left( 1, \frac{C}{4} \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2} \right) \\ C_5 = \frac{C}{64 \|\varepsilon_1\|_{\psi_1}^2} \\ C_6 = \frac{1}{8} \min \left( \frac{\sqrt{C} \sqrt{p}}{\sqrt{2} \|\varepsilon_1\|_{\psi_1}} \right) \end{cases}$$

for any  $\kappa_4 > 0$  and for  $N$  large enough.

*Proof of Lemma 4.2.* Since  $t > n$ , based on Assumption A, we can write that:

$$\widehat{Cr}(t) = \frac{a(t)}{t-p} \left\| (I(t) - H(t)) \varepsilon(t) + (I(t) - H(t)) D(t) \right\|^2 \text{ with } D(t) = (0, \dots, 0, \delta_1, \dots, \delta_{t-n})'$$

where  $(\delta_i)_{i \geq 1}$  is defined in Assumption A.

We therefore have:

$$\begin{aligned} \mathbb{P}\left(\widehat{Cr}(t) < \widehat{Cr}(n)\right) &= \mathbb{P}\left(\frac{a(t)}{t-p} \left\| (I(t) - H(t))(\varepsilon(t) + D(t)) \right\|^2 < \frac{a(n)}{n-p} \left\| (I(n) - H(n))\varepsilon(n) \right\|^2\right) \\ &= \mathbb{P}\left(\frac{a(t)}{t-p} \left\| (V_n + W_{t-n}) + P_{[Z]_t^\perp} D(t) \right\|^2 < \frac{a(n)}{n-p} \|V_n\|^2\right) \\ &= \mathbb{P}\left(\frac{a(t)}{t-p} (\|V_n + P_{[Z]_t^\perp} D(t)\|^2 + \|W_{t-n}\|^2) < \frac{a(n)}{n-p} \|V_n\|^2\right) \\ &= \mathbb{P}\left(B(t) \|V_n\|^2 + \|W_{t-n}\|^2 + 2D(t)' P_{[Z]_t^\perp} V_n + \|P_{[Z]_t^\perp} D(t)\|^2 < 0\right), \end{aligned}$$

where  $B(t) = 1 - \frac{(t-p)a(n)}{(n-p)a(t)}$ , and  $V_n$  and  $W_{t-n}$  defined as in the first part of the proof.

As in Lemma 4.1, it is interesting to write:

$$B(t) \|V_n\|^2 + \|W_{t-n}\|^2 = \varepsilon(t)' R_{n,t} \varepsilon(t),$$

where  $R_{n,t}$  is a symmetric matrix. We can therefore write that:

$$\begin{aligned} \mathbb{P}\left(\widehat{Cr}(t) < \widehat{Cr}(n)\right) &= \mathbb{P}\left(\varepsilon(t)' R_{n,t} \varepsilon(t) + 2D(t)' P_{[Z]_t^\perp} V_n < -\|P_{[Z]_t^\perp} D(t)\|^2\right) \\ &\leq \mathbb{P}\left(\varepsilon(t)' R_{n,t} \varepsilon(t) - \mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] < -\mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] - \frac{1}{2} \|P_{[Z]_t^\perp} D(t)\|^2\right) \\ &\quad + \mathbb{P}\left(D(t)' P_{[Z]_t^\perp} V_n < -\frac{1}{4} \|P_{[Z]_t^\perp} D(t)\|^2\right). \end{aligned} \quad (4.12)$$

We will use conditional probabilities given the random vector  $D(t)$ . Consider first  $I_1$  such as:

$$I_1 = \mathbb{P}\left(\varepsilon(t)' R_{n,t} \varepsilon(t) - \mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] < -\mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] - \frac{1}{2} \|P_{[Z]_t^\perp} D(t)\|^2 \mid D(t)\right). \quad (4.13)$$

We will again use the the Hanson-Wright concentration inequality (4.3) for quadratic forms of independent sub-exponential random variables to upper bound this probability. Thus, we have:

- $\mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] = \sigma_\varepsilon^2 \text{tr}(R_{n,t}) = \sigma_\varepsilon^2 ((n-p)B(t) + (t-n)) = \sigma_\varepsilon^2 (t-p) \frac{a(t) - a(n)}{a(t)}$ ;
- $\|R_{n,t}\|_2 \leq \max(1, |B(t)|, |B(t) + 1|)$ ;
- $\|R_{n,t}\|_F^2 = (n-p)B^2(t) + (t-n)$ ;
- $\|P_{[Z]_t^\perp} D(t)\|^2 \geq \frac{1}{2} C(t-n)$  for  $N$  large enough from Lemma 4.1 of Bardet (2025).

With  $t \in \{n+1, \dots, N\}$ , we have from a Taylor-Lagrange expansion, with  $n \leq \theta_{n,t} \leq t$  and for any  $\kappa > 0$

$$\begin{aligned} B(t) &= \frac{n-t}{(n-p)a(t)} \left(1 + \frac{MN g'(c)}{n \log^\beta(n) t \log^\beta(t)}\right) \text{ with } n < c \leq t \text{ and } g(t) = (t-p)t \log^\beta(t) \\ &\implies (1 + \kappa) \frac{n-t}{n-p} \leq B(t) \leq (1 - \kappa) \frac{n-t}{n-p}, \end{aligned}$$

for  $N$  large enough. Moreover, for any  $\kappa > 0$ ,

$$\begin{aligned} \mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] &= \sigma_\varepsilon^2 M \frac{N}{a(t)} \left( \frac{1}{t \log^\beta(t)} - \frac{1}{n \log^\beta(n)} \right) \\ &\simeq \sigma_\varepsilon^2 \frac{MN}{n \log^\beta(n)} (n-t) \\ \implies -\mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] - \frac{1}{2} \|P_{[Z]_t^\perp} D(t)\|^2 &\leq (1-\kappa) \frac{1}{4} C (n-t) \\ \text{and this implies } |\delta| &\geq (1-\kappa) \frac{1}{4} C (t-n) \end{aligned} \quad (4.14)$$

when  $N$  is large enough.

Moreover,

$$\begin{cases} \|R_{n,t}\|_2 &\leq \max\left((1+\kappa) \frac{t-n}{n}, 1\right) \\ \|R_{n,t}\|_F^2 &\leq (1+\kappa) \frac{t}{n} (t-n) \end{cases} \quad (4.15)$$

As a consequence, from (4.14) and (4.15), for any  $\kappa' > 0$ , for  $N$  large enough,

$$\begin{aligned} \min\left(\frac{\delta^2}{\|\varepsilon_1\|_{\psi_1}^4 \|R_{t,n}\|_F^2}, \frac{|\delta|}{\|\varepsilon_1\|_{\psi_1}^2 \|R_{t,n}\|_2}\right) \\ \geq (1-\kappa') C \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2} \frac{\tau}{4(1-\tau)} \min\left(1, \frac{C}{4} \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2}\right) (t-n). \end{aligned} \quad (4.16)$$

And from (4.13), applying an expectation, we deduce:

$$\begin{aligned} \mathbb{P}\left(\varepsilon(t)' R_{n,t} \varepsilon(t) - \mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] < -\mathbb{E}[\varepsilon(t)' R_{n,t} \varepsilon(t)] - \frac{1}{2} \|P_{[Z]_t^\perp} D(t)\|^2\right) \\ \leq \exp\left(-C_{HW} (1-\kappa') C \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2} \frac{\tau}{4(1-\tau)} \min\left(1, \frac{C}{4} \frac{\sigma_\varepsilon^2}{\|\varepsilon_1\|_{\psi_1}^2}\right) (t-n)\right). \end{aligned} \quad (4.17)$$

Now, for bounding the term  $\mathbb{P}\left(D(t)' P_{[Z]_t^\perp} V_n < -\frac{1}{4} \|P_{[Z]_t^\perp} D(t)\|^2\right)$  of (4.12), define

$$I_2 = \mathbb{P}\left(D(t)' P_{[Z]_t^\perp} V_n < -\frac{1}{4} \|P_{[Z]_t^\perp} D(t)\|^2 \mid D(t)\right). \quad (4.18)$$

Here, we will use a Bernstein Inequality for linear combination of sub-exponential zero-mean i.i.d. random variables  $(Z_j)_{1 \leq j \leq t}$  as it can be found in Chapter 2 of Boucheron *et al.* (2013). Hence, there exists a constant  $C_B > 0$  such that for any  $U \in \mathbb{R}^t$  and any  $\delta' > 0$ ,

$$\mathbb{P}(U' Z < -\delta') \leq \exp\left(-C_B \min\left(\frac{\delta'}{\|\varepsilon_1\|_{\psi_1} \|U\|_\infty}, \frac{\delta'^2}{\|\varepsilon_1\|_{\psi_1}^2 \|U\|_2^2}\right)\right). \quad (4.19)$$

Using Assumption (2.2) and from Lemma 4.1 of Bardet (2025), we have  $\delta' \geq \frac{1}{8} \Delta_{t-n}$  and  $\|U\|_\infty \leq$

$\sqrt{\frac{1}{p} \Delta_{t-n}}$  for  $N$  large enough, where  $\Delta_k := \sum_{i=1}^k \delta_i^2$ . Finally, we also have  $\delta'^2 \geq \frac{1}{32} \Delta_{t-n}$  and  $\|U\|_2^2 \leq \Delta_{t-n}$ . Therefore, from (4.19),

$$I_2 \leq \exp\left(-C_B \min\left(\frac{\sqrt{C} \sqrt{p}}{8 \sqrt{2} \|\varepsilon_1\|_{\psi_1}} \sqrt{t-n}, \frac{C}{64 \|\varepsilon_1\|_{\psi_1}^2} (t-n)\right)\right).$$

Once again, using an expectation, we deduce that

$$\mathbb{P}\left(D(t)' P_{[Z]_t^\perp} V_n < -\frac{1}{4} \|P_{[Z]_t^\perp} D(t)\|^2\right) \leq \exp\left(-\frac{C_B}{8} \min\left(\frac{\sqrt{C} \sqrt{p}}{\sqrt{2} \|\varepsilon_1\|_{\psi_1}} \sqrt{t-n}, \frac{C}{8 \|\varepsilon_1\|_{\psi_1}^2} (t-n)\right)\right). \quad (4.20)$$

Using the decomposition (4.12) and exponential concentration inequalities (4.17) and (4.20), we deduce (4.11).  $\square$

*Proof of Theorem 3.1.* For any  $\beta' \geq 2\beta$  and any  $\alpha > 0$ , we have:

$$\begin{aligned} \mathbb{P}(|\hat{n} - n| \geq \log^{\beta'}(N)) &\leq \mathbb{P}\left(\max\{\widehat{C}r(t), t \in \{[\log^\alpha(N)], \dots, N\} \text{ such that } |t-n| \geq \log^{\beta'}(N)\} < \widehat{C}r(n)\right) \\ &\leq \sum_{t=[\log^\alpha(N)]}^{n-\log^{\beta'}(N)} \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) + \sum_{t=n+\log^{\beta'}(N)}^N \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right). \end{aligned} \quad (4.21)$$

Now Lemmas 4.1 and 4.2 can be applied. As a consequence,

$$\begin{aligned} \sum_{t=[\log^\alpha(N)]}^{n-\log^{\beta'}(N)} \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) &\leq \sum_{t=[\log^\alpha(N)]}^{[N \log^{-\beta}(N)]-1} e^{-C_1 t} + \sum_{t=[N \log^{-\beta}(N)]}^{[n/2]-1} e^{-\frac{C_2 N}{\log^{2\beta}(N)}} + \sum_{t=[n/2]-1}^{n-\log^{\beta'}(N)} e^{-\frac{C_3 (n-t)}{\log^{2\beta}(N)}} \\ &\leq \frac{e^{-C_1 [\log^\alpha(N)]}}{1 - e^{-C_1}} + \frac{\tau}{2} N e^{-\frac{C_2 N}{\log^{2\beta}(N)}} + \frac{e^{-C_3 \log^{\beta'} - 2\beta}(N)}{1 - e^{-C_3 \log^{-2\beta}(N)}} \\ &\leq \frac{e^{-C_1 [\log^\alpha(N)]}}{1 - e^{-C_1}} + \frac{\tau}{2} N e^{-\frac{C_2 N}{\log^{2\beta}(N)}} + \frac{2}{C_3} \log^{2\beta}(N) e^{-C_3 (\log(N))^{\beta'} - 2\beta} \\ &\xrightarrow{N \rightarrow \infty} 0 \end{aligned} \quad (4.22)$$

for  $N$  large enough. Similarly, we have:

$$\begin{aligned} \sum_{t=n+\log^{\beta'}(N)}^N \mathbb{P}\left(\widehat{C}r(t) < \widehat{C}r(n)\right) &\leq \sum_{t=n+\log^{\beta'}(N)}^N e^{-C_4 (t-n)} + e^{-\min(C_5 (t-n), C_6 \sqrt{t-n})} \\ &\leq e^{-C_4 \log^{\beta'}(N)} + e^{-C_5 \log^{\beta'}(N)} + \sum_{k=\log^{\beta'}(N)}^{N-n} e^{-C_6 \sqrt{k}} \\ &\leq e^{-C_4 \log^{\beta'}(N)} + e^{-C_5 \log^{\beta'}(N)} + \frac{2}{C_6} \log^{\beta'/2}(N) e^{-C_6 \log^{\beta'/2}(N)} \\ &\xrightarrow{N \rightarrow \infty} 0, \end{aligned} \quad (4.23)$$

using a sum-integral comparison. Therefore, from (4.21), from (4.22) and (4.23), we deduce Theorem 3.1.  $\square$

*Proof of Corollary 1.* From Huber (1981), page 157, the Lindeberg central limit theorem allows us to write that under Assumption A, for all  $t \leq n$ ,

$$(Z'(t) Z(t))^{1/2} (\hat{\theta}(t) - \theta) \xrightarrow[t \rightarrow \infty]{\mathcal{L}} \mathcal{N}_p(0, \sigma^2 I(p)), \quad (4.24)$$

with  $\hat{\theta}(t) = (Z'(t) Z(t))^{-1} Z'(t) Y^{(N)}(t)$  the ordinary least square estimator.

Let  $x \in \mathbb{R}^p$  and denote  $\widehat{U}^{(N)} = (Z'(\tilde{n}) Z(\tilde{n}))^{1/2} (\hat{\theta} - \theta)$ . Then, using the notation  $F_{\widehat{U}^{(N)}}(x) = \mathbb{P}(\widehat{U}^{(N)} \leq x) = \mathbb{P}(\widehat{U}_1^{(N)} \leq x_1 \cap \widehat{U}_2^{(N)} \leq x_2 \cap \dots \cap \widehat{U}_p^{(N)} \leq x_p)$  when  $x = (x_1, \dots, x_p)'$ , we obtain:

$$F_{\widehat{U}^{(N)}}(x) = \mathbb{P}(\widehat{U}^{(N)} \leq x \mid \tilde{n} \leq n) \mathbb{P}(\tilde{n} \leq n) + \mathbb{P}(\widehat{U}^{(N)} \leq x \cap \tilde{n} \leq n).$$

But given  $\tilde{n} \leq n$ , then  $\widehat{U}^{(N)} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_p(0, \sigma^2 I(p))$ . Moreover, from the proof of Theorem 3.1,  $\mathbb{P}(\widehat{n} \geq n + N \log^{3\beta}(N)) \xrightarrow[N \rightarrow \infty]{} 0$  and therefore

$$\mathbb{P}(\tilde{n} \leq n) = 1 - \mathbb{P}(\widehat{n} \leq n + N \log^{3\beta}(N)) \xrightarrow[N \rightarrow \infty]{} 1.$$

Moreover  $\mathbb{P}(\widehat{U}^{(N)} \leq x \cap \tilde{n} \leq n) \leq \mathbb{P}(\tilde{n} \leq n) \xrightarrow[N \rightarrow \infty]{} 0$ . Thus, with  $U \sim \mathcal{N}_p(0, \sigma^2 I(p))$ , we have for any  $x \in \mathbb{R}^p$

$$\mathbb{P}(\widehat{U}^{(N)} \leq x) \xrightarrow[N \rightarrow \infty]{} F_U(x),$$

and the asymptotic normality of  $\widehat{\theta}$  can be deduced.  $\square$

## References

- [1] Arlot, A., Celisse, A. and Harchaoui, Z. (2019). A Kernel Multiple Change-point Algorithm via Model Selection. *Journal of Machine Learning Research*, 20, 1-56.
- [2] Avanesov, V. (2019). Nonparametric Change Point Detection in Regression. *Preprint arXiv:1903.02603*.
- [3] Bai, J. (1998). Estimation of multiple-regime regressions with least absolute deviation, *Journal of Statistical Planning and Inference*, **74**, 103–134.
- [4] Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.
- [5] Bardet, J.-M. (2025). A criterion for estimating the largest linear homoscedastic zone in Gaussian data. *Journal of Statistical Planning and Inference*, 235, 106223.
- [6] Ben Hariz, S., Wylie J. and Zhang Q. (2007). Optimal rate of convergence for nonparametric change-point estimators for nonstationary sequences. *The Annals of Statistics* 35, 1802-1826.
- [7] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, Academic Press.
- [8] Carlstein, E. (1988). Nonparametric change-point estimation. *The Annals of Statistics* 16, 188-197.
- [9] Chow, G. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28, p. 591-605.
- [10] Hanson, D.L. and Wright, F.T. (1971). A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Stat.* 42, 1079-1083.
- [11] Huber, P.J. (1981). *Robust statistics*, Wiley Series in Probability and Mathematical Statistics, Wiley & Sons Inc., New York.

- [12] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stoc. Proc. Appl.* 83, 79-102.
- [13] Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* 18, 82.