Supplementary material for ”Statistical Inference for Multiple Change-Point Models”

Wu Wang[a], Xuming He[b], Zhongyi Zhu[c]

[a]*Statistics Program, King Abdullah University of Science and Technology*
[b]*Department of Statistics, University of Michigan*
[c]Department of Statistics, Fudan University

Appendices

A More on simulation study

The details of how do we implement the competing methods are listed as follows,

BS Implemented by the routine `sbs` of the package `wbs`(Baranowski & Fryzlewicz, 2015). The number of change-points is selected by the BIC criterion.

WBS Implemented by the routine `wbs` of the package `wbs`(Baranowski & Fryzlewicz, 2015). The number of change-points is selected by the BIC criterion. The code we used is
`obj<-wbs(x);changepoints(obj, penalty="bic.penalty").`

FLASSO Implemented by the routine `fusedlasso1d` of the package `genlasso`(Arnold & Tibshirani, 2014). The number of change-points is selected by the BIC criterion.

SMUCE Implemented by the routine `smuceR` of the package `stepR`(Pein et al., 2017). We use the default settings which set the parameter α at 0.5. The code we used is `stepFit(x).`

cumSeg Implemented by the routine `jumpoints` of the package `cumSeg`(Muggeo, 2012). We use the default settings. The code we used is `jumpoints(x)`. The number of change-points is selected by the BIC criterion.

PELT Implemented by the routine `cpt.mean` of the package `changepoint`(Killick et al., 2016). The number of change-points is selected by the BIC criterion. The code we used is `cpt.mean(x/mad(diff(x)/sqrt(2)), method="PELT", penalty = "BIC")`, where `mad` function calculates the median absolute deviation. This implementation was advocated by Fryzlewicz (2014).

S3IB Implemented by the routine `Segmentor` of the package `Segmentor3IsBack`(Cleynen et al., 2016). The number of change-points is selected by the BIC criterion. The code we used is

`obj<-Segmentor(x, model=2);SelectModel(obj,penalty='BIC')`.

SaRa The routine is coded by ourselves. The number of change-points is selected by the BIC criterion.

Model 1 is a standard testing signal which is also used in Frick et al. (2014) and Fryzlewicz (2014). Model 2 and 3 have signals in the shape of teeth. Model 2 is a mix of more pronounced changes with short distances between change-points and less pronounced changes with greater distances between change-points. Model 3 has more frequent change-points while the size of the jump is constant. Model 4 has signals in the shape of stairs, the change occurs every 10th observation. The models are considered because they illustrate various patterns of change-points.

We show more simulation results using the same models as in Section 4, and now the errors are generated from an autoregressive (AR) model,

$$\epsilon_t = 0.3\epsilon_{t-1} + \sigma u_t,$$

where σ is the same as the model in Section 4, and u_t is standard normally distributed. The results are collected in Table A.1,A.2 and Figure A.1. Generally speaking, all methods perform worse than the case where the errors are independent because the correlation structure is ignored. Compared to other methods, the proposed method performs better regarding change-point number estimation and change-point location estimation. Regarding the empirical coverage probabilities of the percentile interval, the smoothed interval, and the adaptive interval, they are all below the nominal level. This is mainly because the bootstrapping method and the standard formula for the bagging estimator may not be the best for correlated data. Also notice that the percentile interval, the smoothed interval, and the adaptive interval have higher coverage probabilities compared to the SMUCE interval.

Table A.1: Summary statistics of $\hat{N} - N$ and the Hausdorff distance d_H for model 1 and 2. The error are generated from the AR model.

	$\hat{N} - N$							d_H	
	Frequency(%)								
Method	≤ -2	$= -1$	$= 0$	$= 1$	≥ 2	Mean	Med	Mean	SD
				Model 1					
BootCp	3.6	8.0	**62.8**	8.6	17.0	0.37	0.0	31.48	36.56
BS	37.0	9.4	38.2	12.4	3.0	-0.67	0.0	29.83	28.99
WBS	6.0	5.8	**64.8**	13.6	9.8	0.21	0.0	26.94	34.38
FLASSO	4.2	0.8	1.0	2.4	91.6	6.01	6.0	69.24	34.02
SMUCE	0.0	0.6	8.4	19.8	71.2	2.43	2.0	92.43	41.87
cumSeg	77.6	8.8	10.8	2.2	0.6	-1.61	-2.0	35.89	16.25
PELT	0.0	0.6	10.6	14.2	74.6	3.65	3.0	81.12	52.91
S3IB	0.6	2.8	45.0	24.0	27.6	0.91	1.0	40.71	46.67
SaRa	5.2	11.8	32.6	17.4	33.0	0.90	1.0	55.93	44.74
				Model 2					
BootCp	24.2	20.6	29.8	11.6	13.8	-0.31	0.0	78.09	57.20
BS	60.2	18.2	12.2	5.4	4.0	-2.12	-2.0	162.70	88.65
WBS	36.6	21.0	27.4	9.2	5.8	-0.91	-1.0	86.64	63.95
FLASSO	97.4	0.0	0.6	0.6	1.4	-12.34	-13.0	259.83	140.77
SMUCE	12.0	31.2	36.2	15.0	5.6	-0.29	0.0	48.15	32.18
cumSeg	99.6	0.4	0.0	0.0	0.0	-8.91	-9.0	174.77	89.12
PELT	2.2	5.4	16.2	18.4	57.8	2.34	2.0	48.82	36.32
S3IB	18.8	18.2	**35.8**	27.2	0.0	-0.33	0.0	73.08	56.03
SaRa	5.8	8.6	18.6	15.4	51.6	1.88	2.0	52.96	37.98

Table A.2: Summary statistics of $\hat{N} - N$ and the Hausdorff distance d_H for model 3 and 4. The error are generated from the AR model.

	$\hat{N} - N$							d_H	
	Frequency(%)								
Method	≤ -2	$= -1$	$= 0$	$= 1$	≥ 2	Mean	Med	Mean	SD
				Model 3					
BootCp	32.6	7.4	**53.2**	6.2	0.6	-1.15	0.0	9.08	11.06
BS	99.8	0.0	0.2	0.0	0.0	-12.61	-13.0	111.22	22.06
WBS	34.2	5.4	40.4	13.8	6.2	-2.32	0.0	11.63	20.23
FLASSO	100.0	0.0	0.0	0.0	0.0	-12.96	-13.0	96.50	26.68
SMUCE	83.0	11.6	5.2	0.2	0.0	-3.69	-4.0	13.33	7.09
cumSeg	100.0	0.0	0.0	0.0	0.0	-12.94	-13.0	109.28	23.69
PELT	17.4	9.8	44.2	20.2	8.4	-0.15	0.0	7.07	5.92
S3IB	77.2	6.2	16.4	0.2	0.0	-3.14	-3.0	19.70	16.21
SaRa	30.4	13.4	48.0	6.8	1.4	-0.85	0.0	7.39	6.51
				Model 4					
BootCp	0.2	8.4	**79.0**	12.0	0.4	0.04	0.0	2.74	2.56
BS	9.6	24.2	59.8	6.2	0.2	-0.41	0.0	4.67	3.49
WBS	0.0	1.8	43.0	36.8	18.4	0.80	1.0	3.37	1.97
FLASSO	16.0	1.8	2.4	5.2	74.6	1.61	4.0	37.99	8.04
SMUCE	47.6	26.6	25.2	0.6	0.0	-1.47	-1.0	6.33	3.23
cumSeg	10.8	22.4	60.2	6.6	0.0	-0.45	0.0	4.97	3.22
PELT	0.0	1.8	51.4	31.0	15.8	0.70	0.0	3.12	2.16
S3IB	3.6	23.0	73.4	0.0	0.0	-0.30	0.0	3.58	3.45
SaRa	4.4	22.4	61.8	10.2	1.2	-0.19	0.0	4.60	3.39

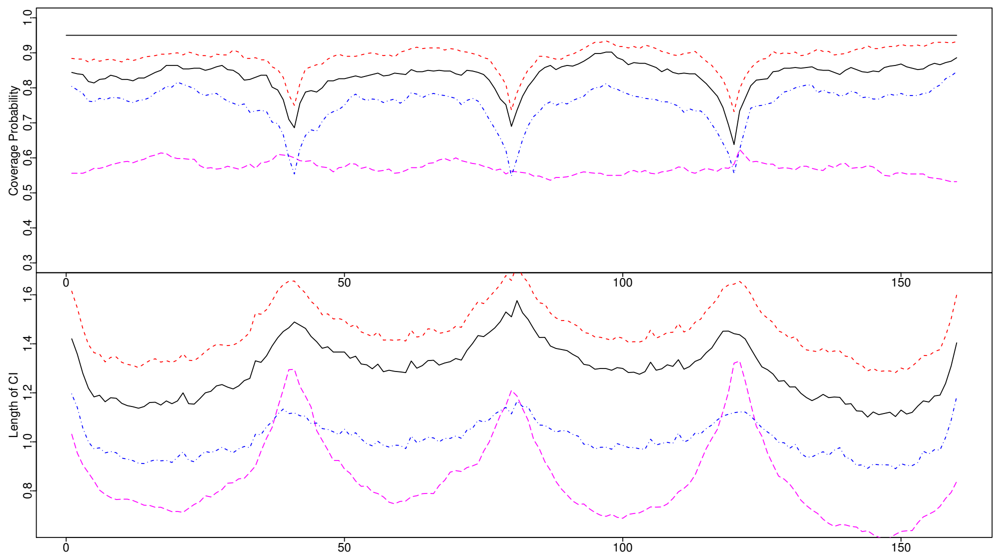

Figure A.1: *The empirical coverage probabilities and average length of the confidence intervals based on 500 simulation runs. From top to bottom, the lines are percentile intervals, adaptive intervals, smoothed intervals, SMUCE intervals.*

B Proof of the theorems

We will prove the theorems under a model with two change-points. Since the change-points are estimated sequentially, the theory can be generalized to a model with multiple change-points easily. This technique for proving the convergence rate of the change-points estimator is advocated by Bai (1997). The model under consideration in this section is

$$\begin{aligned} X_t &= \beta_1^0 + \epsilon_t, \quad 1 \le t \le t_1^0, \\ X_t &= \beta_2^0 + \epsilon_t, \quad t_1^0 + 1 \le t \le t_2^0, \\ X_t &= \beta_3^0 + \epsilon_t, \quad t_2^0 + 1 \le t \le n. \end{aligned}$$

Because the BS algorithm is implemented sequentially in this paper, we will consider the convergence rate of the first change-point detected by the BS algorithm. The weighted residual sum of squares is defined as

$$L_n^w(k) = \sum_{t=1}^{k} w_t (X_t - \bar{X}_k^w)^2 + \sum_{t=k+1}^{n} w_t (X_t - \bar{X}_{-k}^w)^2, \tag{B.1}$$

where $\bar{X}_k^w = \sum_{t=1}^k w_t X_t / \sum_{t=1}^k w_t$ and $\bar{X}_{-k}^w = \sum_{t=k+1}^n w_t X_t / \sum_{t=k+1}^n w_t$ for an integer

$k \in [1, n-1]$. The first detected change-point is defined as

$$\hat{t}^w = \underset{k=1,\ldots,n-1}{\arg\min}\, L_n^w(k). \tag{B.2}$$

To prove Theorem 1, we need to investigate the convergence property of $L_n^w(k)$ defined in (B.1). We first study the asymptotic behavior of $U_n^w(\tau) = n^{-1} L_n^w(\lfloor n\tau \rfloor)$ under the assumptions A1-A3, where $\lfloor n\tau \rfloor$ is the integer part of $n\tau$. The intermediate results are stated in Lemma 1 to Lemma 5.

Lemma 1. *Under Assumptions A1-A3,* $\sup_{\tau\in[0,1]} |U_n^w(\tau) - U(\tau)| = o_{\mathbb{P}_w}(1)$ *for a deterministic function* $U(\tau)$.

Lemma 2. *Under Assumptions A1-A3,* $\max\limits_{1\le k<n} |U_n^w(k/n) - \bar{U}_n(k/n)| = o_{\mathbb{P}_w}(n^{-1/2}\log(n))$ *for a deterministic function* $\bar{U}_n(k/n)$.

See (B.9) for the definition of $\bar{U}_n(k/n)$. The function $U(\tau)$ has different expressions over three different regimes,

$$U(\tau) = \begin{cases} (\tau_1^0 - \tau)a(\tau)^2 + (\tau_2^0 - \tau_1^0)b(\tau)^2 + (1-\tau_2^0)c(\tau)^2 & \text{if } \tau \in [0, \tau_1^0], \\ \frac{\tau_1^0(\tau-\tau_1^0)}{\tau}(\beta_2^0 - \beta_1^0)^2 + \frac{(\tau_2^0-\tau)(1-\tau_2^0)}{1-\tau}(\beta_3^0 - \beta_2^0)^2 & \text{if } \tau \in (\tau_1^0, \tau_2^0), \\ \tau_1^0 d(\tau)^2 + (\tau_2^0 - \tau_1^0)e(\tau)^2 + (\tau - \tau_2^0)f(\tau)^2 & \text{if } \tau \in [\tau_2^0, 1], \end{cases} \tag{B.3}$$

where $a(\tau) = \frac{1-\tau_1^0}{1-\tau}(\beta_1^0 - \beta_2^0) + \frac{1-\tau_2^0}{1-\tau}(\beta_2^0 - \beta_3^0)$, $b(\tau) = \frac{\tau_1^0-\tau}{1-\tau}(\beta_2^0 - \beta_1^0) + \frac{1-\tau_2^0}{1-\tau}(\beta_2^0 - \beta_3^0)$, $c(\tau) = \frac{\tau_1^0-\tau}{1-\tau}(\beta_2^0 - \beta_1^0) + \frac{\tau_2^0-\tau}{1-\tau}(\beta_3^0 - \beta_2^0)$, $d(\tau) = \frac{\tau-\tau_1^0}{\tau}(\beta_1^0 - \beta_2^0) + \frac{1-\tau_2^0}{\tau}(\beta_2^0 - \beta_3^0)$, $e(\tau) = \frac{\tau_1^0}{\tau}(\beta_2^0 - \beta_1^0) + \frac{\tau-\tau_2^0}{\tau}(\beta_2^0 - \beta_3^0)$, and $f(\tau) = \frac{\tau_1^0}{\tau}(\beta_2^0 - \beta_1^0) + \frac{\tau_2^0}{\tau}(\beta_3^0 - \beta_2^0)$. Note that $U(\tau)$ is also the limit of $U_n(\tau) = L_n(\lfloor n\tau \rfloor; 0, n)/n$, see Bai (1997). From Lemma 2, we know that $U_n^w(\tau)$ is asymptotically close to $\bar{U}_n(k/n)$ given almost all sample paths of $X_1, X_2, \ldots$ at the rate $o_{\mathbb{P}_w}(n^{-1/2}\log(n))$. We expect that the minimizer of $U_n^w(\tau)$ will be close to the minimizer of $\bar{U}_n(k/n)$ too. The function $U(\tau)$ has two local minima at τ_1^0 or τ_2^0 as explained by Bai (1997). A convenient condition would be that $U(\tau)$ has a well pronounced minimizer. In the following, we will assume the following assumption.

Assumption A4. $U(\tau_1^0) < U(\tau_2^0)$.

Bai (1997) discussed this case where $U(\tau_1^0) = U(\tau_2^0)$ in-depth, and he found that in this case the first change-point detected by the BS algorithm converges to a random variable with equal mass at t_1^0 and t_2^0. It is possible to adapt the method of our proof to the case $U(\tau_1^0) = U(\tau_2^0)$. However, we focus on the most interesting case by assuming Assumption A4.

Lemma 3. *Under Assumption A1-A4, there exists a positive constant C which does not depend on n such that*

$$n(\bar{U}_n(k/n) - \bar{U}_n(t_1^0/n)) \geq C|k - t_1^0| \; for \; all \; large \; n,$$

where $k \in [1, n-1]$ denotes a location.

Lemma 3 is essentially a restatement of the Lemma 3 of Bai (1997), with minor differences. Lemma 3 confirms that t_1^0 is the minimizer of $\bar{U}_n(k/n)$. Furthermore, it quantifies the lower bound on the difference $\bar{U}_n(k/n) - \bar{U}_n(t_1^0/n)$, which is essential in proving the convergence rate of $\hat{t}^w$ to t_1^0. Recall that $\hat{t}^w$ is the minimizer of $L_n^w(.)$ defined in (B.2), and it is the first detected change-point in our sequential implementation of the BS algorithm.

Lemma 4. *Under Assumptions A1-A4, $\hat{t}^w - t_1^0 = o_{\mathbb{P}_w}(n^{1/2}\log(n))$.*

The convergence rate of $\hat{t}^w$ to t_1^0 is $o_{\mathbb{P}_w}(n^{1/2}\log(n))$ as shown in Lemma 4. This rate is too slow for doing inference on the mean of X_t. The convergence rate can be improved by carefully examining the behavior of $L_n^w(k)$ for those k is small neighborhood of t_1^0. For a fixed $\varepsilon > 0$, denote $D_{n\varepsilon} = \{k : n\eta \leq k \leq n\tau_2^0(1-\eta),\ |k - t_1^0| > \varepsilon\log(n)\}$, where $\eta > 0$ is chosen such that $t_1^0 \in (n\eta, n\tau_2^0(1-\eta))$. The following lemma is helpful in improving the convergence rate of $\hat{t}^w$.

Lemma 5. *Under Assumptions A1-A4, for every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_w\left(\inf_{k\in D_{n\varepsilon}} L_n^w(k) \leq L_n^w(t_1^0)\right) = 0 \quad a.s..$$

To prove the lemmas, we exploit the Háyek-Rényi inequality (Hájek & Rényi, 1955) for independent random variables. Note that the Háyek-Rényi inequality has been generazlied to linear processes (Bai, 1994), martingale difference sequences, ρ-mixing sequences (Fazekas & Klesov, 2001), and negatively associated sequences (Hu & Hu, 2006). The Háyek-Rényi inequality can be used to derive uniform convergence rate of partial sums.

Lemma 6. *(Hájek & Rényi, 1955) Let $X_1, \ldots, X_k, \ldots$ be a sequence of independent random variables with zero mean ($\mathbb{E}(X_k) = 0$) and finite variance ($\mathbb{E}(X_k^2) < \infty$). Denote the partial sum of X_k as $S_k = X_1 + \cdots + X_k$. Let $c_k, k = 1, 2, \ldots$ be a non-increasing sequence of positive numbers. For any $\alpha > 0$ and positive numbers $n < m$,*

$$\mathbb{P}\left(\max_{n \le k \le m} c_k |S_k| \ge \alpha\right) \le \frac{1}{\alpha^2}\left[c_n^2 \sum_{k=1}^{n} \mathbb{E}(X_k^2) + \sum_{k=n+1}^{m} c_k^2 \mathbb{E}(X_k^2)\right].$$

The following conditional Háyek-Rényi inequality, derived by Rao (2009), will be useful in deriving convergence rates of bootstrap estimators.

Lemma 7. *(Rao, 2009, Theorem 5) Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. Conditional on $\mathcal{G}$, let $X_1, \ldots, X_k, \ldots$ be a sequence of conditional independent random variables with zero conditional mean ($\mathbb{E}(X_k|\mathcal{G}) = 0$) and finite conditional variance ($\mathbb{E}(X_k^2|\mathcal{G}) < \infty$). Denote the partial sum of X_k as $S_k = X_1 + \cdots + X_k$. Let $c_k, k = 1, 2, \ldots$ be a non-increasing sequence of positive $\mathcal{G}$-measurable random variables. For any $\mathcal{G}$-measurable random variable $\alpha > 0, a.s.$ and positive numbers $n < m$,*

$$\mathbb{P}\left(\max_{n \le k \le m} c_k |S_k| \ge \alpha \,\middle|\, \mathcal{G}\right) \le \frac{1}{\alpha^2}\left[c_n^2 \sum_{k=1}^{n} \mathbb{E}(X_k^2|\mathcal{G}) + \sum_{k=n+1}^{m} c_k^2 \mathbb{E}(X_k^2|\mathcal{G})\right] \; a.s..$$

By Lemma 6 and 7, we can establish uniform convergence of partial sums. Lemma 8 collects some uniform convergence rates which are handy in proving the lemmas.

Lemma 8. *Let $0 < \phi(n) \le n$ be a non-decreasing and unbounded function. Under*

Assumptions A1-A3, we have,

$$(i) \quad \max_{1\le k\le t_1^0} \Big| \sum_{t=k+1}^{n} (w_t - 1) \Big| = O_{\mathbb{P}}(\sqrt{n}) = O_{\mathbb{P}_w}(\sqrt{n}),$$

$$\text{and} \max_{1\le k\le t_1^0} \Big| \sum_{t=k+1}^{t_1^0} (w_t - 1) \Big| = O_{\mathbb{P}_w}(\sqrt{n});$$

$$(ii) \quad \max_{1\le k\le t_1^0} \Big| \sum_{t=k+1}^{n} (w_t - 1)\epsilon_t \Big| = O_{\mathbb{P}_w}(\sqrt{n});$$

$$(iii) \quad \max_{1\le k\le \phi(n)} \Big| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} (w_t - 1)\epsilon_t \Big| = o_{\mathbb{P}_w}(\log \phi(n));$$

$$(iv) \quad \max_{1\le k\le \phi(n)} \Big| \sum_{t=1}^{k} \epsilon_t \Big| = o_{\mathbb{P}_w}(\sqrt{\phi(n)} \log \phi(n));$$

$$(v) \quad \max_{1\le k\le \phi(n)} \Big| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} \epsilon_t \Big| = o_{\mathbb{P}_w}(\log \phi(n)).$$

Proof. (i) By the Háyek-Rényi inequality (Lemma 6), we have for $M > 0$,

$$\mathbb{P}\left(\max_{1\le k\le t_1^0} \Big| \sum_{k+1}^{n} (w_t - 1) \Big| > M\sqrt{n} \right) \le \frac{1}{M^2 n} \big((n - t_1^0) + (t_1^0 - 1) \big) \le \frac{1}{M^2}.$$

Therefore, $\max_{1\le k\le t_1^0} \Big| \sum_{k+1}^{n} (w_t - 1) \Big| = O_{\mathbb{P}}(\sqrt{n})$. Because the bootstrap weights $w_t, t = 1, \ldots, n$ and the data $X_t, t = 1, \ldots, n$ are independent by Assumption 3, $\max_{1\le k\le t_1^0} \Big| \sum_{k+1}^{n} (w_t - 1) \Big| = O_{\mathbb{P}_w}(\sqrt{n})$ as well. The second part of (i) is proved in the same way, for any $M > 0$,

$$\mathbb{P}\left(\max_{1\le k\le t_1^0} \Big| \sum_{1}^{t_1^0} (w_t - 1) \Big| > M\sqrt{n} \right) \le \frac{t_1^0}{M^2 n} \le \frac{1}{M^2}.$$

(ii) By the conditional Háyek-Rényi inequality (Lemma 7), we have for $M > 0$,

$$
\begin{aligned}
&\mathbb{P}_w\left(\max_{1\le k\le t_1^0} \mid \sum_{k+1}^{n}(w_t-1)\epsilon_t \mid > M\sqrt{n}\right)\\
&\le \frac{1}{nM^2}\left(\sum_{t=t_1^0+1}^{n}\mathbb{E}_w((w_t-1)\epsilon_t)^2+\sum_{t=1}^{t_1^0}\mathbb{E}_w((w_t-1)\epsilon_t)^2\right) \quad a.s.\\
&= \frac{1}{M^2}\left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\right) \quad a.s.\\
&\le \frac{1}{M^2}(\mathbb{E}(\epsilon_1^2)+1) \quad a.s.
\end{aligned}
$$

where the last step follows from the strong law of large numbers. By the definition of the notation $O_{\mathbb{P}_w}(.)$, (ii) is proved.

(iii) By the conditional Háyek-Rényi inequality (Lemma 7), we have for $\varepsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}_w\left(\max_{1\le k\le \phi(n)} \mid \frac{1}{\sqrt{k}}\sum_{1}^{k}(w_t-1)\epsilon_t \mid > \varepsilon\log\phi(n)\right)\\
&\le \frac{1}{\varepsilon^2(\log\phi(n))^2}\left(\sum_{t=1}^{\phi(n)}\frac{\mathbb{E}_w((w_t-1)\epsilon_t)^2}{t}\right) \quad a.s.\\
&= \frac{1}{\varepsilon^2}\left(\frac{1}{(\log\phi(n))^2}\sum_{t=1}^{\phi(n)}\frac{\epsilon_t^2-\sigma^2}{t}+\frac{1}{(\log\phi(n))^2}\sum_{t=1}^{\phi(n)}\frac{\sigma^2}{t}\right) \quad a.s..
\end{aligned}
$$

The series $\sum_{t=1}^{\infty}(\epsilon_t^2-\sigma^2)/(t(\log(t))^2)$ converges almost surely by invoking the Kolmogorov's three series theorem. In view of the Kronecker lemma, the first term in the bracket converges almost surely to 0. The second term in the bracket converges to 0 by noting that $\sum_{t=1}^{\phi(n)} t^{-1} \le \log(\phi(n)) + 1$.

(iv) By a formula discovered by Chung (1948, page 206), we have for any $\varepsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\limsup_n\left\{\max_{1\le k\le\phi(n)} \mid \sum_{t=1}^{k}\epsilon_t \mid > (\sqrt{\phi(n)}\log\phi(n))\varepsilon\right\}\right)\\
&= \mathbb{P}\left(\limsup_n\left\{\mid \sum_{t=1}^{\phi(n)}\epsilon_t \mid > \sqrt{\phi(n)}\log\phi(n)\varepsilon\right\}\right).
\end{aligned}
$$

That is, we can translate the strong convergence result of sum of random variables to the maximum of sums. Therefore, by applying the law of iterated logarithms, we have $\max\limits_{1\le k\le \phi(n)} |\sum_{t=1}^{k}\epsilon_t|/(\sqrt{\phi(n)}\log\phi(n))$ converges to 0 almost surely. Lastly, we prove that if Z_n is $\sigma(X_1,\ldots,X_n)$ measurable, then $Z_n = o_{\mathbb{P}_w}(1)$ is equivalent to Z_n converges to zero almost surely. This follows from the following chain of arguments,

$$
\begin{aligned}
&\forall\varepsilon>0,\ \lim_{n\to\infty}\mathbb{P}_w(|Z_n|>\varepsilon)=0,\ a.s.\\
\Leftrightarrow\ &\forall\varepsilon>0,\ \lim_{n\to\infty}1(|Z_n|>\varepsilon)=0,\ a.s.\\
\Leftrightarrow\ &\forall k=1,2,\ldots,\ \lim_{n\to\infty}1(|Z_n|>1/k)=0,\ a.s.\\
\Leftrightarrow\ &\lim_{n\to\infty}|Z_n|=0,\ a.s.,
\end{aligned}
$$

where $1(.)$ is the indicator function.

(v) The proof of (v) is the same as that of (iv). Using Chung (1948)'s formula, we have for any $\varepsilon>0$,

$$
\begin{aligned}
&\mathbb{P}\left(\limsup_n\left\{\max_{1\le k\le\phi(n)}\Big|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}\epsilon_t\Big|>\log\phi(n)\varepsilon\right\}\right)\\
&=\mathbb{P}\left(\limsup_n\left\{\Big|\frac{1}{\sqrt{\phi(n)}}\sum_{t=1}^{\phi(n)}\epsilon_t\Big|>\log\phi(n)\varepsilon\right\}\right).
\end{aligned}
$$

The lemma is proved by invoking the law of iterated logarithm.

□

The proofs of the Lemma 1-5 and theorems are presented below.

Proof of Lemma 1 and Lemma 2. We first examine $L_n^w(k)$ for $k\in[1,n]$. For $k\le t_1^0$,

$$
\begin{aligned}
\bar{X}_k^w &= \frac{\sum_{t=1}^{k}w_tX_t}{\sum_{t=1}^{k}w_t}=\beta_1^0+\frac{\sum_{t=1}^{k}w_t\epsilon_t}{\sum_{t=1}^{k}w_t},\\
\bar{X}_{-k}^w &= \frac{\sum_{t=k+1}^{t_1^0}w_t}{\sum_{t=k+1}^{n}w_t}\beta_1^0+\frac{\sum_{t=t_1^0+1}^{t_2^0}w_t}{\sum_{t=k+1}^{n}w_t}\beta_2^0+\frac{\sum_{t=t_2^0+1}^{n}w_t}{\sum_{t=k+1}^{n}w_t}\beta_3^0+\frac{\sum_{t=k+1}^{n}w_t\epsilon_t}{\sum_{t=k+1}^{n}w_t}.
\end{aligned}
$$

To simplify the notations, denote

$$
\bar{\epsilon}_k^w=\frac{\sum_{t=1}^{k}w_t\epsilon_t}{\sum_{t=1}^{k}w_t},\quad \bar{\epsilon}_{-k}^w=\frac{\sum_{t=k+1}^{n}w_t\epsilon_t}{\sum_{t=k+1}^{n}w_t}.
$$

and

$$a_k^w = \frac{1}{\sum_{k+1}^n w_t} \left\{ \sum_{t_1^0+1}^{n} w_t(\beta_1^0 - \beta_2^0) + \sum_{t_2^0+1}^{n} w_t(\beta_2^0 - \beta_3^0) \right\},$$
$$b_k^w = \frac{1}{\sum_{k+1}^n w_t} \left\{ \sum_{k+1}^{t_1^0} w_t(\beta_2^0 - \beta_1^0) + \sum_{t_2^0+1}^{n} w_t(\beta_2^0 - \beta_3^0) \right\},$$
$$c_k^w = \frac{1}{\sum_{k+1}^n w_t} \left\{ \sum_{k+1}^{t_1^0} w_t(\beta_2^0 - \beta_1^0) + \sum_{k+1}^{t_2^0} w_t(\beta_3^0 - \beta_2^0) \right\}.$$

Using these notations, we have

$$\sum_{t=1}^{k} w_t(X_t - \bar{X}_k^w)^2 = \sum_{t=1}^{k} w_t(\epsilon_t - \bar{\epsilon}_k^w)^2,$$

and

$$\begin{aligned}
&\sum_{t=k+1}^{n} w_t(X_t - \bar{X}_{-k}^w)^2 \\
&\quad= \sum_{t=k+1}^{t_1^0} w_t(X_t - \bar{X}_{-k}^w)^2 + \sum_{t=t_1^0+1}^{t_2^0} w_t(X_t - \bar{X}_{-k}^w)^2 + \sum_{t=t_2^0+1}^{n} w_t(X_t - \bar{X}_{-k}^w)^2 \\
&\quad= \sum_{t=k+1}^{t_1^0} w_t \left(a_k^w + \epsilon_t - \bar{\epsilon}_{-k}^w\right)^2 + \sum_{t=t_1^0+1}^{t_2^0} w_t \left(b_k^w + \epsilon_t - \bar{\epsilon}_{-k}^w\right)^2 + \sum_{t=t_2^0+1}^{n} w_t \left(c_k^w + \epsilon_t - \bar{\epsilon}_{-k}^w\right)^2 \\
&\quad= (a_k^w)^2 \sum_{k+1}^{t_1^0} w_t + (b_k^w)^2 \sum_{t_1^0+1}^{t_2^0} w_t + (c_k^w)^2 \sum_{t_2^0+1}^{n} w_t + 2a_k^w \sum_{k+1}^{t_1^0} w_t(\epsilon_t - \bar{\epsilon}_{-k}^w) \\
&\qquad+ 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t(\epsilon_t - \bar{\epsilon}_{-k}^w) + 2c_k^w \sum_{t_2^0+1}^{n} w_t(\epsilon_t - \bar{\epsilon}_{-k}^w) + \sum_{k+1}^{n} w_t(\epsilon_t - \bar{\epsilon}_{-k}^w)^2.
\end{aligned}$$

Thus, we obtain for $k \le t_1^0$,

$$
\begin{aligned}
U_n^w(k/n) = L_n^w(k) &= \sum_{t=1}^{k} w_t (X_t - \bar{X}_k^w)^2 + \sum_{t=k+1}^{n} w_t (X_t - \bar{X}_{-k}^w)^2 \\
&= \frac{1}{n}\left[(a_k^w)^2 \sum_{k+1}^{t_1^0} w_t + (b_k^w)^2 \sum_{t_1^0+1}^{t_2^0} w_t + (c_k^w)^2 \sum_{t_2^0+1}^{n} w_t\right] \qquad \text{(B.4)} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} w_t \epsilon_t^2 + R_n^w(k),
\end{aligned}
$$

where

$$
\begin{aligned}
R_n^w(k) &= \frac{1}{n}\left(2a_k^w \sum_{k+1}^{t_1^0} w_t\epsilon_t + 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t + 2c_k^w \sum_{t_2^0+1}^{n} w_t\epsilon_t\right) \\
&\quad - \frac{1}{n}\left(2a_k^w \sum_{k+1}^{t_1^0} w_t + 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t + 2c_k^w \sum_{t_2^0+1}^{n} w_t\right)\bar{\epsilon}_{-k}^w \\
&\quad - \frac{1}{n}\sum_{t=1}^{k} w_t (\bar{\epsilon}_k^w)^2 - \frac{1}{n}\sum_{t=k+1}^{n} w_t (\bar{\epsilon}_{-k}^w)^2 \\
&:= I(k) + II(k) + III(k) + IV(k).
\end{aligned}
$$

In the following, we derive the convergence rate of $R_n^w(k)$. Note that the bootstrap weights $w_t, t = 1, \ldots, n$ are positive by Assumption A3, therefore $\max\{|a_k^w|, |b_k^w|, |c_k^w|\} \le$

$2(|\beta_1^0| + |\beta_2^0| + |\beta_3^0|)$. Thus,

$$
\begin{aligned}
I(k) &= \frac{1}{n}\left(2a_k^w \sum_{k+1}^{t_1^0} w_t\epsilon_t + 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t + 2c_k^w \sum_{t_2^0+1}^{n} w_t\epsilon_t\right) \\
&= \frac{1}{n}\left(2a_k^w \sum_{k+1}^{n} w_t\epsilon_t + 2(b_k^w - a_k^w) \sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t + 2(c_k^w - a_k^w) \sum_{t_2^0+1}^{n} w_t\epsilon_t\right) \\
&= \frac{1}{n}\left(2a_k^w \sum_{k+1}^{n} (w_t - 1)\epsilon_t + 2(b_k^w - a_k^w) \sum_{t_1^0+1}^{t_2^0} (w_t - 1)\epsilon_t + 2(c_k^w - a_k^w) \sum_{t_2^0+1}^{n} (w_t - 1)\epsilon_t\right) \\
&\quad + \frac{1}{n}\left(2a_k^w \sum_{k+1}^{n} \epsilon_t + 2(b_k^w - a_k^w) \sum_{t_1^0+1}^{t_2^0} \epsilon_t + 2(c_k^w - a_k^w) \sum_{t_2^0+1}^{n} \epsilon_t\right) \\
&:= I_1(k) + I_2(k).
\end{aligned}
$$

Now, we apply Lemma 8 (ii) to the first term $I_1(k)$, by the boundedness of a_k^w, b_k^w, and c_k^w, we conclude that $\sup_{1\le k\le t_1^0} I_1(k) = O_{\mathbb{P}_w}(n^{-1/2})$. Note that $\max_{1\le k\le t_1^0} |\sum_{k+1}^n \epsilon_t| \le \max_{1\le k\le n} |\sum_{t=k}^n \epsilon_t| = o_{\mathbb{P}_w}(\sqrt{n}\log n)$ by Lemma 8 (iv). Thus, we have $\sup_{1\le k\le t_1^0} I_2(k) = o_{\mathbb{P}_w}(n^{-1/2}\log n)$. By combining the convergence rates of $I_1(k)$ and $I_2(k)$, we obtain the uniform convergence rate of $I(k)$ as $\sup_{1\le k\le t_1^0} I(k) = o_{\mathbb{P}_w}(n^{-1/2}\log n)$.

Similarly,

$$
\begin{aligned}
II(k) &= \frac{1}{n}\left(2a_k^w \sum_{k+1}^{t_1^0} w_t + 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t + 2c_k^w \sum_{t_2^0+1}^{n} w_t\right) \bar{\epsilon}_{-k}^w \\
&= \frac{1}{n} \frac{\left(2a_k^w \sum_{k+1}^{t_1^0} w_t + 2b_k^w \sum_{t_1^0+1}^{t_2^0} w_t + 2c_k^w \sum_{t_2^0+1}^{n} w_t\right)}{\sum_{k+1}^n w_t} \left(\sum_{t=k+1}^{n} w_t\epsilon_t\right) \\
&\le \frac{2}{n}\max\{|a_k^w|, |b_k^w|, |c_k^w|\} \left(\Big|\sum_{t=k+1}^{n} (w_t - 1)\epsilon_t\Big| + \Big|\sum_{t=k+1}^{n} \epsilon_t\Big|\right) \\
&\le \frac{4}{n}(|\beta_1^0| + |\beta_2^0| + |\beta_3^0|) \left(\Big|\sum_{t=k+1}^{n} (w_t - 1)\epsilon_t\Big| + \Big|\sum_{t=k+1}^{n} \epsilon_t\Big|\right).
\end{aligned}
$$

By applying Lemma 8 (ii) and (iv), we have $\sup_{1\le k\le t_1^0} II(k) = o_{\mathbb{P}_w}(n^{-1/2}\log n)$.

The uniform convergence rate of $III(k)$ can be derived as follows,

$$\begin{aligned}
\sup_{1\le k\le t_1^0} III(k) &= \frac{1}{n}\sup_{1\le k\le t_1^0}\sum_{t=1}^{k} w_t(\bar{\epsilon}_k^w)^2 = \frac{1}{n}\sup_{1\le k\le t_1^0}\frac{\left(\sum_{t=1}^{k} w_t\epsilon_t\right)^2}{\sum_{t=1}^{k} w_t} \\
&= \frac{1}{n}\sup_{1\le k\le t_1^0}\frac{\left(\frac{1}{\sqrt{k}}\sum_{t=1}^{k}(w_t-1)\epsilon_t + \frac{1}{\sqrt{k}}\sum_{t=1}^{k}\epsilon_t\right)^2}{\frac{1}{k}\sum_{t=1}^{k} w_t} \\
&\le \frac{1}{n}\sup_{1\le k\le \log n}\frac{\left(\sum_{t=1}^{k}(w_t-1)\epsilon_t + \sum_{t=1}^{k}\epsilon_t\right)^2}{\sum_{t=1}^{k} w_t} \\
&\quad + \frac{1}{n}\sup_{\log n\le k\le t_1^0}\frac{\left(\frac{1}{\sqrt{k}}\sum_{t=1}^{k}(w_t-1)\epsilon_t + \frac{1}{\sqrt{k}}\sum_{t=1}^{k}\epsilon_t\right)^2}{\frac{1}{k}\sum_{t=1}^{k}(w_t-1)+1} \\
&:= III_1 + III_2.
\end{aligned}$$

By the conditional Háyek-Rényi inequality (Lemma 7), we have for $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}_w\left(\max_{1\le k\le \log(n)} \Big|\sum_{t=1}^{k}(w_t-1)\epsilon_t\Big| > \varepsilon\log(n)\right) \\
&\le \frac{1}{\varepsilon^2\log(n)}\left(\frac{1}{\log(n)}\sum_{t=1}^{\log(n)}\epsilon_t^2\right) \to 0 \quad a.s.,
\end{aligned} \tag{B.5}$$

where the last step follows from the strong law of large numbers. Take $\phi(n) = \log(n)$ in Lemma 8 (iv), we have $\sup_{1\le k\le\log n}|\sum_{t=1}^{k}\epsilon_t| = o_{\mathbb{P}_w}(\log n)$. Together with (B.5), we have $III_1 = o_{\mathbb{P}_w}(\log(n)^2/n)$.

Next, we derive the convergence rate of III_2. By Lemma 6 and the Assumption A3 that the bootstrap weights and the observations are independent, we have for any $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{P}_w\left(\max_{\log(n)\le k\le t_1^0}\Big|\frac{1}{k}\sum_{t=1}^{k}(w_t-1)\Big| > \varepsilon\right) \\
&\le \frac{1}{\varepsilon^2}\left(\frac{1}{\log(n)}\sum_{t=1}^{\log(n)}\frac{1}{k^2} + \sum_{t=\log(n)+1}^{t_1^0}\frac{1}{k^2}\right) \le \frac{1}{\varepsilon^2}\frac{3}{\log(n)}.
\end{aligned} \tag{B.6}$$

Therefore $\max_{\log(n)\le k\le t_1^0}\Big|\frac{1}{k}\sum_{t=1}^{k}(w_t-1)\Big| = o_{\mathbb{P}_w}(1)$. Take $\phi(n) = t_1^0 = n\tau_1^0$ in Lemma

8 (iii) and (v), we obtain that

$$\max_{\log(n) \le k \le t_1^0} \left| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} (w_t - 1)\epsilon_t \right| = o_{\mathbb{P}_w}(\log(n)), \tag{B.7}$$

and

$$\max_{\log(n) \le k \le t_1^0} \left| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} \epsilon_t \right| = o_{\mathbb{P}_w}(\log(n)). \tag{B.8}$$

By combining (B.6), (B.7), and (B.8), we obtain $III_2 = o_{\mathbb{P}_w}(\log(n)^2/n)$, and therefore $\sup_{1 \le k \le t_1^0} III(k) = o_{\mathbb{P}_w}(\log(n)^2/n)$.

Using the same techniques for deriving the convergence rate of III_2, we can prove that $\sup_{1 \le k \le t_1^0} IV(k) = o_{\mathbb{P}_w}(\log(n)^2/n)$. Combining the convergence rates of $I(k)$, $II(k)$, $III(k)$, and $IV(k)$, we obtain $\sup_{1 \le k \le t_1^0} R_n^w(k) = o_{\mathbb{P}_w}(\log(n)/\sqrt{n})$.

Recall the decomposition of $U_n^w(k/n)$ in (B.4) for $k \in [1, t_1^0]$ is

$$\begin{aligned} U_n^w(k/n) &= \frac{1}{n} \left[(a_k^w)^2 \sum_{k+1}^{t_1^0} w_t + (b_k^w)^2 \sum_{t_1^0+1}^{t_2^0} w_t + (c_k^w)^2 \sum_{t_2^0+1}^{n} w_t \right] + \frac{1}{n} \sum_{t=1}^{n} w_t \epsilon_t^2 + R_n^w(k), \\ &= \left(\frac{t_1^0 - k}{n} + \frac{\sum_{k+1}^{t_1^0} (w_t - 1)}{n} \right) (a_k^w)^2 + \left(\frac{t_2^0 - t_1^0}{n} + \frac{\sum_{t_1^0+1}^{t_2^0} (w_t - 1)}{n} \right) (b_k^w)^2 \\ &\quad + \left(\frac{n - t_2^0}{n} + \frac{\sum_{t_2^0+1}^{n} (w_t - 1)}{n} \right) (c_k^w)^2 + \frac{1}{n} \sum_{t=1}^{n} w_t \epsilon_t^2 + R_n^w(k). \end{aligned}$$

For $k \in [1, t_1^0]$, denote

$$\bar{U}_n(k/n) = \frac{t_1^0 - k}{n} a_k^2 + \frac{t_2^0 - t_1^0}{n} b_k^2 + \frac{n - t_2^0}{n} c_k^2 + \sigma^2, \tag{B.9}$$

where $a_k = ((n - t_1^0)(\beta_1^0 - \beta_2^0) + (n - t_2^0)(\beta_2^0 - \beta_3^0))/(n - k)$, $b_k = ((t_1^0 - k)(\beta_2^0 - \beta_1^0) + (n - t_2^0)(\beta_2^0 - \beta_3^0))/(n - k)$, and $c_k = ((t_1^0 - k)(\beta_2^0 - \beta_1^0) + (t_2^0 - k)(\beta_3^0 - \beta_2^0))/(n - k)$.

We calculate the difference between $U_n^w(k/n)$ and $\bar{U}_n(k/n)$ as

$$\begin{aligned} U_n^w(k/n) - \bar{U}_n(k/n) = & \frac{t_1^0 - k}{n}\left((a_k^w)^2 - a_k^2\right) + \frac{\sum_{k+1}^{t_1^0}(w_t - 1)}{n}(a_k^w)^2 \\ & + \frac{t_2^0 - t_1^0}{n}\left((b_k^w)^2 - b_k^2\right) + \frac{\sum_{t_1^0+1}^{t_2^0}(w_t - 1)}{n}(b_k^w)^2 \\ & + \frac{n - t_2^0}{n}\left((c_k^w)^2 - c_k^2\right) + \frac{\sum_{t_2^0+1}^{n}(w_t - 1)}{n}(c_k^w)^2 \\ & + \frac{1}{n}\sum_{t=1}^{n}(w_t - 1)\epsilon_t^2 + \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t^2 - \sigma^2) + R_n^w(k). \end{aligned} \tag{B.10}$$

To derive the uniform convergence rate of $U_n^w(k/n) - \bar{U}_n(k/n)$, we first investigate the convergence rate of a_k^w to a_k, b_k^w to b_k, and c_k^w to c_k. For a_k^w, simple algebras lead to

$$\begin{aligned} a_k^w - a_k = & \left(\frac{\sum_{t_1^0+1}^{n} w_t}{\sum_{k+1}^{n} w_t} - \frac{n - t_1^0}{n-k}\right)(\beta_1^0 - \beta_2^0) + \left(\frac{\sum_{t_2^0+1}^{n} w_t}{\sum_{k+1}^{n} w_t} - \frac{n - t_2^0}{n-k}\right)(\beta_2^0 - \beta_3^0) \\ = & \frac{\frac{1}{n-t_1^0}\sum_{t_1^0+1}^{n}(w_t - 1) - \frac{1}{n-k}\sum_{k+1}^{n}(w_t - 1)}{\left(\frac{1}{n-k}\sum_{k+1}^{n}(w_t - 1) + 1\right)\left(\frac{n-k}{n-t_1^0}\right)}(\beta_1^0 - \beta_2^0) \\ & + \frac{\frac{1}{n-t_2^0}\sum_{t_2^0+1}^{n}(w_t - 1) - \frac{1}{n-k}\sum_{k+1}^{n}(w_t - 1)}{\left(\frac{1}{n-k}\sum_{k+1}^{n}(w_t - 1) + 1\right)\left(\frac{n-k}{n-t_2^0}\right)}(\beta_2^0 - \beta_3^0). \end{aligned}$$

By Lemma 6, we have for some $M > 0$

$$\mathbb{P}_w\left(\max_{1\le k\le t_1^0}\Big|\ \frac{1}{n-k}\sum_{k+1}^{n}(w_t - 1)\ \Big| > Mn^{-1/2}\right) \le \frac{1}{M^2}\frac{nC}{n - t_1^0}, \tag{B.11}$$

for some constant $C > 0$. Therefore, $\max_{1\le k\le t_1^0}\big|\ (n-k)^{-1}\sum_{k+1}^{n}(w_t - 1)\ \big| = O_{\mathbb{P}_w}(n^{-1/2})$. We thus proved that $\max_{1\le k\le t_1^0}|a_k^w - a_k| = O_{\mathbb{P}_w}(n^{-1/2})$. Similarly, for b_k^w and c_k^w we have $\max_{1\le k\le t_1^0}|b_k^w - b_k| = O_{\mathbb{P}_w}(n^{-1/2})$ and $\max_{1\le k\le t_1^0}|c_k^w - c_k| = O_{\mathbb{P}_w}(n^{-1/2})$.

By the convergence rate of a_k^w to a_k, b_k^w to b_k, c_k^w to c_k, and Lemma 8 (i), we have proved that the first six terms of (B.10) have uniform convergence rate of $O_{\mathbb{P}_w}(n^{-1/2})$ for

$k \in [1, t_1^0]$. The seventh term of (B.10) can be bounded as

$$\begin{aligned}
&\mathbb{P}_w\left(\Big| \frac{1}{n}\sum_{t=1}^{n}(w_t - 1)\epsilon_t^2 \Big| > n^{-1/2}\log(n)\varepsilon \right) \\
&\leq \frac{1}{n\log(n)^2\varepsilon^2}\sum_{t=1}^{n}\epsilon_t^4 \to 0 \quad a.s.,
\end{aligned}$$

for any $\varepsilon > 0$, where the convergence follows from the strong law of large numbers. The convergence rate of the eighth term is $o_{\mathbb{P}_w}(n^{-1/2}\log(n))$ by the law of iterated logarithm. Together with the already proved rate $\sup_{1\leq k\leq t_1^0} R_n^w(k) = o_{\mathbb{P}_w}(\log(n)/\sqrt{n})$, we obtain the convergence rate

$$\sup_{1\leq k\leq t_1^0} |U_n^w(k/n) - \bar{U}_n(k/n)| = o_{\mathbb{P}_w}(\log(n)/\sqrt{n}).$$

With the same technique the results can be proved for $k \in [t_1^0{+}1, t_2^0]$ and $k \in [t_2^0{+}1, n]$. These results imply Lemma 2.

For $k = \lfloor n\tau \rfloor$ and $\tau \leq \tau_1^0$, denote

$$\begin{aligned}
a(\tau) &= \frac{1-\tau_1^0}{1-\tau}(\beta_1^0 - \beta_2^0) + \frac{1-\tau_2^0}{1-\tau}(\beta_2^0 - \beta_3^0) \\
b(\tau) &= \frac{\tau_1^0 - \tau}{1-\tau}(\beta_2^0 - \beta_1^0) + \frac{1-\tau_2^0}{1-\tau}(\beta_2^0 - \beta_3^0) \\
c(\tau) &= \frac{\tau_1^0 - \tau}{1-\tau}(\beta_2^0 - \beta_1^0) + \frac{\tau_2^0 - \tau}{1-\tau}(\beta_3^0 - \beta_2^0),
\end{aligned}$$

and let

$$U(\tau) = (\tau_1^0 - \tau)a(\tau)^2 + (\tau_2^0 - \tau_1^0)b(\tau)^2 + (1-\tau_2^0)c(\tau)^2. \tag{B.12}$$

Since $\lim_{n\to\infty}\bar{U}_n(k/n) = U(\tau)$ uniformly for $\tau \in [0, \tau_1^0]$, Lemma 1 is proved.

Proof of Lemma 4.

$$\begin{aligned}
U_n^w(\hat{t}^w/n) - U_n^w(t_1^0/n) &= (U_n^w(\hat{t}^w/n) - \bar{U}_n(\hat{t}^w/n)) - (U_n^w(t_1^0/n) - \bar{U}_n(t_1^0/n)) \\
&\quad + \bar{U}_n(\hat{t}^w/n) - \bar{U}_n(t_1^0/n) \\
&\geq -2\sup_{1\leq k\leq n}|U_n^w(k/n) - \bar{U}_n(k/n)| + C|\hat{t}^w - t_1^0|/n,
\end{aligned}$$

where the last inequality is true by Lemma 3. Since $U_n^w(\hat{t}^w/n) - U_n^w(t_1^0/n) \leq 0$, we have

$$|\hat{t}^w - t_1^0| \leq C^{-1}2n\sup_{1\leq k\leq n}|U_n^w(k/n) - \bar{U}_n(k/n)|.$$

Thus, Lemma 4 is proved by the result of Lemma 2.

Proof of Lemma 5. In the proof, we consider the case $k < t_1^0$. The case $k > t_1^0$ is similar. With a little bit abuse of notation, we denote $D_{n\varepsilon} = \{k : n\eta \le k \le n\tau_2^0(1-\eta), t_1^0 - k > \varepsilon \log(n)\}$. We decompose $L_n^w(k) - L_n^w(t_1^0)$ as follows,

$$\begin{aligned} L_n^w(k) - L_n^w(t_1^0) = &\sum_{k+1}^{t_1^0} w_t (a_k^w)^2 + \sum_{t_1^0+1}^{t_2^0} w_t((b_k^w)^2 - (b_{t_1^0}^w)^2) \\ &+ \sum_{t_2^0+1}^{n} w_t((c_k^w)^2 - (c_{t_1^0}^w)^2) + n(R_n^w(k) - R_n^w(t_1^0)). \end{aligned}$$

Denote $A(w, \boldsymbol{\beta}^0) = \sum_{t_1^0+1}^n w_t(\beta_1^0 - \beta_2^0) + \sum_{t_2^0+1}^n w_t(\beta_2^0 - \beta_3^0)$. We rewrite $a_k^w,\ b_k^w,\ c_k^w$ as

$$\begin{aligned} a_k^w &= \frac{A(w, \boldsymbol{\beta}^0)}{\sum_{k+1}^n w_t}, \\ b_k^w &= \frac{A(w, \boldsymbol{\beta}^0)}{\sum_{k+1}^n w_t} + (\beta_2^0 - \beta_1^0), \\ c_k^w &= \frac{A(w, \boldsymbol{\beta}^0)}{\sum_{k+1}^n w_t} + (\beta_3^0 - \beta_1^0). \end{aligned}$$

After some algebras, we obtain

$$
\begin{aligned}
&\sum_{k+1}^{t_1^0} w_t(a_k^w)^2 + \sum_{t_1^0+1}^{t_2^0} w_t((b_k^w)^2 - (b_{t_1^0}^w)^2) + \sum_{t_2^0+1}^{n} w_t((c_k^w)^2 - (c_{t_1^0}^w)^2) \\
&= \frac{\sum_{k+1}^{t_1^0} w_t}{(\sum_{k+1}^{n} w_t)^2} A(w,\boldsymbol{\beta}^0)^2 + \sum_{t_1^0+1}^{t_2^0} w_t \left\{ \frac{A(w,\boldsymbol{\beta}^0)^2}{(\sum_{k+1}^{n} w_t)^2} + \frac{2(\beta_2^0 - \beta_1^0)A(w,\boldsymbol{\beta}^0)}{\sum_{k+1}^{n} w_t} \right\} \\
&\quad - \sum_{t_1^0+1}^{t_2^0} w_t \left\{ \frac{A(w,\boldsymbol{\beta}^0)^2}{(\sum_{t_1^0+1}^{n} w_t)^2} + \frac{2(\beta_2^0 - \beta_1^0)A(w,\boldsymbol{\beta}^0)}{\sum_{t_1^0+1}^{n} w_t} \right\} \\
&\quad + \sum_{t_2^0+1}^{n} w_t \left\{ \frac{A(w,\boldsymbol{\beta}^0)^2}{(\sum_{k+1}^{n} w_t)^2} + \frac{2(\beta_3^0 - \beta_1^0)A(w,\boldsymbol{\beta}^0)}{\sum_{k+1}^{n} w_t} \right\} \\
&\quad - \sum_{t_2^0+1}^{n} w_t \left\{ \frac{A(w,\boldsymbol{\beta}^0)^2}{(\sum_{t_1^0+1}^{n} w_t)^2} + \frac{2(\beta_3^0 - \beta_1^0)A(w,\boldsymbol{\beta}^0)}{\sum_{t_1^0+1}^{n} w_t} \right\} \\
&= \frac{-2\sum_{k+1}^{t_1^0} w_t A(w,\boldsymbol{\beta}^0)}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t} \left(\sum_{t_1^0+1}^{t_2^0} w_t(\beta_2^0 - \beta_1^0) + \sum_{t_2^0+1}^{n} w_t(\beta_3^0 - \beta_1^0) \right) \\
&\quad - \frac{\sum_{k+1}^{t_1^0} w_t A(w,\boldsymbol{\beta}^0)^2}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t} \\
&= \frac{\sum_{k+1}^{t_1^0} w_t A(w,\boldsymbol{\beta}^0)^2}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t}.
\end{aligned}
$$

Therefore, we have

$$
L_n^w(k) - L_n^w(t_1^0) = \frac{\sum_{k+1}^{t_1^0} w_t A(w,\boldsymbol{\beta}^0)^2}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t} + n(R_n^w(k) - R_n^w(t_1^0)).
$$

We first derive a lower bound of $A(w,\boldsymbol{\beta}^0)^2/(\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t)$ for $k \in [1, t_1^0]$. Let $B = (1-\tau_1^0)(\beta_1^0 - \beta_2^0) + (1-\tau_2^0)(\beta_2^0 - \beta_3^0)$, and $\Delta = B^2/(1-\tau_1^0)$. We have

$$
\begin{aligned}
&\frac{A(w,\boldsymbol{\beta}^0)^2}{(\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t)} \\
&= \frac{\left(\frac{1}{n}\sum_{t_1^0+1}^{n}(w_t - 1)(\beta_1^0 - \beta_2^0) + \frac{1}{n}\sum_{t_2^0+1}^{n}(w_t - 1)(\beta_2^0 - \beta_3^0) + B \right)^2}{\left(\frac{1}{n}\sum_{k+1}^{n}(w_t - 1) + (n-k)/n \right) \left(\frac{1}{n}\sum_{t_1^0+1}^{n}(w_t - 1) + (1-\tau_1^0) \right)}.
\end{aligned}
\tag{B.13}
$$

By Lemma 8 (ii), all the stochastic terms in (B.13) are $O_{\mathbb{P}_w}(n^{-1/2})$ uniformly over $k \in [1, t_1^0]$. It is easily seen that for a fixed $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_w \left(\min_{1\le k\le t_1^0} \frac{A(w, \boldsymbol{\beta}^0)^2}{(\sum_{k+1}^n w_t \sum_{t_1^0+1}^n w_t)} > \Delta/2 \right) = 0 \quad a.s.. \tag{B.14}$$

Therefore, we have

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}_w \left(\inf_{k\in D_{n\varepsilon}} L_n^w(k) \le L_n^w(t_1^0) \right) &= \lim_{n\to\infty} \mathbb{P}_w \left(\bigcup_{k\in D_{n\varepsilon}} \{ L_n^w(k) - L_n^w(t_1^0) \le 0 \} \right) \\ = \lim_{n\to\infty} \mathbb{P}_w \left(\bigcup_{k\in D_{n\varepsilon}} \left\{ \frac{n(R_n^w(k) - R_n^w(t_1^0))}{\sum_{k+1}^{t_1^0} w_t} \le \frac{-A(w, \boldsymbol{\beta}^0)^2}{(\sum_{k+1}^n w_t \sum_{t_1^0+1}^n w_t)} \right\} \right) \\ \le \lim_{n\to\infty} \mathbb{P}_w \left(\bigcup_{k\in D_{n\varepsilon}} \left\{ \frac{n(R_n^w(k) - R_n^w(t_1^0))}{\sum_{k+1}^{t_1^0} w_t} \le -\frac{\Delta}{2} \right\} \right), \end{aligned} \tag{B.15}$$

where the last inequality follows from (B.14). Therefore, to prove Lemma 5, it is suffices to show that for any fixed $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_w \left(\sup_{k\in D_{n\varepsilon}} \left| \frac{n(R_n^w(k) - R_n^w(t_1^0))}{\sum_{k+1}^{t_1^0} w_t} \right| > \Delta/2 \right) = 0 \quad a.s., \tag{B.16}$$

which can be implied by proving $\sup_{k\in D_{n\varepsilon}} \left| n(R_n^w(k) - R_n^w(t_1^0))/\sum_{k+1}^{t_1^0} w_t \right| = o_{\mathbb{P}_w}(1)$.

We decompose $n(R_n^w(k) - R_n^w(t_1^0))$ for $k < t_1^0$ as follows,

$$\begin{aligned} & n(R_n^w(k) - R_n^w(t_1^0)) \\ &= 2a_k^w \sum_{k+1}^{t_1^0} w_t\epsilon_t + 2(b_k^w - b_{t_1^0}^w) \sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t + 2(c_k^w - c_{t_1^0}^w) \sum_{t_2^0+1}^{n} w_t\epsilon_t \\ &\quad - 2a_k^w \bar\epsilon_{-k}^w \sum_{k+1}^{t_1^0} w_t + 2(b_k^w \bar\epsilon_{-k}^w - b_{t_1^0}^w \bar\epsilon_{-t_1^0}^w) \sum_{t_1^0+1}^{t_2^0} w_t + 2(c_k^w \bar\epsilon_{-k}^w - c_{t_1^0}^w \bar\epsilon_{-t_1^0}^w) \sum_{t_2^0+1}^{n} w_t \\ &\quad - \left[\sum_{t=1}^{k} w_t(\bar\epsilon_k^w)^2 - \sum_{t=1}^{t_1^0} w_t(\bar\epsilon_{t_1^0}^w)^2 \right] - \left[\sum_{t=k+1}^{n} w_t(\bar\epsilon_{-k}^w)^2 - \sum_{t=t_1^0+1}^{n} w_t(\bar\epsilon_{-t_1^0}^w)^2 \right]. \end{aligned} \tag{B.17}$$

We will show that each term of $n(R_n^w(k) - R_n^w(t_1^0))$ divided by $\sum_{k+1}^{t_1^0} w_t$ is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in D_{n\varepsilon}$. The first term of (B.17) divided by $\sum_{k+1}^{t_1^0} w_t$ is

$$\frac{2a_k^w \sum_{k+1}^{t_1^0} w_t}{\sum_{k+1}^{t_1^0} w_t} = 2a_k^w \frac{\frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}(w_t-1)\epsilon_t + \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}\epsilon_t}{\frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}(w_t-1)+1}. \tag{B.18}$$

By Lemma 7, for an arbitrary $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}_w\left(\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \left| \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}(w_t-1)\epsilon_t \right| > \delta\right) \\
&\le \frac{1}{\delta^2}\left(\frac{1}{t_1^0-\varepsilon\log(n)}\sum_{k=t_1^0-\varepsilon\log(n)}^{t_1^0}\frac{\epsilon_t^2}{(t_1^0-k)^2} + \sum_{k=n\eta}^{t_1^0-\varepsilon\log(n)}\frac{\epsilon_t^2}{(t_1^0-k)^2}\right) \; a.s..
\end{aligned} \tag{B.19}$$

The right hand side converges almost surely to 0 as $n \to 0$ by invoking the Kolmogorov's three series theorem. Next, we prove that $\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}\epsilon_t \to 0, a.s..$ For any $\delta > 0$,

$$\begin{aligned}
&\lim_{m\to\infty}\mathbb{P}\left(\bigcup_{n=m}^{\infty}\left\{\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \left| \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}\epsilon_t \right| > \delta\right\}\right) \\
&\le \lim_{m\to\infty}\mathbb{P}\left(\max_{k\ge\varepsilon\log(m)} \left| \frac{1}{k}\sum_{t=1}^{k}\epsilon_t \right| > \delta\right) \le \lim_{m\to\infty}\frac{1}{\varepsilon\log(m)}\sum_{k=1}^{\infty}\frac{1}{k^2},
\end{aligned} \tag{B.20}$$

where the last inequality follows from Lemma 6. Thus, by (B.20) we have proved that $\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}\epsilon_t \to 0, a.s..$ Similarly, for any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}_w\left(\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \left| \frac{1}{t_1^0-k}\sum_{k+1}^{t_1^0}(w_t-1) \right| > \delta\right) \\
&\le \frac{1}{\delta^2}\left(\frac{1}{t_1^0-\varepsilon\log(n)}\sum_{k=t_1^0-\varepsilon\log(n)}^{t_1^0}\frac{1}{(t_1^0-k)^2} + \sum_{k=n\eta}^{t_1^0-\varepsilon\log(n)}\frac{1}{(t_1^0-k)^2}\right),
\end{aligned} \tag{B.21}$$

the right hand side converges to 0 obviously. By combining (B.19), (B.20), (B.21) and that a_k^w is bounded from above, we have

$$\max_{n\eta\le k\le t_1^0-\varepsilon\log(n)} \frac{2a_k^w\sum_{k+1}^{t_1^0} w_t}{\sum_{k+1}^{t_1^0} w_t} = o_{\mathbb{P}_w}(1).$$

For the second term of (B.17), note that

$$\begin{aligned}\frac{(b_k^w - b_{t_1^0}^w)\sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t}{\sum_{k+1}^{t_1^0} w_t} &= \frac{A(w,\boldsymbol{\beta}^0)\sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t} \\ &\leq 2(|\beta_1^0|+|\beta_2^0|+|\beta_3^0|)\frac{\frac{1}{n-t_1^0}\sum_{t_1^0+1}^{t_2^0}(w_t-1)\epsilon_t + \frac{1}{n-t_1^0}\sum_{t_1^0+1}^{t_2^0}\epsilon_t}{\frac{1}{n-t_1^0}\sum_{t_1^0+1}^{n}(w_t-1)+1}\end{aligned} \tag{B.22}$$

Therefore, by Lemma 8 (i), (ii), and the strong law of large numbers, we have

$$\max_{n\eta\leq k\leq t_1^0-\varepsilon\log(n)} \frac{(b_k^w - b_{t_1^0}^w)\sum_{t_1^0+1}^{t_2^0} w_t\epsilon_t}{\sum_{k+1}^{t_1^0} w_t} = o_{\mathbb{P}_w}(1).$$

The third term of (B.17) can be treated similarly as the second term. To treat the fourth term, we first deal with $\bar{\epsilon}_k^w$,

$$\begin{aligned}\bar{\epsilon}_k^w &= \frac{\sum_{t=k+1}^{n} w_t\epsilon_t}{\sum_{t=k+1}^{n} w_t} \\ &= \frac{\frac{1}{n-k}\sum_{t=k+1}^{n}(w_t-1)\epsilon_t + \frac{1}{n-k}\sum_{t=k+1}^{n}\epsilon_t}{\frac{1}{n-k}\sum_{t=k+1}^{n}(w_t-1)+1}.\end{aligned} \tag{B.23}$$

By the same technique in deriving (B.19) and (B.20), we know that $\max_{n\eta\leq k\leq t_1^0-\varepsilon\log(n)} \bar{\epsilon}_k^w = o_{\mathbb{P}_w}(1)$. Therefore, the fourth term is $o_{\mathbb{P}_w}(1)$ uniformly over $k\in[n\eta, t_1^0-\varepsilon\log(n)]$.

The fifth term of (B.17) can be written as

$$\begin{aligned}&(b_k^w\bar{\epsilon}_{-k}^w - b_{t_1^0}^w\bar{\epsilon}_{-t_1^0}^w)\sum_{t_1^0+1}^{t_2^0} w_t \\ &= (b_k^w - b_{t_1^0}^w)\bar{\epsilon}_{-k}^w\sum_{t_1^0+1}^{t_2^0} w_t + b_{t_1^0}^w(\bar{\epsilon}_{-k}^w - \bar{\epsilon}_{-t_1^0}^w)\sum_{t_1^0+1}^{t_2^0} w_t \\ &= \frac{A(w,\boldsymbol{\beta}^0)\sum_{t_1^0+1}^{t_2^0} w_t}{\sum_{k+1}^{n} w_t\sum_{t_1^0+1}^{n} w_t}\sum_{k+1}^{t_1^0} w_t\bar{\epsilon}_{-k}^w + b_{t_1^0}^w\sum_{t_1^0+1}^{t_2^0} w_t\frac{\sum_{k+1}^{t_1^0} w_t\epsilon_t}{\sum_{k+1}^{n} w_t} \\ &\quad - b_{t_1^0}^w\sum_{t_1^0+1}^{t_2^0} w_t\frac{\sum_{k+1}^{t_1^0} w_t\sum_{t_1^0+1}^{n} w_t\epsilon_t}{\sum_{k+1}^{n} w_t\sum_{t_1^0+1}^{n} w_t} := V_1(k)+V_2(k)+V_3(k).\end{aligned}$$

For the first part, $|V_1(k)|/\sum_{k+1}^{t_1^0} w_t \le C|\bar{\epsilon}_{-k}^w|$. Thus, by (B.23), we obtain

$$\max_{n\eta \le k \le t_1^0 - \varepsilon \log(n)} |V_1(k)|/\sum_{k+1}^{t_1^0} w_t = o_{\mathbb{P}_w}(1).$$

For the second part,

$$\frac{V_2(k)}{\sum_{k+1}^{t_1^0} w_t} = b_{t_1^0}^w \frac{\sum_{t_1^0+1}^{t_2^0} w_t}{\sum_{k+1}^{t_1^0} w_t} \frac{\sum_{k+1}^{t_1^0} w_t \epsilon_t}{\sum_{k+1}^{n} w_t},$$

which is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in [n\eta, t_1^0 - \varepsilon \log(n)]$ because (B.18) does so. For the third part, we have

$$\begin{aligned} \frac{V_3(k)}{\sum_{k+1}^{t_1^0} w_t} &= b_{t_1^0}^w \frac{\sum_{t_1^0+1}^{t_2^0} w_t \sum_{t_1^0+1}^{n} w_t}{\sum_{k+1}^{n} w_t \sum_{t_1^0+1}^{n} w_t} \frac{\sum_{t_1^0+1}^{n} w_t \epsilon_t}{\sum_{t_1^0+1}^{n} w_t} \\ &\le 2(|\beta_1^0| + |\beta_2^0| + |\beta_3^0|) \frac{\sum_{t_1^0+1}^{n} w_t \epsilon_t}{\sum_{t_1^0+1}^{n} w_t}, \end{aligned}$$

which is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in [n\eta, t_1^0 - \varepsilon \log(n)]$. By combining the convergence result of $V_1(k)$, $V_2(k)$, and $V_3(k)$, we obtain that the fifth term of (B.17) divided by $\sum_{k+1}^{t_1^0} w_t$ is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in [n\eta, t_1^0 - \varepsilon \log(n)]$.

The sixth term of (B.17) is treated similarly as the fifth term. The second to the last term divided by $\sum_{k+1}^{t_1^0} w_t$ can be decomposed as

$$\begin{aligned} &\left(\sum_{t=1}^{k} w_t (\bar{\epsilon}_k^w)^2 - \sum_{t=1}^{t_1^0} w_t (\bar{\epsilon}_{t_1^0}^w)^2 \right) / \sum_{k+1}^{t_1^0} w_t \\ &= \frac{\left(\sum_{t=1}^{k} w_t \epsilon_t \right)^2}{\sum_{k+1}^{t_1^0} w_t \sum_{t=1}^{k} w_t} - \frac{\left(\sum_{t=1}^{t_1^0} w_t \epsilon_t \right)^2}{\sum_{k+1}^{t_1^0} w_t \sum_{t=1}^{t_1^0} w_t}. \end{aligned}$$

which is $o_{\mathbb{P}_w}(1)$ uniformly over $k \in [n\eta, t_1^0 - \varepsilon \log(n)]$ using the same techniques in proving (B.7) and (B.8). The last term is treated similarly as the second to the last term.

By combining all the convergence results we obtained, we have

$$\max_{n\eta \le k \le t_1^0 - \varepsilon \log(n)} \frac{n(R_n^w(k) - R_n^w(t_1^0))}{\sum_{k+1}^{t_1^0} w_t} = o_{\mathbb{P}_w}(1),$$

which leads to (B.16). The proof of Lemma 5 is completed.

Proof of Theorem 1. Note that the binary segmentation algorithm is implemented sequentially in this paper, we first study the convergence rate of $\hat{t}^w$, which is first change-point detected by the BS algorithm defined in Section 2.2. For a fixed $\varepsilon > 0$, recall that $D_{n\varepsilon} = \{k : n\eta \leq k \leq n\tau_2^0(1-\eta),\ |k - t_1^0| > \varepsilon \log(n)\}$,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}_w(|\hat{t}^w - t_1^0| > \log(n)\varepsilon) \\
&\leq \lim_{n\to\infty} \mathbb{P}_w(\hat{t}^w \leq n\eta \text{ or } \hat{t}^w \geq n\tau_2^0(1-\eta)) + \lim_{n\to\infty} \mathbb{P}_w(\hat{t}^w \in D_n) \\
&\leq \lim_{n\to\infty} \mathbb{P}_w(\hat{t}^w \leq n\eta \text{ or } \hat{t}^w \geq n\tau_2^0(1-\eta)) + \lim_{n\to\infty} \mathbb{P}_w(\inf_{k\in D_{n\varepsilon}} L_n^w(k) \leq L_n^w(t_1^0)) \\
&= 0 \quad a.s.,
\end{aligned}$$

where the last equality follows from Lemma 4 and Lemma 5. Thus, we obtain $|\hat{t}^w - t_1^0| = o_{\mathbb{P}_w}(\log(n))$.

Have obtained the convergence rate of the first change-point $\hat{t}^w$, we can proceed to analyze the convergence rate a further change-point in the sub-sample $[\hat{t}^w, n]$. Since we have proved that $\hat{t}^w - t_1^0 = o_{\mathbb{P}_w}(\log(n))$, the difference between the residual sum of squares using the sub-sample $[\hat{t}^w, n]$ and the sub-sample $[t_1^0, n]$ is of order $o_{\mathbb{P}_w}(\log(n))$, which is ignorable in the proof. Thus, we can use the same technique used in proving the convergence rate of $\hat{t}^w$ to prove that the second estimated change-point converges to t_2^0 with the rate $o_{\mathbb{P}_w}(\log(n))$.

Proof of Theorem 2. Denote the weighted mean of $X_t,\ t = k_1, \ldots, k_2$ by the weights $w_t,\ t = k_1, \ldots, k_2$ by $\bar{X}^w_{k_1,k_2} = (\sum_{k_1+1}^{k_2} w_t)^{-1} \sum_{k_1+1}^{k_2} w_t X_t$, denote the weighted-CUSUM statistic for the sample $X_t,\ t = k_1, \ldots, k_2$ by

$$\begin{aligned}
V^w_{k_1,k_2}(k) &= \left| \sqrt{\frac{\sum_{t=k_1}^{k_2} w_t}{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}} \left[\sum_{t=k_1}^{k} w_t (X_t - \bar{X}^w_{k_1,k_2}) \right] \right| \\
&= \left| \sqrt{\frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=k_1}^{k_2} w_t}} \left[\bar{X}^w_{k_1,k} - \bar{X}^w_{k,k_2} \right] \right|.
\end{aligned}$$

We first study the difference between $(\hat{\sigma}^w_k)^2$ and $(\hat{\sigma}^w_{k-1})^2$, the subscript k and $k-1$ denote the number of change-points for the candidate model. Note that the estimated set of change-points are nested, thus $(\hat{\sigma}^w_{k-1})^2 - (\hat{\sigma}^w_k)^2$ must be of the form

$$
\begin{aligned}
&(\hat{\sigma}^w_{k-1})^2 - (\hat{\sigma}^w_k)^2 = \\
&\quad \frac{\sum_{t=k_1}^{k_2} w_t (X_t - \bar{X}^w_{k_1,k_2})^2 - \sum_{t=k_1}^{k} w_t (X_t - \bar{X}^w_{k_1,k})^2 - \sum_{t=k}^{k_2} w_t (X_t - \bar{X}^w_{k,k_2})^2}{\sum_{t=1}^{n} w_t} \\
&= \frac{(V^2_{k_1,k_2})^2}{\sum_{t=1}^{n} w_t} = \frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=1}^{n} w_t \sum_{t=k_1}^{k_2} w_t} \left[\bar{X}^w_{k_1,k} - \bar{X}^w_{k,k_2}\right]^2,
\end{aligned}
$$

where k, k_1, k_2 are change-points estimated by the binary segmentation algorithm. Note that the number of change-points is $N = 2$ in the true model, we use N in the proof for clarity. By Theorem 1, the change-points are estimated by the precision $o_{\mathbb{P}_w}(\log(n))$,

$$
\begin{aligned}
(\hat{\sigma}^w_{N-1})^2 - (\hat{\sigma}^w_N)^2 &= \frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=1}^{n} w_t \sum_{t=k_1}^{k_2} w_t} \left[\bar{X}^w_{k_1,k} - \bar{X}^w_{k,k_2}\right]^2 \\
&= \frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=1}^{n} w_t \sum_{t=k_1}^{k_2} w_t} \left[\beta^0_1 - \beta^0_2 + \bar{\epsilon}^w_{k_1,k} - \bar{\epsilon}^w_{k,k_2} + o_{\mathbb{P}_w}(n^{-1}\log(n))\right]^2 \\
&= (\beta_1 - \beta_2)^2 + o_{\mathbb{P}_w}(1),
\end{aligned}
$$

where $\bar{\epsilon}^w_{k_1,k_2} = (\sum_{k_1}^{k_2} w_t)^{-1} \sum_{k_1}^{k_2} w_t \epsilon_t$, β^0_1 is the mean of X_t in the segment $X_t, t = k_1, \ldots, k$, and β^0_2 is the mean of X_t in the segment $X_t, t = k, \ldots, k_2$. For $k < N$,

$$
\begin{aligned}
&(\hat{\sigma}^w_k)^2 - (\hat{\sigma}^w_N)^2 = (\hat{\sigma}^w_k)^2 - (\hat{\sigma}^w_{k+1})^2 +, \ldots, + (\hat{\sigma}^w_{N-1})^2 - (\hat{\sigma}^w_N)^2 \\
&\geq (\hat{\sigma}^w_{N-1})^2 - (\hat{\sigma}^w_N)^2 = (\beta_1 - \beta_2)^2 + o_{\mathbb{P}_w}(1).
\end{aligned}
$$

We have for $k < N$ and some positive constant $C > 0$,

$$
\begin{aligned}
BIC(k) - BIC(N) &= \frac{n}{2} \log \frac{(\hat{\sigma}^w_k)^2}{(\hat{\sigma}^w_k)^2} + (k - N)\log(n) \\
&= \frac{n}{2} \log \left(1 + \frac{(\hat{\sigma}^w_k)^2 - (\hat{\sigma}^w_k)^2}{(\hat{\sigma}^w_k)^2}\right) + (k - N)\log(n) \\
&\geq Cn \frac{(\hat{\sigma}^w_k)^2 - (\hat{\sigma}^w_k)^2}{(\hat{\sigma}^w_k)^2} + (k - N)\log(n) \quad a.s. \\
&\geq Cn + (k - N)\log(n) > 0 \quad a.s.,
\end{aligned}
$$

which is true for sufficiently large n. Next we consider the case where $k > N$, for $k =$

$N+1$ we have,

$$(\hat{\sigma}_N^w)^2 - (\hat{\sigma}_{N+1}^w)^2 = \frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=1}^{n} w_t \sum_{t=k_1}^{k_2} w_t} \left[\bar{X}_{k_1,k}^w - \bar{X}_{k,k_2}^w\right]^2$$
$$= \frac{\sum_{t=k_1}^{k} w_t \sum_{t=k+1}^{k_2} w_t}{\sum_{t=1}^{n} w_t \sum_{t=k_1}^{k_2} w_t} \left[\bar{\epsilon}_{k_1,k}^w - \bar{\epsilon}_{k,k_2}^w + o_p(n^{-1})\right]^2 = o_{\mathbb{P}_w}(n^{-1}\log(n)).$$

Thus for $k > N$, we have $(\hat{\sigma}_k^w)^2 - (\hat{\sigma}_N^w)^2 = o_{\mathbb{P}_w}(n^{-1}\log(n))$, and

$$\begin{aligned} BIC(k) - BIC(N) &= \frac{n}{2}\log\frac{(\hat{\sigma}_k^w)^2}{(\hat{\sigma}_k^w)^2} + (k-N)\log(n) \\ &\geq Cn\frac{(\hat{\sigma}_k^w)^2 - (\hat{\sigma}_k^w)^2}{(\hat{\sigma}_k^w)^2} + (k-N)\log(n) \\ &\geq o_{\mathbb{P}_w}(\log(n)) + (k-N)\log(n), \end{aligned}$$

which is positive for large n almost surely. Thus for large n, $BIC(k)$ is minimized at $k = N = 2$ almost surely.

Proof of Theorem 3. Without loss of generality, we prove the asymptotic consistency of the confidence interval for μ_t^0 for a fixed location t such that $t_1^0 - t > \log(n)$. Note that in the proof t is used to indicate a fixed notation. By Theorem 2, we have that the probability of the estimated number of change-points for the bootstrap samples equals the true number of change-points goes to 1 conditional on the observations almost surely. Take $\varepsilon = 1/2$ in Theorem 1, we have $\lim_{n\to\infty} \mathbb{P}_w(|\hat{t}_1^w - t_1^0| < 1/2\log(n)) = 1$ almost surely. Therefore, $\lim_{n\to\infty} \mathbb{P}_w(\hat{t}_1^w - t > 1/2\log(n)) = 1$ almost surely. Therefore, $\lim_{n\to\infty} \mathbb{P}_w(\hat{\mu}_t^w = \sum_{s=1}^{\hat{t}_1^w} w_s X_s / \sum_{s=1}^{\hat{t}_1^w} w_s) = 1$ almost surely, where $\hat{\mu}_t^w$ denotes the bootstrap analog of $\hat{\mu}_t = 1/\hat{t}_1 \sum_{s=1}^{\hat{t}_1} X_s$.

Denote $\check{\mu}_t^w = \sum_{s=1}^{t_1^0} w_s X_s / \sum_{s=1}^{t_1^0} w_s$, we shall show that $|\hat{\mu}_t^w - \check{\mu}_t^w| = o_{\mathbb{P}_w}(\log(n)^2/n)$. In view of the inequality

$$|\hat{\mu}_t^w - \check{\mu}_t^w| \leq \frac{\left|\sum_{s=1}^{t_1^0} w_s X_s \sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s\right| + \left|\sum_{s=1}^{t_1^0} w_s \sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s X_s\right|}{\sum_{s=1}^{t_1^0} w_s \sum_{s=1}^{\hat{t}_1^w} w_s},$$

we only need to prove that

$$\sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s = o_{\mathbb{P}_w}(\log(n)^2), \text{ and } \sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s X_s = o_{\mathbb{P}_w}(\log(n)^2). \tag{B.24}$$

For any $\varepsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}_w\left(\sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s > \varepsilon \log(n)^2\right) \\
&\leq \mathbb{P}_w\left(\sum_{s=\hat{t}_1^w+1}^{t_1^0} w_s > \varepsilon \log(n)^2, |\hat{t}_1^w - t_1^0| < \varepsilon \log(n)\right) \\
&\quad + \mathbb{P}_w(|\hat{t}_1^w - t_1^0| \geq \epsilon \log(n)) \\
&\leq \mathbb{P}_w\left(\sum_{s=t_1^0-\log(n)+1}^{t_1^0} w_s > \varepsilon \log(n)^2, |\hat{t}_1^w - t_1^0| < \varepsilon \log(n)\right) \\
&\quad + \mathbb{P}_w(|\hat{t}_1^w - t_1^0| \geq \varepsilon \log(n)) \\
&\leq \frac{\log(n)}{\varepsilon \log(n)^2} + \mathbb{P}_w(|\hat{t}_1^w - t_1^0| \geq \varepsilon \log(n)) \to 0 \quad a.s.,
\end{aligned}
$$

where the second inequality follows from positiveness of w_s, the third inequality follows from the Chebyshev's inequality, and the convergence follows from Theorem 1. The proof of second part of (B.24) is the same as the first part except we need to apply a strong law of large numbers to sum of $|X_s|$.

Theorem 3.6.13 of van der Vaart & Weller (1996) develops the general theory for the weighted bootstrap measures, with the continuous mapping theorem (van der Vaart & Weller, 1996, Theorem 1.11.1), we have

$$\sqrt{t_1^0}\left(\check{\mu}_t^w - \frac{1}{t_1^0}\sum_{s=1}^{t_1^0} X_s\right) \rightsquigarrow N(0, \sigma^2),$$

in probability, given the almost all the observations $X_1, X_2, \ldots$, where $\rightsquigarrow$ stands for weak converge. For the exact meaning of the above statement, we refer the reader to Section 2.9 and Section 3.6 of van der Vaart & Weller (1996). Together with the already established result $|\hat{\mu}_t^w - \check{\mu}_t^w| = o_{\mathbb{P}_w}(\log(n)^2/n)$, we have

$$\sqrt{t_1^0}\left(\hat{\mu}_t^w - \frac{1}{t_1^0}\sum_{s=1}^{t_1^0} X_s\right) \rightsquigarrow N(0, \sigma^2),$$

in probability, given the almost all the observations $X_1, X_2, \ldots$. The asymptotic consistency of the percentile interval follows from Lemma 23.3 of Van der Vaart (2000).

Note that the bagging estimator is the conditional expectation of $\hat{\mu}_t^w$ given the observations. By standard probability manipulations similar to the proof of (B.24), we can show that $n/\log(n)^2\mathbb{E}_w(|\hat{\mu}_t^w - \check{\mu}_t^w|) \to 0$ almost surely. Also note that $\mathbb{E}_w(\check{\mu}_t^w) = 1/t_1^0 \sum_{s=1}^{t_1^0} X_t$. The delta method approximation of the variance of the sample mean is simply the sample variance, see Chapter 6 of Efron (1982). Therefore, the consistency of the smoothed interval follows from the central limit theorem. Because the adaptive interval is either the smoother interval or the percentile interval, the adaptive interval is also consistent.

References

Arnold, T. B. & Tibshirani, R. J. (2014). *genlasso: Path algorithm for generalized lasso problems*. R package version 1.3.
URL: *https://CRAN.R-project.org/package=genlasso*

Bai, J. (1994). Least squares estimation of a shift in linear processes, *J. Time Series Anal.* **15**(5): 453–472.

Bai, J. (1997). Estimating multiple breaks one at a time, *Econometric Theory* **13**(3): 315–352.

Baranowski, R. & Fryzlewicz, P. (2015). *wbs: Wild Binary Segmentation for Multiple Change-Point Detection*. R package version 1.3.
URL: *https://CRAN.R-project.org/package=wbs*

Chung, K. L. (1948). On the maximum partial sums of sequences, *Trans. Amer. Math. Soc.* **64**: 205–233.

Cleynen, A., Rigaill, G. & Koskas, M. (2016). *Segmentor3IsBack: A Fast Segmentation Algorithm*. R package version 2.0.
URL: *https://CRAN.R-project.org/package=Segmentor3IsBack*

Efron, B. (1982). *The Jackknife, the Bootstrap, and other Resampling Plans*, Philadelphia: SIAM.

Fazekas, I. & Klesov, O. (2001). A general approach to the strong law of large numbers, *Theory Probab. Appl.* **45**(3): 436–449.

Frick, K., Axel, M. & Hannes, S. (2014). Multiscale change point inference, *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**(3): 495–580.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection, *Ann. Statist.* **42**(6): 2243–2281.

Hájek, J. & Rényi, A. (1955). Generalization of an inequality of kolmogorov, *Acta Math. Hungar.* **6**(3-4): 281–283.

Hu, S. & Hu, M. (2006). A general approach rate to the strong law of large numbers, *Statist. Probab. Lett.* **76**(8): 843–851.

Killick, R., Haynes, K. & Eckley, I. A. (2016). *changepoint: An R package for changepoint analysis*. R package version 2.2.2.
URL: *https://CRAN.R-project.org/package=changepoint*

Muggeo, V. M. (2012). *cumSeg: Change point detection in genomic sequences*. R package version 1.1.
URL: *https://CRAN.R-project.org/package=cumSeg*

Pein, F., Hotz, T., Sieling, H. & Aspelmeier, T. (2017). *stepR: Multiscale change-point inference*. R package version 2.0-1.
URL: *https://CRAN.R-project.org/package=stepR*

Rao, B. P. (2009). Conditional independence, conditional mixing and conditional association, *Ann. Inst. Statist. Math.* **61**(2): 441–460.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge University Press.

van der Vaart, A. W. & Weller, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer-Verlag.