SUPPLEMENT TO "STATISTICAL INFERENCE FOR FOUR-REGIME SEGMENTED REGRESSION MODELS"

BY HAN YAN[1,a] AND SONG XI CHEN[2,b]

[1]*Guanghua School of Management, Peking University,* [a]*hanyan@stu.pku.edu.cn*

[2]*Department of Statistics and Data Science, Tsinghua University,* [b]*sxchen@tsinghua.edu.cn*

The supplementary materials contain additional details, theoretical proofs, and additional results on simulations and case studies of the paper " Statistical Inference for Four-regime Segmented Regression Models".

Organization

A Auxiliary lemmas 2
B Proofs for Section 3 10
C Proof for Section 4 and additional algorithms 38
D Proofs for Section 5 42
E Proofs for Section 6 54
F Auxiliary assumptions 57
G Extension to general segmented regressions 59
H Additional simulation results 59
I Additional case study results 64
References 70

Notations. Throughout the supplementary material, we use $c_1, c_2, C_1, C_2, \cdots$ to denote generic finite positive constants, which may differ from line to line. We use $\mathbb{1}(\mathcal{A})$ as the indicator function of an event $\mathcal{A}$. For any vector $\boldsymbol{v} = (v_1, \cdots, v_d)^{\mathrm{T}}$, let $\|\boldsymbol{v}\| = (\sum_{i=1}^d v_i^2)^{1/2}$ be its L_2-norm. For any $r > 0$, we define $\mathcal{N}(\boldsymbol{v}_0; r) = \{\boldsymbol{v} : \|\boldsymbol{v} - \boldsymbol{v}_0\| \le r\}$. Denote by $\boldsymbol{v}_{-1}$ as the sub-vector of $\boldsymbol{v}$ excluding its first element, i.e., $\boldsymbol{v}_{-1} = (v_2, \cdots, v_d)^{\mathrm{T}}$. For any two sets A, B, we let $A \setminus B = A \cap B^c$, where B^c is the complement of B, and $A \triangle B = (A \setminus B) \cup (B \setminus A)$. The empirical measure $\mathbb{E}_T(\cdot)$ denotes the sample average of a sequence of random elements with T observations, i.e., $\mathbb{E}_T(\boldsymbol{X}_t) = T^{-1} \sum_{t=1}^T \boldsymbol{X}_t$. We also denote $\mathbb{G}_T(\cdot) = \sqrt{T}\{\mathbb{E}_T(\cdot) - \mathbb{E}(\cdot)\}$.

For the four-regime regression model

$$Y_t = \sum_{k=1}^4 \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\beta}_k \mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\} + \varepsilon_t,$$

we define the indicator functions for the t-th observation on the k-th regions as

$$\mathbb{1}_t^{(k)}(\boldsymbol{\gamma}) := \mathbb{1}\left\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\right\} \text{ for } k \in \{1, \cdots, 4\};$$

and for $l = 1$ and 2, let

$$\mathbb{1}_{l,t}(\boldsymbol{\gamma}) := \mathbb{1}(\boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} > 0) \quad \text{and} \quad \mathbb{1}_{l,t}(\boldsymbol{\gamma}, \tilde{\boldsymbol{\gamma}}) := \mathbb{1}(\boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} \le 0 < \boldsymbol{Z}_{l,t}^{\mathrm{T}} \tilde{\boldsymbol{\gamma}}). \tag{1}$$

For each $1 \le k \le 4$, that $\boldsymbol{z} = (\boldsymbol{z}_1, \boldsymbol{z}_2) \in R_k(\boldsymbol{\gamma})$ or not depends on the signs of $\boldsymbol{z}_1^{\mathrm{T}} \boldsymbol{\gamma}_1$ and $\boldsymbol{z}_2^{\mathrm{T}} \boldsymbol{\gamma}_2$. As results, for each $l = 1$ and 2 and $1 \le k \le 4$, we denote

$$s_l^{(k)} = \operatorname{sign}\left(\boldsymbol{z}_l^{\mathrm{T}} \boldsymbol{\gamma}_l\right), \quad \text{for} \quad (\boldsymbol{z}_1, \boldsymbol{z}_2) \in R_k(\boldsymbol{\gamma}), \tag{2}$$

for each $l \in \{1,2\}$ and $k \in \{1,\cdots,4\}$, which is well-defined since any $\boldsymbol{z} \in R_k(\boldsymbol{\gamma})$ has the same $\operatorname{sign}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\gamma}_l)$. Specifically, in the four-regime model, we have $s_1^{(1)} = s_2^{(1)} = 1; s_1^{(2)} = -1, s_2^{(2)} = 1; s_1^{(3)} = s_2^{(3)} = -1$; and $s_1^{(4)} = 1, s_2^{(4)} = -1$. We now define the pairs of *adjacent* sub-regions. For the l-th splitting hyperplane, we let

$$\mathcal{S}(l) = \left\{(j,k) : s_l^{(j)} \neq s_l^{(k)} \text{ and } s_i^{(j)} = s_i^{(k)} \text{ if } i \neq l\right\}, \tag{3}$$

that is, $\{\boldsymbol{z}_l^{\mathrm{T}}\boldsymbol{\gamma}_l = 0\}$ is the only splitting hyperplane that $R_j(\boldsymbol{\gamma})$ and $R_k(\boldsymbol{\gamma})$ are on opposite directions of it. Specifically, in the four-regime model, $\mathcal{S}(1) = \{(1,2),(3,4),(2,1),(4,3)\}$ and $\mathcal{S}(2) = \{(1,4),(2,3),(4,1),(3,2)\}$. Let

$$m(\boldsymbol{W}_t,\boldsymbol{\theta}) = \{Y_t - \sum_{k=1}^{4} \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta}_k \mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\boldsymbol{\gamma}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\boldsymbol{\gamma}_2)\}^2.$$

We denote by $\mathbb{M}_T(\boldsymbol{\theta}) = \mathbb{E}_T\{m(\boldsymbol{W}_t,\boldsymbol{\theta})\}$ and $\mathbb{M}(\boldsymbol{\theta}) = \mathbb{E}\{m(\boldsymbol{W}_t,\boldsymbol{\theta})\}$ for any $\boldsymbol{\theta} \in \Theta$.

APPENDIX A: AUXILIARY LEMMAS

In this section, we provide some useful lemmas that will be constantly used in the proofs of main results.

A.1. Lemmas for moment inequalities and empirical processes.

The following lemma establishes a uniform law of large numbers for the segmented linear models with an α-mixing sequence of observations.

LEMMA A.1 (Glivenko-Cantelli). *Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\mathrm{T}}, \boldsymbol{\gamma}_2^{\mathrm{T}})^{\mathrm{T}} \in \prod_{l=1}^{2}\Gamma_l$. Let $U_t = U(\boldsymbol{W}_t)$ be a function of $\boldsymbol{W}_t$ with $\sup_t \mathbb{E}\|U_t\|^4 < \infty$. Then under the α-mixing condition in Assumption 1, for each $k \in \{1,\cdots,4\}$ we have*

$$\sup_{\boldsymbol{\gamma}\in\prod_{l=1}^{2}\Gamma_l} |\mathbb{E}_T\{U_t \mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\}\} - \mathbb{E}\{U_t \mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\}\}| = o_p(1).$$

REMARK A.1. In this lemma, the geometric decaying rate of the α-mixing coefficient in Assumption 1 can be relaxed as a polynomial rate satisfying $\sum_{t=1}^{\infty}\alpha(t)^{1-\frac{2}{r}} < \infty$ for some $r > 2$.

PROOF. Let $\mathcal{F}_l = \{\boldsymbol{z}_l : \boldsymbol{z}_l^{\mathrm{T}}\boldsymbol{\gamma} < 0, \boldsymbol{\gamma} \in \Gamma_l\}$. By Example 2.6.1 of van der Vaart and Wellner (1996) we know that the VC-dimension of $\mathcal{F}_l$ is $\mathrm{VC}(\mathcal{F}_l) = d_l$, where d_l is the dimension of $\boldsymbol{z}_l$ for $l = 1$ and 2. Let $R_k = \{R_k(\boldsymbol{\gamma}), \boldsymbol{\gamma} \in \prod_{l=1}^{2}\Gamma_l\}$. Then, R_k consists of intersection of sets in $\{\mathcal{F}_l, l \in \{1,2\}\}$ or their complements. Then, according to Lemma 2.6.17 of van der Vaart and Wellner (1996), R_k is a VC-class which can pick out at most $O(n^{\sum_{l=1}^{2}d_l-2})$ subsets of any given set $\{\mathbf{x}_i\}_{i=1}^{n}$ for $\mathbf{x}_i \in \mathbb{R}^{\sum_{l=1}^{2}d_l}$. Hence, by Lemma 2.6.18 of van der Vaart and Wellner (1996), the function class $\mathcal{G}_k = \{g(u,\boldsymbol{z}) = u\mathbb{1}(\boldsymbol{z} \in R), R \in R_k\}$ is a VC-subgraph function class, which implies that $\mathcal{G}_k$ has a finite uniform covering numbers.

For any fixed $\boldsymbol{\gamma} \in \prod_{l=1}^{2}\Gamma_l$, by the ergodic thoeorem for the α-mixing processes (see Theorem 10.2.1 of Doob, 1953), we have $|\mathbb{E}_T\{U_t\mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\}\} - \mathbb{E}\{U_t\mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma})\}\}| = o_p(1)$ for each $k \in [4]$. Because the covering number of $\mathcal{G}_k$ is finite, using the same arguments as in Theorem 2.4.1 of van der Vaart and Wellner (1996), the uniform weak law of large numbers is established. □

The next lemma provides useful moment inequalities about perturbations of $\boldsymbol{\gamma}_0$ around its neighborhoods.

LEMMA A.2. *Suppose that U is a random variable that satisfies $M_0 < \mathbb{E}(U|\boldsymbol{Z}_\ell^{\mathrm{T}}\boldsymbol{\gamma}=0) < M_1$ almost surely with some constants $M_0, M_1 > 0$ for any $\ell \in \{1,2\}$, where $\boldsymbol{\gamma} \in \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta)$ for some $\delta > 0$.*

(i) Under Assumption 3.(ii), there exist constants $c_1, \delta_1 > 0$, such that if $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta_1)$, then

$$\mathbb{E}\left\{U\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_1)-\mathbb{1}_\ell(\boldsymbol{\gamma}_2)\right|\right\} \le c_1\|\gamma_1-\gamma_2\|. \tag{A.1}$$

(ii) Under Assumption 4.(i), there exist constants $c_2, \delta_2 > 0$, such that if $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta_2)$, then

$$\mathbb{E}\left\{U\mathbb{1}(\boldsymbol{Z}_\ell \in R)\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})-\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)\right|\right\} \ge c_2\|\boldsymbol{\gamma}_{\ell 0}-\boldsymbol{\gamma}_\ell\|. \tag{A.2}$$

where $R = R_k(\boldsymbol{\gamma}_0) \cup R_h(\boldsymbol{\gamma}_0)$ with $(k,h) \in \mathcal{S}(\ell)$.

(iii) Under Assumption 4.(iii), there exist constants $c_3, \delta_3 > 0$, such that if $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{N}(\boldsymbol{\gamma}_{10};\delta_3)$ and $\boldsymbol{\gamma}_3, \boldsymbol{\gamma}_4 \in \mathcal{N}(\boldsymbol{\gamma}_{20};\delta_3)$, then

$$\mathbb{E}\left\{U\left|\mathbb{1}_1(\boldsymbol{\gamma}_1)-\mathbb{1}_1(\boldsymbol{\gamma}_2)\right|\left|\mathbb{1}_2(\boldsymbol{\gamma}_3)-\mathbb{1}_2(\boldsymbol{\gamma}_4)\right|\right\} \le c_3\|\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2\|\|\boldsymbol{\gamma}_3-\boldsymbol{\gamma}_4\|. \tag{A.3}$$

PROOF. (i) Let $\delta_1 = \min(\delta,\delta_0)$, where δ_0 is specified in Assumption 3 (ii) and δ is in the assumption of Lemma A.2 (i). Denote $\mathcal{N}_{1\ell} = \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta_1)$. Since for any $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{N}_{1\ell}$, the event $|\mathbb{1}_\ell(\boldsymbol{\gamma}_1)-\mathbb{1}_\ell(\boldsymbol{\gamma}_2)| > 0$ implies that there exists $\boldsymbol{\gamma}_3 = \lambda\boldsymbol{\gamma}_1 + (1-\lambda)\boldsymbol{\gamma}_2$ with $\lambda \in (0,1)$ such that $\boldsymbol{Z}_\ell^{\mathrm{T}}\boldsymbol{\gamma}_3 = 0$, we have

$$\begin{aligned}\mathbb{E}\left\{U\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_1)-\mathbb{1}_\ell(\boldsymbol{\gamma}_2)\right|\right\} &\le \mathbb{E}_{\boldsymbol{Z}_\ell}\left\{\sup_{\boldsymbol{\gamma}_3\in\mathcal{N}_{1\ell}}\mathbb{E}(U|\boldsymbol{Z}_\ell^{\mathrm{T}}\boldsymbol{\gamma}_3=0)\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_1)-\mathbb{1}_\ell(\boldsymbol{\gamma}_2)\right|\right\}\\ &\le M_1\mathbb{E}\left(\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_1)-\mathbb{1}_\ell(\boldsymbol{\gamma}_2)\right|\right) \le c_1M_1\|\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2\|,\end{aligned}$$

where the last inequality is due to Assumption 3.(ii), which verifies (A.1).

(ii) For each $\ell = 1$ and 2, let $\mathcal{N}_{2\ell} = \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta)$. Let M_R be a positive constant such that $P_R = \mathbb{P}(\boldsymbol{Z}_\ell \in \mathcal{A}_R) > 0$, where $\mathcal{A}_R = \{\|\boldsymbol{Z}_\ell\| \le M_R, \boldsymbol{Z}_\ell \in R\}$. Then, for any $\boldsymbol{\gamma}_\ell \in \mathcal{N}_{2\ell}$, we have

$$\begin{aligned}&\mathbb{E}\left\{U\mathbb{1}(\boldsymbol{Z}_\ell \in R)\left|\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)-\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})\right|\right\}\\ =&\mathbb{E}_{\boldsymbol{Z}_\ell}[\mathbb{E}(U|\boldsymbol{Z}_\ell)\{|\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)-\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})|\,\mathbb{1}(\boldsymbol{Z}_\ell \in R)\}]\\ \ge&\mathbb{E}_{\boldsymbol{Z}_\ell}[\inf_{\boldsymbol{\gamma}_3\in\mathcal{N}_{2\ell}}\mathbb{E}(U|\boldsymbol{Z}_\ell^{\mathrm{T}}\boldsymbol{\gamma}_3=0)\{|\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)-\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})|\,\mathbb{1}(\boldsymbol{Z}_\ell \in R)\}]\\ \ge&M_0\mathbb{E}\left\{|\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)-\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})|\,\mathbb{1}(\boldsymbol{Z}_\ell \in R)\right\},\\ \ge&M_0\mathbb{E}\left\{|\mathbb{1}_\ell(\boldsymbol{\gamma}_\ell)-\mathbb{1}_\ell(\boldsymbol{\gamma}_{\ell 0})|\,\mathbb{1}(\|\boldsymbol{Z}_\ell\| \le M_R, \boldsymbol{Z}_\ell \in R)\right\}\\ =&M_0\mathbb{E}\left\{\mathbb{1}(|q_\ell| < |\boldsymbol{Z}_\ell^{\mathrm{T}}\Delta\boldsymbol{\gamma}_\ell|)\mathbb{1}(\|\boldsymbol{Z}_\ell\| \le M_R, \boldsymbol{Z}_\ell \in R)\right\}\\ =&M_0P_R\mathbb{E}\left\{\mathbb{1}(|q_\ell| < |\boldsymbol{Z}_\ell^{\mathrm{T}}\Delta\boldsymbol{\gamma}_\ell|) \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right\},\end{aligned} \tag{A.4}$$

where $\Delta\boldsymbol{\gamma}_\ell = \boldsymbol{\gamma}_\ell - \boldsymbol{\gamma}_{\ell 0}$. Take $\delta_3 = \min(\delta_2/M_R, \delta)$, where δ_2 is specified in Assumption 4.(i). Then, for any $\boldsymbol{\gamma}_\ell \in \mathcal{N}(\boldsymbol{\gamma}_{\ell 0};\delta_2)$, we have $|\boldsymbol{Z}_\ell^{\mathrm{T}}\Delta\boldsymbol{\gamma}_\ell| \le \delta_2$. Since the first elements of $\boldsymbol{\gamma}_{\ell 0}$ and $\boldsymbol{\gamma}_\ell$ are 1, $\boldsymbol{Z}_\ell^{\mathrm{T}}\Delta\boldsymbol{\gamma}_\ell = \boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\Delta\boldsymbol{\gamma}_{-1,\ell}$. Hence, by Assumption 4.(i),

$$\begin{aligned}&\mathbb{E}\left\{\mathbb{1}(|q_\ell| < |\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\Delta\boldsymbol{\gamma}_{-1,\ell}|)|\boldsymbol{Z}_\ell \in \mathcal{A}_R\right\}\\ \ge&c_2\mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\Delta\boldsymbol{\gamma}_{-1,\ell}|| \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right)\\ \ge&c_2\|\Delta\boldsymbol{\gamma}_{-1,\ell}\|\inf_{\|\boldsymbol{\gamma}_{-1}\|=1}\mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\boldsymbol{\gamma}_{-1}| \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right)\end{aligned}$$

$$=c_2\|\gamma_{\ell 0}-\gamma_\ell\| \inf_{\|\gamma_{-1}\|=1} \mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1}| \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right). \tag{A.5}$$

We next show that $\inf_{\|\gamma_{-1}\|=1} \mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1}| \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right) > 0$. If otherwise, there exists some γ_* such that $\|\gamma_{-1,*}\| = 1$ and $\mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1,*}|| \mid \boldsymbol{Z}_{-1,\ell} \in \mathcal{A}_R\right) = 0$. This means that $\mathbb{P}(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1,*}| = 0|\boldsymbol{Z}_\ell \in \mathcal{A}_R) = 1$, which further implies that $\mathbb{P}(|\boldsymbol{Z}_{-1\ell}^{\mathrm{T}}\gamma_{-1,*}| = 0) \geq \mathbb{P}(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1,*}| = 0|\boldsymbol{Z}_\ell \in \mathcal{A}_R)\mathbb{P}(\boldsymbol{Z}_\ell \in \mathcal{A}_R) = P_R$, and contradicts with Assumption 3.(ii). Therefore, it must hold that

$$\inf_{\|\gamma_{-1}\|=1} \mathbb{E}\left(|\boldsymbol{Z}_{-1,\ell}^{\mathrm{T}}\gamma_{-1}| \mid \boldsymbol{Z}_\ell \in \mathcal{A}_R\right) > 0. \tag{A.6}$$

Combining (A.4)–(A.6) completes the proof of Part (ii) of Lemma A.2.

(iii) It follows from similar arguments as in (i) and thus is omitted. □

The following moment inequalities are for partial sums, built upon Lemma A.2 and Rosenthal-type moment inequalities for mixing sequences provided in Peligrad (1982).

LEMMA A.3 (Moment inequalities). *Let $U_t = U(\boldsymbol{W}_t)$ be a function of $\boldsymbol{W}_t$. Under Assumptions 1.(i), 3.(ii) and 4.(iii), and suppose that $\sup_{\gamma\in\mathcal{N}(\gamma_{l0};\delta_l)} \mathbb{E}(|U_t|^4 \mid \boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma = 0) < M$ for almost surely $\boldsymbol{Z}_{l,t}$ for each $l = 1$ and 2, where δ_1 and M are positive constants. Then, there exist constants $c_1, c_2 > 0$ such that for each $l \in \{1, 2\}$, if $\gamma_1, \gamma_2 \in \mathcal{N}(\gamma_{l0}; \delta_l)$, then*

$$\mathbb{E}\left|\mathbb{G}_T\left[U_t\left\{\mathbb{1}_{l,t}(\gamma_1) - \mathbb{1}_{l,t}(\gamma_2)\right\}\right]\right|^4 \leq c_1\|\gamma_1 - \gamma_2\|^2 \tag{A.7}$$

and if $\gamma_1, \gamma_2 \in \mathcal{N}(\gamma_{10}; \delta_1)$ and $\gamma_3, \gamma_4 \in \mathcal{N}(\gamma_{20}; \delta_1)$, then

$$\mathbb{E}\left|\mathbb{G}_T\left[U_t\left\{\mathbb{1}_{1,t}(\gamma_1) - \mathbb{1}_{1,t}(\gamma_2)\right\}\left\{\mathbb{1}_{2,t}(\gamma_3) - \mathbb{1}_{2,t}(\gamma_4)\right\}\right]\right|^4 \leq c_2\|\gamma_1 - \gamma_2\|^2\|\gamma_3 - \gamma_4\|^2. \tag{A.8}$$

PROOF. Denote by $U_t\{\mathbb{1}_{l,t}(\gamma_1) - \mathbb{1}_{l,t}(\gamma_2)\} = \widetilde{U}_t(\gamma_1, \gamma_2)$. Then according to Lemma 3.6 of Peligrad (1982), there is a constant $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}\left|\sum_{t=1}^{T}\{\widetilde{U}_t(\gamma_1,\gamma_2) - \mathbb{E}\widetilde{U}_t(\gamma_1,\gamma_2)\}\right|^4 \\
\leq &C\left(T^2\|\widetilde{U}_t(\gamma_1,\gamma_2)\|_2^4 + T\|\widetilde{U}_t(\gamma_1,\gamma_2)\|_4^4\right),
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{E}\left|\mathbb{G}_T\{\widetilde{U}_t(\gamma_1,\gamma_2)\}\right|^4 \leq &2C[\mathbb{E}\{\widetilde{U}_t(\gamma_1,\gamma_2)\}^2]^2 \\
= &2C\{\mathbb{E}(U_t^2|\mathbb{1}_{l,t}(\gamma_1) - \mathbb{1}_{l,t}(\gamma_2)|)\}^2 \\
\leq &C'\|\gamma_1 - \gamma_2\|^2,
\end{aligned} \tag{A.9}$$

for some constant $C' > 0$, where the last inequality is from (A.1) in Lemma A.2. Therefore, (A.7) is verified. Similarly, (A.8) can be shown by using Lemma 3.6 of Peligrad (1982) and the moment inequality (A.3). □

The next lemma is a maximal inequality for empirical processes with regime indicators under the α-mixing condition.

LEMMA A.4 (Maximal inequalities). *Suppose that the conditions in Lemma A.3 hold. Then there exist constants $c_1, c_2 > 0$ such that for any λ and $\varepsilon > 0$, it holds that*

$$\mathbb{P}\left\{\sup_{\gamma_1,\gamma_2\in\mathcal{N}(\gamma_{l0};\varepsilon)} |\mathbb{G}_T\left[U_t\left\{\mathbb{1}_{l,t}(\gamma_1)-\mathbb{1}_{l,t}(\gamma_2)\right\}\right]| > \lambda\right\} \le \frac{c_1}{\lambda^2}\varepsilon^2, \ \text{for } l=1,2 \ \text{and} \tag{A.10}$$

$$\mathbb{P}\left\{\sup_{\substack{\gamma_1,\gamma_2\in\mathcal{N}(\gamma_{10};\varepsilon)\\ \gamma_3,\gamma_4\in\mathcal{N}(\gamma_{20};\varepsilon)}} |\mathbb{G}_T\left[U_t\left\{\mathbb{1}_{1,t}(\gamma_1)-\mathbb{1}_{1,t}(\gamma_2)\right\}\left\{\mathbb{1}_{2,t}(\gamma_3)-\mathbb{1}_{2,t}(\gamma_4)\right\}\right]| > \lambda\right\} \le \frac{c_2}{\lambda^4}\varepsilon^4. \tag{A.11}$$

PROOF. The first part (A.10) follows similar arguments as that in proof of Lemma I.1 of Lee et al. (2021). We now show (A.11) by adapting the proof of Lemma I.1 of Lee et al. (2021), which mainly employed Theorem 1 of Bickel and Wichura (1971).

First, by applying (A.8) of Lemma A.3, we know that for some $\delta > 0$ and any $\gamma_j, \gamma_j' \in \mathcal{N}(\gamma_{j0};\delta)$ and any $\gamma_k, \gamma_k' \in \mathcal{N}(\gamma_{k0};\delta)$,

$$\mathbb{E}\left|\mathbb{G}_T\left\{U_t\left|\mathbb{1}_{j,t}(\gamma_j)-\mathbb{1}_{j,t}(\gamma_j')\right|\left|\mathbb{1}_{k,t}(\gamma_k)-\mathbb{1}_{k,t}(\gamma_k')\right|\right\}\right|^4 \le C_1\|\gamma_j-\gamma_j'\|^2\|\gamma_k-\gamma_k'\|^2, \tag{A.12}$$

for some constant $C_1 > 0$. Let $\gamma_0 = (\gamma_{j0}^{\mathrm{T}}, \gamma_{k0}^{\mathrm{T}})^{\mathrm{T}}$, $\gamma = (\gamma_j^{\mathrm{T}}, \gamma_k^{\mathrm{T}})^{\mathrm{T}}$ and

$$J_T(\gamma) = \mathbb{G}_T\left\{U_t\left|\mathbb{1}_{j,t}(\gamma_j)-\mathbb{1}_{j,t}(\gamma_{j0})\right|\left|\mathbb{1}_{k,t}(\gamma_k)-\mathbb{1}_{k,t}(\gamma_{k0})\right|\right\}. \tag{A.13}$$

By equation (1) of Bickel and Wichura (1971),

$$\sup_{\gamma:\|\gamma-\gamma_0\|\le\varepsilon} |J_T(\gamma)| \le d\cdot M'' + |J_T(\widetilde{\gamma})|, \tag{A.14}$$

where $d = d_j + d_k$ and $\widetilde{\gamma} = \gamma_0 + \varepsilon\mathbf{1}$ is the elementwise increment of γ_0 by a positive constant ε, and the supremum is taken over a hyper-cube $\{\gamma : 0 \le \gamma_i - \gamma_{i,0} \le \varepsilon, i \in [d]\}$, and the precise definition and an upper bound of M'' are referred to Bickel and Wichura (1971). It is sufficient to show that each of M'' and $J_T(\widetilde{\gamma})$ satisfies the conclusion of the lemma since $|a| + |b| > 2c$ implies either $|a| > c$ or $|b| > c$.

To apply Theorem 1 of Bickel and Wichura (1971), we need to consider the increment of the process J_T around a block in the tube $T_\varepsilon = \{\gamma : \|\gamma - \gamma_0\| \le \varepsilon\}$. For a block $B = (\gamma_1, \gamma_2] = (\gamma_{11}, \gamma_{21}] \times \cdots \times (\gamma_{1d}, \gamma_{2d}]$ in the tube T_ε , let

$$\begin{aligned} J_T(B) &= \sum_{k_1=0,1}\cdots\sum_{k_d=0,1}(-1)^{d-k_1-\cdots k_d}J_T\left(\gamma_{11}+k_1(\gamma_{21}-\gamma_{11}),\cdots,\gamma_{1d}+k_d(\gamma_{2d}-\gamma_{1d})\right) \\ &= \sum_{k_2=0,1}\cdots\sum_{k_d=0,1}(-1)^{d-k_2-\cdots k_d}\left\{J_T\left(\gamma_{11},\gamma_{12}+k_2(\gamma_{22}-\gamma_{12})\cdots,\gamma_{1d}+k_d(\gamma_{2d}-\gamma_{1d})\right)\right. \\ &\qquad \left.-J_T\left(\gamma_{21},\gamma_{12}+k_2(\gamma_{22}-\gamma_{12})\cdots,\gamma_{1d}+k_d(\gamma_{2d}-\gamma_{1d})\right)\right\}. \end{aligned}$$

It follows from the C_r-inequality that there exists some positive constants C_2 and C_3 such that

$$\begin{aligned} \mathbb{E}\left|J_T(B)\right|^4 \le & C_2\sum_{k_2=0,1}\cdots\sum_{k_d=0,1}\mathbb{E}\{|J_T\left(\gamma_{11},\gamma_{12}+k_2(\gamma_{22}-\gamma_{12})\cdots,\gamma_{1d}+k_d(\gamma_{2d}-\gamma_{1d})\right) \\ & - J_T\left(\gamma_{21},\gamma_{12}+k_2(\gamma_{22}-\gamma_{12})\cdots,\gamma_{1d}+k_d(\gamma_{2d}-\gamma_{1d})\right)|^4\}. \end{aligned} \tag{A.15}$$

Let $\boldsymbol{\gamma}_1(\boldsymbol{\psi}) = (\gamma_{11}, \boldsymbol{\psi}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\gamma}_2(\boldsymbol{\psi}) = (\gamma_{21}, \boldsymbol{\psi}^{\mathrm{T}})^{\mathrm{T}}$, which are identical except for the first element, such that $\|\boldsymbol{\psi} - \boldsymbol{\gamma}_{-1,0}\| \le \varepsilon$. Then, (A.15) implies that

$$\mathbb{E}\,|J_T(B)|^4 \le C_3 \sup_{\boldsymbol{\psi}\in\mathcal{N}(\boldsymbol{\gamma}_{-1,0};\varepsilon)} \mathbb{E}\,|J_T\{\boldsymbol{\gamma}_1(\boldsymbol{\psi})\} - J_T\{\boldsymbol{\gamma}_2(\boldsymbol{\psi})\}|^4 \tag{A.16}$$

for some positive constant C_3. Let $\widetilde{\boldsymbol{\psi}}_k$ be the last d_k elements $\boldsymbol{\psi}$, and let $\boldsymbol{\gamma}_{1,j}(\boldsymbol{\psi})$ and $\boldsymbol{\gamma}_{2,j}(\boldsymbol{\psi})$ be the vectors of the first d_j elements of $\boldsymbol{\gamma}_1(\boldsymbol{\psi})$ and $\boldsymbol{\gamma}_2(\boldsymbol{\psi})$, respectively. Then, note that for any $\boldsymbol{\psi}$, by the triangle inequality,

$$\begin{aligned} &|J_T\{\boldsymbol{\gamma}_1(\boldsymbol{\psi})\} - J_T\{\boldsymbol{\gamma}_2(\boldsymbol{\psi})\}|^4 \\ \le &\left|\mathbb{G}_T\left\{|U_t|\,|\mathbb{1}_{j,t}(\boldsymbol{\gamma}_{1,j}(\boldsymbol{\psi})) - \mathbb{1}_{j,t}(\boldsymbol{\gamma}_{2,j}(\boldsymbol{\psi}))|\left|\mathbb{1}_{k,t}(\widetilde{\boldsymbol{\psi}}_k) - \mathbb{1}_{k,t}(\boldsymbol{\gamma}_{k0})\right|\right\}\right|^4. \end{aligned} \tag{A.17}$$

Since $\|\boldsymbol{\gamma}_{1,j}(\boldsymbol{\psi}) - \boldsymbol{\gamma}_{2,j}(\boldsymbol{\psi})\| \le |\gamma_{11} - \gamma_{21}|$ and $\|\widetilde{\boldsymbol{\psi}}_k - \boldsymbol{\gamma}_{k0}\| \le \varepsilon$ for any $\|\boldsymbol{\psi} - \boldsymbol{\gamma}_0\| \le \varepsilon$, it follows from (A.12), (A.16), and (A.17) that there exists some poistive constant C_4 such that

$$\mathbb{E}\,|J_T(B)|^p \le C_4\,|\gamma_{11} - \gamma_{21}|^2\,\varepsilon^2 \le C_5\,|\gamma_{11} - \gamma_{21}|^4,$$

where $C_5 \ge C_4\varepsilon/|\gamma_{11} - \gamma_{21}|$. Now, without loss of generality, we can assume that $\mu(B) \ge C_5\,|\gamma_{11} - \gamma_{21}|^d$, where μ denotes the Legesque measure in $\mathbb{R}^d$, since we can derive the same bound by choosing the smallest side length of B as $|\gamma_{11} - \gamma_{21}|$. This implies that $\mathbb{E}\,|J_T(B)|^4 \le C_5\,\{\mu(B)\}^{\frac{p}{d}}$ for any block $B \subset T_\varepsilon$. Therefore, we can take $\gamma_1 = \gamma_2 = 2$ and $\beta_1 = \beta_2 = \frac{2}{d}$ in the equation (3) of Bickel and Wichura (1971), implying that their equation (2) holds with $\gamma = 4$ and $\beta = \frac{4}{d}$. Since $\mu(T_\varepsilon) = \varepsilon^d$, by Theorem 1 of Bickel and Wichura (1971), we conclude that for any λ,

$$\mathbb{P}(M'' > \lambda) \le \frac{C_6}{\lambda^4}\varepsilon^4, \tag{A.18}$$

for some positive constant C_6. Furthermore, by the Markov inequality and the moment bound in (A.12), there exists some positive constant C_7 such that

$$\mathbb{P}\{J_T(\widetilde{\gamma}) > \lambda\} \le \frac{C_7}{\lambda^4}\varepsilon^4. \tag{A.19}$$

Therefore, (A.11) is proved by combining (A.14), (A.18), and (A.19). This completes the proof of Lemma A.4. □

LEMMA A.5. *Suppose that the conditions in Lemma A.3 hold. Then we have*

$$\sup_{\|\boldsymbol{\gamma}_l - \boldsymbol{\gamma}_{l0}\| \lesssim T^{-1}} \sqrt{T}\mathbb{E}_T\{U_t\,|\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l) - \mathbb{1}_{l,t}(\boldsymbol{\gamma}_{l0})|\} = o_p(1), \ \textit{for}\ l = 1, 2\ \textit{and} \tag{A.20}$$

$$\sup_{\substack{\|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_{10}\| \lesssim T^{-1} \\ \|\boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_{20}\| \lesssim T^{-1}}} T\mathbb{E}_T\{U_t\,|\mathbb{1}_{1,t}(\boldsymbol{\gamma}_1) - \mathbb{1}_{1,t}(\boldsymbol{\gamma}_{10})|\,|\mathbb{1}_{2,t}(\boldsymbol{\gamma}_2) - \mathbb{1}_{2,t}(\boldsymbol{\gamma}_{20})|\} = o_p(1). \tag{A.21}$$

PROOF. For each $l = 1$ and 2, letting $\varepsilon = cT^{-1}$ in (A.10) for some constant $c > 0$ implies that

$$\sup_{\|\boldsymbol{\gamma}_l - \boldsymbol{\gamma}_{l0}\| \lesssim T^{-1}} \sqrt{T}(\mathbb{E}_T - \mathbb{E})\{U_t\,|\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l) - \mathbb{1}_{l,t}(\boldsymbol{\gamma}_{l0})|\} = O_p\left(T^{-\frac{2}{p}}\right),$$

for $p \in (4, 4+\beta)$ with β specified in Lemma A.3. According to (A.1) in Lemma A.2,

$$\sup_{\|\boldsymbol{\gamma}_l - \boldsymbol{\gamma}_{l0}\| \lesssim T^{-1}} \sqrt{T}\mathbb{E}\{U_t\,|\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l) - \mathbb{1}_{l,t}(\boldsymbol{\gamma}_{l0})|\} = O\left(T^{-1}\right).$$

Combining the above two equalities leads to (A.20). Similarly, letting $\varepsilon = cT^{-1}$ in (A.11) for some constant $c > 0$ implies that

$$\sup_{\substack{\|\gamma_1-\gamma_{10}\|\lesssim T^{-1}\\ \|\gamma_2-\gamma_{20}\|\lesssim T^{-1}}} \sqrt{T}(\mathbb{E}_T-\mathbb{E})\left\{U_t\left|\mathbb{1}_{1,t}(\gamma_1)-\mathbb{1}_{1,t}(\gamma_{10})\right|\left|\mathbb{1}_{2,t}(\gamma_2)-\mathbb{1}_{2,t}(\gamma_{20})\right|\right\} = O_p\left(T^{-\frac{4}{p}}\right).$$

According to (A.2) in Lemma A.2 we have

$$\sup_{\substack{\|\gamma_1-\gamma_{10}\|\lesssim T^{-1}\\ \|\gamma_2-\gamma_{20}\|\lesssim T^{-1}}} \mathbb{E}\left\{U_t\left|\mathbb{1}_{1,t}(\gamma_1)-\mathbb{1}_{1,t}(\gamma_{10})\right|\left|\mathbb{1}_{2,t}(\gamma_2)-\mathbb{1}_{2,t}(\gamma_{20})\right|\right\} = O_p\left(T^{-2}\right).$$

Combining the above two equations leads to (A.21). □

LEMMA A.6. *Under the conditions of Lemma A.3, for any constatnts $\lambda, c_1, c_2 > 0$ and $j \neq k \in \{1,\cdots,4\}$, we have*

$$\sup_{c_1T^{-1}\le\|\gamma-\gamma_0\|\le c_2}\left\{\left|(\mathbb{E}_T-\mathbb{E})\left(U_t\mathbb{1}_t^{(j)}(\gamma_0)\mathbb{1}_t^{(k)}(\gamma)\right)\right|-\lambda\|\gamma-\gamma_0\|\right\} = O_p(T^{-1}). \quad \text{(A.22)}$$

PROOF. The event that $j \neq k$ can be classed into two cases: (i) $(j,k) \in \mathcal{S}(i)$ for $i = 1$ or 2; and (ii) $(j,k) \notin \mathcal{S}(i)$ for both $i = 1$ and 2.

Case (i): $(j,k) \in \mathcal{S}(i)$ for $i \in \{1,2\}$. Without loss of generality, we take $j = 1, k = 2$ to illustrate. Note that

$$\begin{aligned}\mathbb{1}_t^{(1)}(\gamma_0)\mathbb{1}_t^{(2)}(\gamma) =&\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\mathbb{1}(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20} > 0, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_2 > 0)\\ =&\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\left\{\mathbb{1}(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20} > 0) - \mathbb{1}(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_2 \le 0 < \boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20})\right\},\end{aligned}$$

which implies that

$$\begin{aligned}&\left|(\mathbb{E}_T-\mathbb{E})\left\{U_t\mathbb{1}_t^{(1)}(\gamma_0)\mathbb{1}_t^{(2)}(\gamma)\right\}\right| \le \left|(\mathbb{E}_T-\mathbb{E})\left\{\tilde{U}_t\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\right\}\right|\\ &+\left|(\mathbb{E}_T-\mathbb{E})\left\{U_t\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\mathbb{1}(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_2 \le 0 < \boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20})\right\}\right| =: I_{1,T}(\gamma) + I_{2,T}(\gamma), \quad \text{say},\end{aligned}$$

where $\tilde{U}_t = U_t\mathbb{1}\left(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20} > 0\right)$. Define the "shells"

$$S_{T,j} = \left\{\gamma : c_1jT^{-1} \le \|\gamma-\gamma_0\| < c_1(j+1)T^{-1}\right\}.$$

Then, for any $M > 0$, we have

$$\begin{aligned}&\mathbb{P}\left(\sup_{c_1T^{-1}\le\|\gamma-\gamma_0\|\le c_2} T\left\{I_{1,T}(\gamma) - \lambda\|\gamma-\gamma_0\|/2\right\} > M\right)\\ \le&\sum_{j=1}^{\infty}\mathbb{P}\left\{\gamma\in S_{T,j}, \left|(\mathbb{E}_T-\mathbb{E})\left\{\tilde{U}_t\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\right\}\right| > MT^{-1} + \lambda\|\gamma-\gamma_0\|/2\right\}\\ \le&\sum_{j=1}^{\infty}\mathbb{P}\left\{\gamma\in S_{T,j}, \left|(\mathbb{E}_T-\mathbb{E})\left\{\tilde{U}_t\mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\right\}\right| > (M + c_1j\lambda/2)T^{-1}\right\}\\ \le&\sum_{j=1}^{\infty}\frac{c_3(j+1)^2}{(M+c_1j\lambda/2)^4} = O\left(\frac{1}{M^4}\right), \end{aligned} \quad \text{(A.23)}$$

where the last inequality is by invoking (A.10) in Lemma A.4. Via the similar argument, we obtain

$$\mathbb{P}\left(\sup_{c_1T^{-1}\le\|\gamma-\gamma_0\|\le c_2} T\left\{I_{2,T}(\gamma) - \lambda\|\gamma-\gamma_0\|/2\right\} > M\right) = O\left(\frac{1}{T^2M^4}\right). \quad \text{(A.24)}$$

This together with (A.23) verifies (A.22).

Case (ii): $(j,k)\notin\mathcal{S}(i)$ for either $i=1$ or 2. Without loss of generality, we take $j=1,k=3$ to illustrate. Then

$$\mathbb{1}_t^{(1)}(\gamma_0)\mathbb{1}_t^{(3)}(\gamma) = \mathbb{1}(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_1 \le 0 < \boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10})\mathbb{1}(\boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_2 \le 0 < \boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20}). \tag{A.25}$$

Therefore, $\left|(\mathbb{E}_T-\mathbb{E})\left\{U_t\mathbb{1}_t^{(1)}(\gamma_0)\mathbb{1}_t^{(3)}(\gamma)\right\}\right| = I_2(\gamma)$ and the result follows from (A.24). Combining the two cases completes the proof for the lemma. □

A.2. Lemmas for Poisson point processes. We first introduce some basic notations for the point measures and point processes following the definitions in Resnick (2008).

DEFINITION A.2 (Point measures). Suppose that E is a locally compact space with a countable basis whose Borel σ-algebra of subsets is $\mathcal{E}$. A *point process* on E is a measure m of the following form: for $\{\mathbf{x}_i, i\ge 1\}$, which is a countable collection of points of E, and any Borel set $A\in\mathcal{E}$, $m(A) := \sum_{i=1}\mathbb{1}(\mathbf{x}_i\in A)$. If $m(K)<\infty$ for any compact set $K\in\mathcal{E}$, then m is said to be Radon. Let $M_p(E)$ be the space of all Radon point measures on E. A sequence $\{m_n\}\subset M_p(E)$ is said to converge vaguely to m, if $\int_E f dm_n \to \int_E f dm$ as $n\to\infty$ for all $f\in C_K(E)$, the continuous function space with compact support K. The vague convergence induces a vague topology on $M_p(E)$. Topological space $M_p(E)$ is then metrizable as a complete separable metric space. Define $\mathcal{M}_p(E)$ as the σ-algebra generated by open sets in $M_p(E)$.

DEFINITION A.3 (Point processes and their weak convergence). A *point process* on E is a measurable map from a probability space $(\Omega,\mathcal{A},\mathbb{P})\to(M_p(E),\mathcal{M}_p(E))$, i.e., for every event $\omega\in\Omega$, the realization of the point process $\mathbf{N}(\omega)$ is some point measure in $M_p(E)$. A sequence of point processes $\mathbf{N}_n$ *weak converges* of a point process $\mathbf{N}$, denoted as $\mathbf{N}_n\Rightarrow\mathbf{N}$ if $\mathbb{E}_{\mathbb{P}}\{h(\mathbf{N}_n)\}\to\mathbb{E}_{\mathbb{P}}\{h(\mathbf{N})\}$ for all continuous and bounded functions h mapping $M_p(E)$ to $\mathbb{R}$. Note that if $\mathbf{N}_n\Rightarrow\mathbf{N}$ then $\int_E f(\mathbf{x})d\mathbf{N}_n(\boldsymbol{X}) \xrightarrow{d} \int_E f(\mathbf{x})d\mathbf{N}(\boldsymbol{X})$ for any $f\in C_K(E)$ by the continuous mapping theorem.

DEFINITION A.4 (Poisson point process). A point process $\mathbf{N}$ is called a *Poisson process measure* (PRM) with mean measure μ if $\mathbf{N}$ satisfies

(i) for any $F\in\mathcal{E}$ and any non-negative integer k, $\mathbb{P}(\mathbf{N}(F)=k)=\exp\{-\mu(F)\}\{\mu(F)\}^k/k!$ if $\mu(F)<\infty$ and $\mathbb{P}(\mathbf{N}(F)=k)=0$ if $\mu(F)=\infty$;

(ii) if $F_1,\cdots,F_k$ are mutually disjoint sets in $\mathcal{E}$, then $\{\mathbf{N}(F_i), i\le k\}$ are independent random variables.

The following two lemmas, from Proposition 3.22 of Resnick (2008) and Theorem 1 of Meyer (1973), respectively, provide key tools to study the weak convergence of point processes of extreme events with α-mixing time series.

LEMMA A.7 (Kallenberg's theorem). *Suppose that* $\mathbf{N}$ *is a point process on* E *and* $\mathcal{T}$ *is a basis of relatively compact open sets such that* $\mathcal{T}$ *is closed under finite unions and intersections, and for any* $F\in\mathcal{T}$, $\mathbb{P}\{\mathbf{N}(\partial F)=0\}=1$. *Then* $\widehat{\mathbf{N}}_T\Rightarrow\mathbf{N}$ *if for all* $F\in\mathcal{T}$,

$$\lim_{T\to\infty}\mathbb{P}\left\{\widehat{\mathbf{N}}_T(F)=0\right\}=\mathbb{P}\{\mathbf{N}(F)=0\}, \quad \textit{and} \tag{A.26}$$

$$\lim_{T\to\infty}\mathbb{E}\left\{\widehat{\mathbf{N}}_T(F)\right\}=\mathbb{E}\{\mathbf{N}(F)\}<\infty. \tag{A.27}$$

LEMMA A.8 (Meyer's theorem). *Suppose that the sequence* $\{A_t^n\}_{t=1}^n\,(n=1,2,\cdots)$ *is stationary and* α*-mixing with mixing coefficient* $\alpha_n(k)$ *defined as*

$$\alpha_n(k)=\sup_{\substack{E\in\Omega_1^m,\\ F\in\Omega_{m+k+1}^n}}|\mathbb{P}(EF)-\mathbb{P}(E)\mathbb{P}(F)|\,,\ \textit{where}\ \Omega_j^J=\sigma(A_j^n,\cdots,A_J^n)\ 1\le j<J\le n$$

for any $1\le k\le n$*. Suppose that the probability of the event* A_t^n *is* $\mathbb{P}(A_t^n)=\frac{a}{n}+o(\frac{1}{n})$ *for some* $a>0$*. Moreover, suppose that the following conditions hold: there exist sequences of block sizes* $\{p_m,m\ge1\}$*,* $\{q_m,m\ge1\}$ *and* $\{t_m=m(p_m+q_m),m\ge1\}$ *such that*
(a) for any $r>0$*,* $m^r\alpha_{t_m}(q_m)\to0$ *as* $m\to\infty$*, where* $t_m=m(p_m+q_m)$*,*
(b) $q_m/p_m\to0, p_{m+1}/p_m\to1$ *as* $m\to\infty$*, and*
(c) $I_{p_m}=\sum_{i=1}^{p_m-i}(p_m-i)\mathbb{P}(A_1^{t_m}\cap A_{i+1}^{t_m})=o(\frac{1}{m})$ *as* $m\to\infty$*.*
Then it holds that

$$\mathbb{P}\left(\textit{exactly } k \textit{ events among } \{A_t^n\}_{t=1}^n \textit{ happen}\right)\to\frac{e^{-a}a^k}{k!}\quad \textit{as } n\to\infty.$$

Remark. (i) Note that for any given $n<\infty$, the α-mixing coefficient $\alpha_n(k)$ defined above is upper bounded by the commonly used α-mixing coefficient $\alpha(k)$ (see e.g., Doukhan, 1995), where the supreme of F is taken over Ω_{m+k+1}^{∞} instead of Ω_{m+k+1}^n. (ii) The proof of the above theorem is based on partitioning the observations into consecutive blocks of size p_m and q_m alternately. The condition $I_{p_m}=o(1/m)$ prevents clusters of rare events A_t^n, preventing the compound Poisson processes as the limit.

A.3. Lemmas for epi-convergence. In the investigation of the limiting distribution of $\widehat{\gamma}$ and $\widehat{\boldsymbol{\beta}}$, we will employ the tool of epi-convergence in distribution (Knight, 1999), which is useful in establishing weak convergences of "argmin" functionals, and is more general than uniform convergence, because it allows for more general discontinuity.

DEFINITION A.5 (Epi-convergence in distribution). Suppose that $\{Q_n(\mathbf{x})\}$ is a sequence of random lower semi-continuous (l-sc) functions, namely $Q_n(\mathbf{x})\le\liminf_{\mathbf{x}_j\to x}Q_n(\mathbf{x}_j)$ for any $\mathbf{x}$ and any sequence $\{\mathbf{x}_j\}$ whose limit is $\mathbf{x}$. Let $\mathcal{L}$ be the space of l-sc functions $f:\mathbb{R}^d\to\bar{\mathbb{R}}$, where $\bar{\mathbb{R}}=[-\infty,\infty]$. The space $\mathcal{L}$ can be made into a complete separable metric space. (Rockafellar and Wets, 1998).

A sequence of functions $\{Q_n\}\in\mathcal{L}$ is said to epi-converge in distribution to Q if for any closed rectangles $R_1,\cdots,R_k$ in $\mathbb{R}^d$ with open interiors $R_1^\circ,\cdots,R_k^\circ$, and any real numbers $r_1,\cdots,r_k$:

$$\begin{aligned}\mathbb{P}(\cap_{j=1}^k\{\inf_{\mathbf{x}\in R_j}Q(\mathbf{x})>r_j\})\le&\liminf_{n\to\infty}\mathbb{P}(\cap_{j=1}^k\{\inf_{\mathbf{x}\in R_j}Q_n(\mathbf{x})>r_j\})\\ \le&\limsup_{n\to\infty}\mathbb{P}(\cap_{j=1}^k\{\inf_{\mathbf{x}\in R_j^\circ}Q_n(\mathbf{x})>r_j\})\\ \le&\mathbb{P}(\cap_{j=1}^k\{\inf_{\mathbf{x}\in R_j^\circ}Q(\mathbf{x})>r_j\}).\end{aligned}$$

The above definition of the epi-convergence can be difficult to verify. Instead, we will use an equivalent characterization given by Knight (1999), using the finite-dimensional convergence and stochastic equi-lower-semicontinuouity.

DEFINITION A.6 (Finite-dimensional convergence in distribution). A sequence of random functions $\{Q_n(\mathbf{x})\}$ converges to $Q(\mathbf{x})$ in distribution in the finite-dimensional sense if for any finite positive integer k and any $(\mathbf{x}_1, \cdots, \mathbf{x}_k)$, it holds that

$$(Q_n(\mathbf{x}_1), \cdots, Q_n(\mathbf{x}_k)) \xrightarrow{d} (Q(\mathbf{x}_1), \cdots, Q(\mathbf{x}_k)).$$

DEFINITION A.7 (Stochastic equi-lower-semicontinuous). A sequence $\{Q_n\} \in \mathcal{L}$, where $\mathcal{L}$ is the space of l-sc functions defined in Definition A.5, is said to be stochastic equi-lower-semicontinuous (s.e-l-sc), if for any compact set B and any $\epsilon, \delta > 0$, there exists $\mathbf{x}_1, \cdots, \mathbf{x}_k \in B$, for a finite integer k, and some open sets $\{V(\mathbf{x}_i)\}_{i=1}^k$ covering B and containing $\mathbf{x}_1, \cdots, \mathbf{x}_k$, such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\cup_{j=1}^k \left\{\inf_{\mathbf{x}\in V(\mathbf{x}_j)} Q_n(\mathbf{x}) \le \min(\epsilon^{-1}, Q_n(\mathbf{x}_j) - \epsilon)\right\}\right) < \delta.$$

LEMMA A.9 (Theorem 2 of Knight, 1999). *Let $\{Q_n\}$ be a stochastic e-l-sc sequence of functions. Then $\{Q_n\}$ converges to Q in distribution in the finite-dimensional sense if and only if $\{Q_n\}$ epi-converges in distribution to Q.*

APPENDIX B: PROOFS FOR SECTION 3

B.1. Proof of Proposition 1. The following proof is for Proposition 1 on the identification of $\boldsymbol{\theta}_0$.

PROOF. Note that $\mathbb{M}(\boldsymbol{\theta})$ can be expanded as

$$\begin{aligned}
\mathbb{M}(\boldsymbol{\theta}) &= \mathbb{E}\{m(\boldsymbol{W}, \boldsymbol{\theta})\} \\
&= \mathbb{E}(\varepsilon^2) + \mathbb{E}[\sum_{k=1}^4 \sum_{h=1}^4 \{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_h - \boldsymbol{\beta}_{k0})\}^2 \mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)\mathbb{1}^{(h)}(\boldsymbol{\gamma})] \\
&\quad + 2\mathbb{E}\{\sum_{k=1}^4 \sum_{h=1}^4 \varepsilon \boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_h - \boldsymbol{\beta}_{k0})\mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)\mathbb{1}^{(h)}(\boldsymbol{\gamma})\} \\
&= \mathbb{E}(\varepsilon^2) + \sum_{k=1}^4 \sum_{h=1}^4 \mathbb{E}[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_h - \boldsymbol{\beta}_{k0})\}^2 \mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)\mathbb{1}^{(h)}(\boldsymbol{\gamma})] \\
&= \mathbb{M}(\boldsymbol{\theta}_0) + \sum_{k=1}^4 \sum_{h=1}^4 A_{k,h}(\boldsymbol{\theta}), \quad \text{say,}
\end{aligned} \tag{B.1}$$

where the second equality is because of $\mathbb{E}(\varepsilon | \boldsymbol{X}, \boldsymbol{Z}) = 0$. If $\boldsymbol{\theta} \ne \boldsymbol{\theta}_0$, then one of the following two cases will hold: (1): $\boldsymbol{\gamma} \ne \boldsymbol{\gamma}_0$, or (2): $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ while $\boldsymbol{\beta} \ne \boldsymbol{\beta}_0$. We now consider the two cases respectively.

Case (1). Suppose that $\boldsymbol{\gamma} \ne \boldsymbol{\gamma}_0$. Then for some $l \in \{1, 2\}$ and $h \in \{1, \cdots, 4\}$, the true splitting hyperplane $H_{l0} : \boldsymbol{z}_l^{\mathrm{T}}\gamma_{l0} = 0$ will partition through $R_h(\boldsymbol{\gamma})$. Because Assumption 2 (i) implies that $\mathbb{P}\{|q_l| < \epsilon | \boldsymbol{Z}_{-1,l}\} > 0$ almost surely for any $\epsilon > 0$, meaning there is a positive probability that $\boldsymbol{Z}$ will locate around the neighborhood of the hyperplane $\boldsymbol{z}_l^{\mathrm{T}}\gamma_{l0} = 0$, we have that for some $(k, j) \in \mathcal{S}(l)$, it holds that $\mathbb{P}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0) \cap R_h(\boldsymbol{\gamma})\} > 0$ and $\mathbb{P}\{\boldsymbol{Z} \in R_j(\boldsymbol{\gamma}_0) \cap R_h(\boldsymbol{\gamma})\} > 0$. Therefore,

$$A_{k,h}(\boldsymbol{\theta}) \ge \lambda_0 \|\boldsymbol{\beta}_h - \boldsymbol{\beta}_{k0}\|^2, \;\; A_{j,h}(\boldsymbol{\theta}) \ge \lambda_0 \|\boldsymbol{\beta}_h - \boldsymbol{\beta}_{j0}\|^2$$

according to Assumption 2 (ii). Since $\boldsymbol{\beta}_{k0} \neq \boldsymbol{\beta}_{j0}$, either $A_{k,h}(\boldsymbol{\theta}) > 0$ or $A_{j,h}(\boldsymbol{\theta}) > 0$. Consequently, $\mathbb{M}(\boldsymbol{\theta}) \geq \mathbb{M}(\boldsymbol{\theta}_0) + A_{k,h}(\boldsymbol{\theta}) + A_{j,h}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_0)$.

Case (2). Suppose that $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ while $\boldsymbol{\beta}_{k0} \neq \boldsymbol{\beta}_k$ for some $k \in \{1, \cdots, 4\}$. In such a case,

$$A_{k,k}(\boldsymbol{\theta}) = \mathbb{E}\left[\{\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k0})\}^2 \mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma}_0)\}\right] \geq \lambda_0 \|\boldsymbol{\beta}_k - \boldsymbol{\beta}_{k0}\|^2 > 0,$$

by Assumption 2 (ii). Therefore, $\mathbb{M}(\boldsymbol{\theta}) \geq \mathbb{M}(\boldsymbol{\theta}_0) + A_{k,k}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_0)$. Combining the two cases yields that $\mathbb{M}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_0)$ if $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, which completes the proof. □

B.2. Proof of Theorem 3.1. The following proof is for Theorem 3.1 on the consistency of $\widehat{\boldsymbol{\theta}}$.

PROOF. The consistency of $\widehat{\boldsymbol{\theta}}$ follows the standard approach for M-estimation (van der Vaart, 1998). First, we strengthen the result of Proposition 3.1 by a separable condition (B.2), which can be induced by the continuity of $\mathbb{M}(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$. Note that $\mathbb{M}(\boldsymbol{\theta}) = \mathbb{E}(Y^2) - 2\sum_{k=1}^4 \mathbb{E}\{Y\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_k \mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}))\} + \sum_{k=1}^4 \mathbb{E}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_k)^2 \mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}))\}$. The continuity with respect to $\boldsymbol{\beta}$ is obvious and it remains to show the continuity at $\boldsymbol{\gamma}_0$. Note that for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$,

$$\begin{aligned}
&\left|\mathbb{E}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^2 \mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}))\} - \mathbb{E}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^2 \mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0))\}\right| \\
\leq & \mathbb{E}^{1/2}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^4\} |\mathbb{E}\{\mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}))\} - \mathbb{E}\{\mathbb{1}(\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0))\}|^{1/2} \\
\leq & \mathbb{E}^{1/2}\{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^4\}\{\sum_{l=1}^2 |\mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\boldsymbol{\gamma}_l < 0) - \mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\boldsymbol{\gamma}_{l0} < 0)|\}^{1/2} \lesssim \sqrt{\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|},
\end{aligned}$$

where the last inequality is due to Assumption 3.(ii). Thus, $\mathbb{M}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}_0$, implying that

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \epsilon} \mathbb{M}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_0) \ \ \forall \epsilon > 0. \tag{B.2}$$

As a direct consequence of Lemma A.1. we have the following uniform convergence

$$\sup_{\boldsymbol{\theta} \in \Theta} |\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}_T(\boldsymbol{\theta})| \xrightarrow{P} 0, \tag{B.3}$$

as $T \to \infty$. By the definition of $\widehat{\boldsymbol{\theta}}$, we have $\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) \leq \mathbb{M}_T(\boldsymbol{\theta}_0) + o_p(1)$. Because (B.3) implies that $\mathbb{M}_T(\boldsymbol{\theta}_0) \xrightarrow{P} \mathbb{M}(\boldsymbol{\theta}_0)$. It follows that $\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) \leq \mathbb{M}(\boldsymbol{\theta}_0) + o_p(1)$, whence

$$\begin{aligned}
\mathbb{M}(\widehat{\boldsymbol{\theta}}) - \mathbb{M}(\boldsymbol{\theta}_0) \leq & \mathbb{M}(\widehat{\boldsymbol{\theta}}) - \mathbb{M}_T(\widehat{\boldsymbol{\theta}}) + o_p(1) \\
\leq & \sup_{\boldsymbol{\theta} \in \Theta} |\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}_T(\boldsymbol{\theta})| + o_p(1) \xrightarrow{P} 0.
\end{aligned} \tag{B.4}$$

Because of (B.2), for any $\epsilon > 0$, there exists $\eta > 0$ such that $\mathbb{M}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_0) + \eta$ if $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \epsilon$. Thus, the event $\left\{\|\widehat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_0\| > \epsilon\right\}$ is contained in the event $\left\{\mathbb{M}(\widehat{\boldsymbol{\theta}}) > \mathbb{M}(\boldsymbol{\theta}_0) + \eta\right\}$, whose probability converges to 0 in view of (B.4), which completes the proof for $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \xrightarrow{P} 0$ as $T \to \infty$. □

B.3. Proof of Corollary 3.1.

PROOF. Let $\mathcal{D}_T = \{\boldsymbol{W}_t\}_{t=1}^T$. We prove the corollary for $k=1$ without loss of generality, where $R_1(\gamma_0) = \{\boldsymbol{z}_l^{\mathrm{T}}\gamma_{l0} > 0, l = 1 \text{ and } 2\}$. Then $R_1(\gamma_0) \setminus R_1(\gamma)$ is a subset of $\cup_{l=1}^2 \{\boldsymbol{z} : \boldsymbol{z}_l^{\mathrm{T}}\gamma_{l0} > 0 > \boldsymbol{z}_l^{\mathrm{T}}\gamma_l\}$. Therefore,

$$\begin{aligned}\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \setminus R_1(\widehat{\gamma}) | \mathcal{D}_T\} &\leq \sum_{l=1}^2 \mathbb{P}\left(\boldsymbol{Z}_l^{\mathrm{T}}\gamma_{l0} > 0 > \boldsymbol{Z}_l^{\mathrm{T}}\widehat{\gamma}_l | \mathcal{D}_T\right) \\ &\leq c_1 \sum_{l=1}^2 \|\widehat{\gamma}_l - \gamma_{l0}\|, \end{aligned} \tag{B.5}$$

where the probability is taken over $\boldsymbol{Z}$, and the second inequality is due to the consistency of $\widehat{\gamma}$ and Assumption 3.(ii). Therefore, $\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \setminus R_1(\widehat{\gamma}) | \mathcal{D}_T\} \to 0$ as $T \to \infty$. Similarly, we have $\mathbb{P}\{\boldsymbol{Z} \in R_1(\widehat{\gamma}) \setminus R_1(\gamma_0) | \mathcal{D}_T\} \to 0$. Since $\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \triangle R_1(\widehat{\gamma}) | \mathcal{D}_T\} = \mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \setminus R_1(\widehat{\gamma}) | \mathcal{D}_T\} + \mathbb{P}\{\boldsymbol{Z} \in R_1(\widehat{\gamma}) \setminus R_1(\gamma_0) | \mathcal{D}_T\}$, we obtain

$$\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \triangle R_1(\widehat{\gamma}) | \mathcal{D}_T\} \xrightarrow{P} 0$$

as $T \to \infty$. Because $\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \triangle R_1(\widehat{\gamma}) | \mathcal{D}_T\}$ is uniformly integrable, we have

$$\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \triangle R_1(\widehat{\gamma})\} = \mathbb{E}_{\mathcal{D}_T}\left[\mathbb{P}\{\boldsymbol{Z} \in R_1(\gamma_0) \triangle R_1(\widehat{\gamma}) | \mathcal{D}_T\}\right] \to 0,$$

which completes the proof. □

B.4. Proof of Theorem 3.2. The following proof is for Theorem 3.2 on the convergence rate of $\widehat{\boldsymbol{\theta}}$.

PROOF. The convergence rate will be derived in two steps. In the first step, we establish that there is a metric d such that

$$d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \lesssim \mathbb{E}\{m(\boldsymbol{W}_t, \boldsymbol{\theta}) - m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\} \text{ for any } \boldsymbol{\theta} \in \mathcal{N}(\boldsymbol{\theta}_0; \delta_0), \tag{B.6}$$

for some $\delta_0 > 0$. In the second step, we derive a convergence rate of $\mathbb{E}\{m(\boldsymbol{W}_t, \widehat{\boldsymbol{\theta}}) - m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\}$ by bounding $(\mathbb{E}_T - \mathbb{E})\{m(\boldsymbol{W}_t, \widehat{\boldsymbol{\theta}}) - m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\}$, which combined with Step 1 will lead to the desired convergence rate of $\widehat{\boldsymbol{\theta}}$.

Step 1. Note that we can decompose $\mathbb{E}\{m(\boldsymbol{W}_t, \boldsymbol{\theta}) - m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\}$ as

$$\begin{aligned} &\mathbb{E}\{m(\boldsymbol{W}_t, \boldsymbol{\theta}) - m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\} \\ &= \sum_{j=1}^4 \mathbb{E}\left\{\left(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j)\right)^2 \mathbb{1}_t^{(j)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma)\right\} \\ &\quad + \sum_{i=1}^4 \sum_{k\neq i}^4 \mathbb{E}\left\{\left(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_k)\right)^2 \mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(k)}(\gamma)\right\}, \\ &=: \sum_{j=1}^4 J_j(\boldsymbol{\theta}) + \sum_{i=1}^4 \sum_{k\neq i}^4 G_{ik}(\boldsymbol{\theta}), \quad \text{say,} \end{aligned} \tag{B.7}$$

where the $J_j(\boldsymbol{\theta})$ term corresponds to the part of observations which are classified to the jth region under both the hyperplanes with coefficient $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}$, and the G_{ik} term corresponds of the part of observations which are classified to the ith region under the hyperplanes with coefficient $\boldsymbol{\gamma}_0$, but classified to the kth region under $\boldsymbol{\gamma}$.

First, for each $j \in \{1, \cdots, 4\}$, note that

$$\begin{aligned}\mathbb{P}\{\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0) \cap R_j(\boldsymbol{\gamma})\} =& \mathbb{P}\{\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0)\} - \mathbb{P}\{\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0) \setminus R_j(\boldsymbol{\gamma})\} \\ \overset{(i)}{\geq} & \mathbb{P}\{\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0)\} - c_0\|\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}\| \geq \mathbb{P}\{\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0)\} - c_0\delta \\ \overset{(ii)}{\geq} & \mathbb{P}(\boldsymbol{Z}_t \in R_j(\boldsymbol{\gamma}_0))/2 > 0, \end{aligned} \tag{B.8}$$

uniformly for any $\boldsymbol{\gamma} \in \mathcal{N}(\boldsymbol{\gamma}_0; \delta)$, where (i) is due to (B.5) and (ii) is by taking δ sufficiently small, which is legitimate because of the consistency of $\widehat{\boldsymbol{\gamma}}$. Then by Assumption 2.(ii),

$$J_j(\boldsymbol{\theta}) \geq c_1\|\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j\|^2. \tag{B.9}$$

For each $l \in \{1, 2\}$, we choose one pair $(i_l, k_l) \in \mathcal{S}(l)$. Without loss of generality, let $i_1 = 1, k_1 = 2, i_2 = 1, k_2 = 3$. We now bound the term $G_{i_l k_l}(\boldsymbol{\theta})$ from below,

$$\begin{aligned}G_{i_l k_l}(\boldsymbol{\theta}) =& \mathbb{E}\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{k_l}))^2\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\} \\ =& \mathbb{E}\left\{(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\} + \mathbb{E}\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}))^2\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\} \\ &+ 2\mathbb{E}\left\{\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0}\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l})\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\} \\ \geq& \mathbb{E}\left\{(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\} \\ &- 2\mathbb{E}\left\{|\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0}|\, |\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l})|\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma})\right\}.\end{aligned}$$

Similarly,

$$\begin{aligned}G_{k_l i_l}(\boldsymbol{\theta}) \geq& \mathbb{E}\left\{(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma})\right\} \\ &- 2\mathbb{E}\left\{|\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0}|\, |\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l})|\, \mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma})\right\}.\end{aligned}$$

Let $g_t^{i_l k_l} = (\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}\{\boldsymbol{Z}_t \in R_{i_l}(\boldsymbol{\gamma}_0) \cup R_{k_l}(\boldsymbol{\gamma}_0)\}$. Then

$$(\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma}) + (\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0})^2\, \mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}) = g_t^{i_l k_l}\, |\mathbb{1}_{l,t}(\boldsymbol{\gamma}_{l0}) - \mathbb{1}_{l,t}(\boldsymbol{\gamma}_l)|,$$

whose expectation is bounded by

$$\mathbb{E}\left\{g_t^{i_l k_l}\, |\mathbb{1}_{l,t}(\boldsymbol{\gamma}_{l0}) - \mathbb{1}_{l,t}(\boldsymbol{\gamma}_l)|\right\} \geq c_3\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|, \tag{B.10}$$

for some constants $c_3 > 0$ due to Assumption 4 (ii) and Lemma A.2 (ii).

For the second term of the lower bound of $G_{i_l k_l}(\boldsymbol{\theta})$, note that there exists a positive constant c_4 such that

$$\begin{aligned}&\mathbb{E}\left\{|\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0}|\, |\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l})|\, \mathbb{1}_t^{(i_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_{l0})\right\} \\ &\leq \|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\|\|\boldsymbol{\delta}_{i_l k_l,0}\|\mathbb{E}\left(\|\boldsymbol{X}_t\|^2\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_{l0})\right) \\ &\leq c_4\|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\|\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|,\end{aligned} \tag{B.11}$$

where the first inequality follows from the Cauchy-Schwartz inequality and the second is implied by Lemma A.2. Similarly,

$$\begin{aligned}&\mathbb{E}\left\{|\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{i_l k_l,0}|\, |\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l})|\, \mathbb{1}_t^{(k_l)}(\boldsymbol{\gamma}_0)\mathbb{1}_{l,t}(\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_{l0})\right\} \\ \leq& c_4\|\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l}\|\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|.\end{aligned} \tag{B.12}$$

Combining (B.10)–(B.12) leads to an lower bound of $G_{i_l k_l}(\boldsymbol{\theta}) + G_{k_l i_l}(\boldsymbol{\theta})$. For each given $l = 1$ and 2, these together with (B.9) with $j = k_l$ and i_l lead to

$$\begin{aligned}
&\{J_{k_l}(\boldsymbol{\theta}) + J_{i_l}(\boldsymbol{\theta})\}/2 + G_{i_l k_l}(\boldsymbol{\theta}) + G_{k_l i_l}(\boldsymbol{\theta}) \\
\geq & c_1(\|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\|^2 + \|\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l}\|^2)/2 + c_3\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \\
& - 2c_4\left(\|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\| + \|\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l}\|\right)\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \\
= & \sum_{j_l \in \{i_l, k_l\}} \left(\frac{c_1}{2}\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|^2 + \frac{c_3}{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| - 2c_4\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|\right) \\
= & \sum_{j_l \in \{i_l, k_l\}} L_j^l, \quad \text{say.}
\end{aligned} \tag{B.13}$$

A lower bound for the term L_j^l can be derived by considering the following two cases.
(i) If $c_1\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\| \geq 8c_4\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|$, then

$$L_j^l \geq \frac{c_1}{4}\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|^2 + \frac{c_3}{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|.$$

(ii) If $c_1\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\| < 8c_4\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|$, then

$$\frac{c_3}{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| - 2c_4\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \geq \frac{c_3}{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| - 16\frac{c_4^2}{c_1} \cdot \|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|^2,$$

which can be further bounded from below by $c_3\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|/4$ provided that $\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \leq c_1 c_3/(64c_4^2)$, which is ensured by the consistency of $\widehat{\boldsymbol{\gamma}}$. Therefore, in the case (ii),

$$L_j^l \geq \frac{c_1}{2}\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|^2 + \frac{c_3}{4}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|,$$

provided that $\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \leq c_1 c_3/(64c_4^2)$. Combining Cases (i) and (ii), we have

$$L_j^l \geq c_5(\|\boldsymbol{\beta}_{j_l 0} - \boldsymbol{\beta}_{j_l}\|^2 + \frac{1}{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|),$$

for some generic constant $c_5 > 0$, as long as $\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\| \leq c_4^2/(32c_1)$. By (B.13) we have

$$\begin{aligned}
&\{J_{k_l}(\boldsymbol{\theta}) + J_{i_l}(\boldsymbol{\theta})\}/2 + G_{i_l k_l}(\boldsymbol{\theta}) + G_{k_l i_l}(\boldsymbol{\theta}) \\
\geq & c_5(\|\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l}\|^2 + \|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\|^2 + \|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|),
\end{aligned} \tag{B.14}$$

for some positive constant c_5. Divide the regime index set $\{1, \cdots, 4\}$ to two parts: $\mathcal{K}_1 = \{k_l, i_l : l \in \{1, 2\}\}$ and $\mathcal{K}_2 = \{1, \cdots, 4\}/\mathcal{K}_1$. Then from (B.7), (B.9) and (B.14),

$$\begin{aligned}
\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta}_0) \geq & \sum_{k \in \mathcal{K}_1} J_j(\boldsymbol{\theta}) + \sum_{k \in \mathcal{K}_2} J_j(\boldsymbol{\theta}) + \sum_{i=1}^{K}\sum_{k \neq i}^{K} G_{ik}(\boldsymbol{\theta}) \\
\geq & \sum_{l=1}^{2}\{G_{i_l k_l}(\boldsymbol{\theta}) + G_{k_l i_l}(\boldsymbol{\theta}) + \frac{J_{k_l}(\boldsymbol{\theta}) + J_{i_l}(\boldsymbol{\theta})}{2}\} + \sum_{k \in \mathcal{K}_2} J_j(\boldsymbol{\theta}) \\
\geq & c_5\sum_{l=1}^{2}(\|\boldsymbol{\beta}_{i_l 0} - \boldsymbol{\beta}_{i_l}\|^2 + \|\boldsymbol{\beta}_{k_l 0} - \boldsymbol{\beta}_{k_l}\|^2 + \|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|) + c_1\sum_{k \in \mathcal{K}_2}\|\boldsymbol{\beta}_{k0} - \boldsymbol{\beta}_k\|^2 \\
\geq & c_6(\sum_{k=1}^{4}\|\boldsymbol{\beta}_{k0} - \boldsymbol{\beta}_k\|^2 + \sum_{l=1}^{2}\|\boldsymbol{\gamma}_{l0} - \boldsymbol{\gamma}_l\|),
\end{aligned}$$

where $c_6 = \min\{c_1, c_5\}$. Finally, by the triangle inequality,

$$\mathbb{M}(\boldsymbol{\theta}) - \mathbb{M}(\boldsymbol{\theta}_0) \geq c_6(\|\boldsymbol{\beta}_0 - \boldsymbol{\beta}\|^2 + \|\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}\|), \tag{B.15}$$

provided that $\boldsymbol{\gamma} \in \mathcal{N}(\boldsymbol{\gamma}_0; \delta_0)$ for some $\delta_0 > 0$. Denoting by $d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \sqrt{\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|} + \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$ leads to the desired (B.6).

Step 2. Note that for any $\boldsymbol{\theta} \in \Theta$, we have

$$\begin{aligned}
&(\mathbb{E} - \mathbb{E}_T)\{m(\boldsymbol{W}_t, \boldsymbol{\theta})\} - (\mathbb{E} - \mathbb{E}_T)\{m(\boldsymbol{W}_t, \boldsymbol{\theta}_0)\} \\
=&\sum_{j=1}^{4}(\mathbb{E} - \mathbb{E}_T)[\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j))^2 \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma})\}] \\
&+ \sum_{i=1}^{4}\sum_{k\neq i}^{4}(\mathbb{E} - \mathbb{E}_T)[\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_k))^2 \mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})\}] \\
&+ 2\sum_{j=1}^{4}\mathbb{E}_T[\{\varepsilon_t(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j))\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma})\}] \\
&+ 2\sum_{i=1}^{4}\sum_{k\neq i}^{4}\mathbb{E}_T[\{\varepsilon_t\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_k)\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})\}] \\
=&S_{1,T} + S_{2,T} + S_{3,T} + S_{4,T}, \quad \text{say.}
\end{aligned} \tag{B.16}$$

We now bound the four terms respectively. For $S_{1,T}$, note that $\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}) = 1 - \sum_{k\neq j}^{4}\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})$ and

$$\begin{aligned}
S_{1,T} \leq &\sum_{j=1}^{4}\left|(\mathbb{E} - \mathbb{E}_T)\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j))^2\, \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma})\right\}\right| \\
\leq &\sum_{j=1}^{4}\left|(\mathbb{E} - \mathbb{E}_T)\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j))^2\, \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\right\}\right| \\
&+ \sum_{j=1}^{4}\sum_{k\neq j}^{4}(\mathbb{E} - \mathbb{E}_T)\left|\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_j))^2\, \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})\right\}\right| \\
= &S_{1,a,T} + S_{1,b,T}, \quad \text{say.}
\end{aligned}$$

For $S_{1,a,T}$, by the Cauchy-Schwartz inequality and the ULLN in Lemma A.1, we have $S_{1,a,T} = \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 o_p(1)$. For $S_{1,b,T}$, due to the compactness of the parameter space for $\boldsymbol{\beta}_j$, Assumption 4 (iv) and Lemma A.6, it can be shown that $S_{1,b,T} = \lambda\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| + O_p(T^{-1})$ for any $\lambda > 0$ and $\boldsymbol{\gamma} \in (c_1 T^{-1}, c_2)$ for any $c_1, c_2 > 0$. Therefore,

$$S_{1,T} \leq \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|^2 o_p(1) + \lambda\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| + O_p(T^{-1}). \tag{B.17}$$

For the second term, we have

$$\begin{aligned}
S_{2,T} \leq &2\sum_{i=1}^{4}\sum_{k\neq i}^{4}\left|(\mathbb{E} - \mathbb{E}_T)\left\{(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{i0} - \boldsymbol{\beta}_k))^2\, \mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})\right\}\right| \\
= &\lambda\|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| + O_p(T^{-1}),
\end{aligned} \tag{B.18}$$

for any $\lambda > 0$, $\boldsymbol{\gamma} \in (c_1 T^{-1}, c_2)$, and any $c_1, c_2 > 0$, implied by the same reasoning for the $S_{1,b,T}$ term. For $S_{3,T}$, similar to $S_{1,T}$, it can be decomposed by

$$\begin{aligned} S_{3,T} \leq & 2\sum_{j=1}^{4}\left|\mathbb{E}_T\left\{\varepsilon_t\left(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0}-\boldsymbol{\beta}_j)\right)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\right\}\right| \\ & + 2\sum_{j=1}^{4}\sum_{k\neq j}^{4}\left|\mathbb{E}_T\left\{\varepsilon_t\left(\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_{j0}-\boldsymbol{\beta}_j)\right)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(k)}(\boldsymbol{\gamma})\right\}\right| \\ = & S_{3,a,T} + S_{3,b,T}, \quad \text{say.} \end{aligned}$$

For $S_{3,a,T}$, by the martingale central limit theorem (Hall and Heyde, 1980) we have $S_{3,a,T} = \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| O_p(T^{-1/2})$. For $S_{3,b,T}$, using the same arguments as that for $S_{1,b,T}$, $S_{3,b,T} = \lambda\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| + O_p(T^{-1})$. Therefore,

$$S_{3,T} \leq \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| O_p(T^{-1/2}) + \lambda\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| + O_p(T^{-1}). \tag{B.19}$$

For $S_{4,T}$, following the same reasons for $S_{2,T}$, it can be shown that

$$S_{4,T} \leq \lambda\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| + O_p(T^{-1}). \tag{B.20}$$

Putting (B.17)–(B.20) together, we obtain that if $\boldsymbol{\gamma} \in (c_1 T^{-1}, c_2)$ for some $c_1, c_2 > 0$, then

$$\begin{aligned} (\mathbb{E}-\mathbb{E}_T)\{m(\boldsymbol{W}_t,\boldsymbol{\theta}) - m(\boldsymbol{W}_t,\boldsymbol{\theta}_0)\} \leq & \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| O_p(T^{-1/2}) + \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 o_p(1) \\ & + 4\lambda\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| + O_p(T^{-1}). \end{aligned}$$

Since $\mathbb{E}_T\{m(\boldsymbol{W}_t,\widehat{\boldsymbol{\theta}})\} \leq \mathbb{E}_T\{m(\boldsymbol{W}_t,\boldsymbol{\theta}_0)\}$ and (B.15), we obtain

$$\begin{aligned} C_6(\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 + \|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|) \leq & \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| O_p(T^{-1/2}) + \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|^2 o_p(1) \\ & + 4\lambda\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| + O_p(T^{-1}). \end{aligned}$$

Since the above bound holds for any $\lambda \in (0,1)$, we can take $\lambda < C_6/4$, which delivers

$$C_6\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 + (C_6 - 4\lambda)\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\| \leq \|\boldsymbol{\beta}-\boldsymbol{\beta}_0\| O_p(T^{-1/2}) + \|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 o_p(1) + O_p(T^{-1}),$$

which further implies $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|^2 = O_p(T^{-1})$, and thus, $\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\| = O_p(T^{-1})$. □

Proof of Corollary 3.2

PROOF. It can be seen straightforwardly from the proof of Corollary 3.1 that for each $k \in \{1, \cdots, 4\}$,

$$\mathbb{P}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0) \triangle R_k(\widehat{\boldsymbol{\gamma}}) | \mathcal{D}_T\} \lesssim \sum_{l=1}^{2}\|\widehat{\gamma}_l - \gamma_{l0}\|, \tag{B.21}$$

which is of order $O_p(T^{-1/2})$ by Theorem 3.2. With the uniformly integratability of $\mathbb{P}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0) \triangle R_k(\widehat{\boldsymbol{\gamma}}) \mid \mathcal{D}_T\}$, the conclusion of the corollary follows. □

B.5. Proof of Theorem 3.3. The following proof is for Theorem 3.3 on the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$, which requires the following lemmas. Considering that the proofs for these lemmas are quite lengthy, we provide their proofs later in Subsections B.6–B.10.

For any $(\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{v}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{4p+d_1+d_2}$, we define

$$Q_T(\boldsymbol{u}, \boldsymbol{v}) = \sum_{t=1}^{T} \{m(\boldsymbol{W}_t, \boldsymbol{\beta}_0 + \frac{\boldsymbol{u}}{\sqrt{T}}, \boldsymbol{\gamma}_0 + \frac{\boldsymbol{v}}{T}) - m(\boldsymbol{W}_t, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)\}. \tag{B.22}$$

The following lemma establishes the separability for $Q_T(\boldsymbol{u}, \boldsymbol{v})$, whose proof is available in Section B.6.

LEMMA B.1. *Under Assumptions 1-5, uniformly for $(\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{v}^{\mathrm{T}})^{\mathrm{T}}$ in any compact region of $\mathbb{R}^{4p+d_1+d_2}$, we have*

$$Q_T(\boldsymbol{u}, \boldsymbol{v}) = W_T(\boldsymbol{u}) + D_T(\boldsymbol{v}) + o_p(1), \tag{B.23}$$

where

$$W_T(\boldsymbol{u}) = \sum_{j=1}^{4} [\boldsymbol{u}_j^{\mathrm{T}} \mathbb{E}\{\boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\} \boldsymbol{u}_j - 2\frac{\boldsymbol{u}_j^{\mathrm{T}}}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{X}_t \varepsilon_t \mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)], \tag{B.24}$$

and

$$D_T(\boldsymbol{v}) = \sum_{t=1}^{T} \sum_{l=1}^{2} \sum_{(j,k)\in\mathcal{S}(l)} \xi_t^{(j,k)} \mathbb{1}\left\{s_l^{(j)}\left(Tq_{l,t} + \boldsymbol{Z}_{-1,l,t}^{\mathrm{T}} \boldsymbol{v}_{-1,l}\right) \le 0 < s_l^{(j)} Tq_{l,t}\right\}, \tag{B.25}$$

with

$$\xi_t^{(j,k)} = \left(\boldsymbol{\delta}_{jk,0}^{\mathrm{T}} \boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\delta}_{jk,0} + 2\boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\delta}_{jk,0} \varepsilon_t\right) \{\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0) + \mathbb{1}_t^{(k)}(\boldsymbol{\gamma}_0)\},$$

where $\boldsymbol{\delta}_{jk,0} = \boldsymbol{\beta}_{j0} - \boldsymbol{\beta}_{k0}$, $q_{l,t} = \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma}_{l0}$, $\mathcal{S}(l)$ is the set of indices of adjacent regions split by the l-th hyperplane as defined in (3), and $s_l^{(j)} = \operatorname{sign}(\boldsymbol{z}_l^{\mathrm{T}} \boldsymbol{\gamma}_{l0})$ for $\boldsymbol{z} \in R_j(\boldsymbol{\gamma}_0)$ as defined in (2) of the main text.

The next lemma is to obtain the finite-dimensional weak limit of $D_T(\boldsymbol{v})$, whose behaviour is determined by the point processes induced by the observations which are near the splitting hyperplanes. The following notations are needed for this lemma and its proof. For each $l = 1, 2$ and $(j,k) \in \mathcal{S}(l)$, suppose $(q_l, \boldsymbol{Z}_{-1,l}, \xi^{(j,k)})$ follows the stationary distribution of $(q_{l,t}, \boldsymbol{Z}_{-1,l,t}, \xi_t^{(j,k)})$. We denote $F_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{Z}_{-1,l})$ and $F_{\xi^{(j,k)}|q_l,\boldsymbol{Z}_{-1,l}}(\xi|q_l, \boldsymbol{Z}_{-1,l})$ as the conditional distributions of q_l on $\boldsymbol{Z}_{-1,l}$ and $\xi^{(j,k)}$ on $(q_l, \boldsymbol{Z}_{-1,l})$, respectively, and the corresponding conditional densities are $f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{Z}_{-1,l})$ and $f_{\xi^{(j,k)}|q_l,\boldsymbol{Z}_{-1,l}}(\xi|q_l, \boldsymbol{Z}_{-1,l})$, respectively. Let $\mathcal{Z}_{-1,l}$ be the compact support of the density of $\boldsymbol{Z}_{-1,l}$ as required in Assumption 5.

LEMMA B.2. *Under Assumptions 1-5, the finite-dimensional weak limit of $D_T(\boldsymbol{v})$ in (B.25) is*

$$D(\boldsymbol{v}) = \sum_{l=1}^{2} \sum_{j,k\in\mathcal{S}(l)} \sum_{i=1}^{\infty} \xi_i^{(j,k)} \mathbb{1}\left\{s_l^{(j)}\left(J_{i,l}^{(j,k)} + (\boldsymbol{Z}_{l,i}^{(j,k)})^{\mathrm{T}} \boldsymbol{v}_{-1,l}\right) \le 0 < s_l^{(j)} J_{i,l}^{(j,k)}\right\}, \tag{B.26}$$

for $\boldsymbol{v} = (\boldsymbol{v}_1^{\mathrm{T}}, \boldsymbol{v}_2^{\mathrm{T}})^{\mathrm{T}}$, where $\{(\xi_i^{(j,k)}, \boldsymbol{Z}_{l,i}^{(j,k)})\}_{i=1}^{\infty}$ are independent copies of $(\bar{\xi}^{(j,k)}, \boldsymbol{Z}_{-1,l})$ with $\bar{\xi}^{(j,k)} \sim F_{\xi^{(j,k)}|q_l,\boldsymbol{Z}_{-1,l}}(\xi|0, \boldsymbol{Z}_{-1,l})$, $J_{i,l}^{(j,k)} = \mathcal{J}_{l,i}^{(j,k)} / f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{l,i}^{(j,k)})$ with $\mathcal{J}_{l,i}^{(j,k)} =$

$s_l^{(j)} \sum_{n=1}^{i} \mathcal{E}_{l,n}^{(j,k)}$ *and* $\{\mathcal{E}_{l,n}^{(j,k)}\}_{n=1}^{\infty}$ *are independent unit exponential variables which are independent of* $\{(\xi_i^{(j,k)}, \boldsymbol{Z}_{l,i}^{(j,k)})\}_{i=1}^{\infty}$*. Moreover,* $\{(\xi_i^{(j,k)}, \boldsymbol{Z}_{l,i}^{(j,k)}, J_{l,i}^{(j,k)})\}_{i=1}^{\infty}$ *are independent across* $l = 1, 2$ *and* $(j,k) \in \mathcal{S}(l)$*.*

The following Lemma B.3 establishes the stochastic equi-lower-semicontinuity of $\{D_T(\boldsymbol{v})\}$, which together with the finite-dimensional converges in distribution implies the epi-convergence in distribution.

LEMMA B.3. *Under Assumptions 1-5, the sequence* $\{D_T(\boldsymbol{v})\}$ *defined in* (B.25) *is stochastic equi-lower-semicontinuous, namely that for any compact set* $B \subset \mathbb{R}^{d_1+d_2}$ *and any* $\epsilon, \delta > 0$*, there exists* $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m \in B$*, where* m *is a finite integer depending on* B*, and some open sets* $V(\boldsymbol{v}_1), \cdots, V(\boldsymbol{v}_m)$ *covering* B *and containing* $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m$*, such that*

$$\limsup_{T\to\infty} \mathbb{P}\left(\cup_{j=1}^{m}\left\{\inf_{\boldsymbol{v}\in V(\boldsymbol{v}_j)} D_T(\boldsymbol{v}) \le \min(\epsilon^{-1}, D_T(\boldsymbol{v}_j) - \epsilon)\right\}\right) < \delta.$$

To present our next lemma, we first define the following class of piece-wise constant functions on $\mathbb{R}^d$ as

$$\mathcal{F} = \left\{ f(\boldsymbol{v}) = \sum_{i=0}^{\infty} a_i \mathbb{1}\{\boldsymbol{v} \in F_i\}, \ \ a_i \in \mathbb{R}, F_i \text{ is a connected set in } \mathbb{R}^d, F_i \cap F_j = \emptyset \text{ if } i \ne j \right\}.$$

For each $f \in \mathcal{F}$, let $\tilde{f} = \sum_{i=0}^{\infty} i \mathbb{1}\{\boldsymbol{v} \in F_i\}$ be its associated pure jump process, which has a jump size 1 when moving from F_i to F_{i+1}. We refer to the sets $\{F_i\}$ as the level sets for f and $\tilde{f}$. Note that any realization of both $D_T(\boldsymbol{v})$ and $D(\boldsymbol{v})$ belongs to $\mathcal{F}$. Lemma B.4 below ensures that the centroid of the armgin set of $f \in \mathcal{F}$, when viewed as a functional from $\mathcal{F}$ to $\mathbb{R}$, is a continuous mapping functional under the topology of epi-convergence. It is similar in spirit to Lemma 3.1 of Lan et al. (2009), where they established the smallest and largest argmin functionals are continuous mappings in the univariate Skorohod space, while our result is under the metric induced by the epi-convergence in multivariate space.

LEMMA B.4. *Given a compact space* E *for* $\boldsymbol{v}$*, suppose that (i) on the domain* E*, the sequence* $\{f_n \in \mathcal{F}\}$ *epi-converges to* $f_0 \in \mathcal{F}$ *and its jump process* $\{\tilde{f}_n\}$ *also epi-converges to* $\tilde{f}_0$*; (ii) there are finite numbers of jumps of* $\{\tilde{f}_n\}$ *and* $\tilde{f}_0$ *in* E*; (iii)* f_0 *has a unique level set. Let* G_n *and* G_0 *be the set in* E *on which* f_n *and* f_0 *are minimized, respectively. Then,*

$$\frac{\int \boldsymbol{v}\mathbb{1}(\boldsymbol{v} \in G_n) d\boldsymbol{v}}{\int \mathbb{1}(\boldsymbol{v} \in G_n) d\boldsymbol{v}} \to \frac{\int \boldsymbol{v}\mathbb{1}(\boldsymbol{v} \in G_0) d\boldsymbol{v}}{\int \mathbb{1}(\boldsymbol{v} \in G_0) d\boldsymbol{v}}, \quad \textit{as } n \to \infty. \tag{B.27}$$

Let $\ell^{\infty}(\mathbb{B})$ be the space of all bounded functions equipped with the uniform norm on the domain $\mathbb{B}$, where $\mathbb{B}$ is the parameter space for $\boldsymbol{\beta}$. The following lemma establishes the weak convergence of W_T in $\ell^{\infty}(\mathbb{B})$ and its asymptotic independence with D_T.

LEMMA B.5. *Under Assumptions 1-5, the sequence* $\{W_T\}_{T=1}^{\infty}$ *defined in* (B.24) *weakly converges to* W *in* $\ell^{\infty}(\mathbb{B})$*, where for any* $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathrm{T}}, \cdots, \boldsymbol{u}_4^{\mathrm{T}})^{\mathrm{T}}$*,* $W(\boldsymbol{u}) = \sum_{k=1}^{4} W_k(\boldsymbol{u}_k)$*,*

$$W_k(\boldsymbol{u}_k) = \boldsymbol{u}_k^{\mathrm{T}} \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \mathbb{1}\left\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0)\right\}\right] \boldsymbol{u}_k - 2\boldsymbol{u}_k^{\mathrm{T}} \boldsymbol{H}_k, \tag{B.28}$$

$\boldsymbol{H}_k \sim \boldsymbol{N}(\boldsymbol{0}, \Sigma_k)$ *and* $\Sigma_k = \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} \varepsilon^2 \mathbb{1}\left\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0)\right\}\right]$*. Furthermore, the random function* $W(\boldsymbol{u})$ *is independent of* $D(\boldsymbol{v})$ *defined in* (B.26)*.*

With the above Lemmas B.1–B.5, we are now ready to prove Theorem 3.3 as follows.

Proof of Theorem 3.3

PROOF. Let $\boldsymbol{V}_T = T(\widehat{\gamma} - \gamma_0)$ with $\widehat{\gamma} \in \widehat{G}$ and $\boldsymbol{U}_T = \sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ be standardizations of the LSEs for γ_0 and $\boldsymbol{\beta}_0$, respectively. By the definition of $(\widehat{\gamma}, \widehat{\boldsymbol{\beta}})$,

$$\begin{aligned}(\boldsymbol{V}_T, \boldsymbol{U}_T) &\in \arg\min_{(\boldsymbol{v},\boldsymbol{u})} \left[T\mathbb{E}_T \left\{ m(\boldsymbol{W}_t, \boldsymbol{\beta}_0 + \frac{\boldsymbol{u}}{\sqrt{T}}, \gamma_0 + \frac{\boldsymbol{v}}{T}) - m(\boldsymbol{W}_t, \boldsymbol{\beta}_0, \gamma_0) \right\} \right] \\ &\in \arg\min_{(\boldsymbol{v},\boldsymbol{u})} Q_T(\boldsymbol{v}, \boldsymbol{u}), \end{aligned} \tag{B.29}$$

where $\mathbb{E}_T$ is the empirical average operator, $Q_T(\boldsymbol{v}, \boldsymbol{u})$ is defined in (B.22), $\boldsymbol{v} = (\boldsymbol{v}_1^{\mathrm{T}}, \boldsymbol{v}_2^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{u} = (\boldsymbol{u}_1^{\mathrm{T}}, \cdots, \boldsymbol{u}_4^{\mathrm{T}})^{\mathrm{T}}$. The proof includes the following three steps: (1) the separability and finite-dimensional convergence of $\{Q_T(\boldsymbol{v}, \boldsymbol{u})\}_{T=1}^{\infty}$, (2) the epi-convergence of the random functions $\{Q_T\}_{T=1}^{\infty}$ to Q, and (3) the continuous mapping for the centroid of the argmin set.

Step 1. Separability and finite-dimensional convergence.

According to Lemma B.1, $Q_T(\boldsymbol{v}, \boldsymbol{u})$ can be separated as

$$Q_T(\boldsymbol{v}, \boldsymbol{u}) = W_T(\boldsymbol{u}) + D_T(\boldsymbol{v}) + o_p(1), \tag{B.30}$$

uniformly for $(\boldsymbol{u}^{\mathrm{T}}, \boldsymbol{v}^{\mathrm{T}})^{\mathrm{T}}$ in any compact set of $\mathbb{R}^{4p+d_1+d_2}$, where $W_T(\boldsymbol{u})$ and $D_T(\boldsymbol{v})$ are defined in (B.24) and (B.25), respectively.

Let $Q(\boldsymbol{v}, \boldsymbol{u}) = W(\boldsymbol{u}) + D(\boldsymbol{v})$, where $W(\boldsymbol{u})$ is defined in (B.24) and $D(\boldsymbol{v})$ is given in (B.26). Note that $D(\boldsymbol{v}) = D_1(\boldsymbol{v}_1) + D_2(\boldsymbol{v}_2)$, where

$$D_l(\boldsymbol{v}_l) = \sum_{j,k \in \mathcal{S}(l)} \sum_{i=1}^{\infty} \xi_i^{(j,k)} \mathbb{1}\left\{ s_l^{(j)} \left(J_{i,l}^{(j,k)} + (\boldsymbol{Z}_{l,i}^{(j,k)})^{\mathrm{T}} \boldsymbol{v}_{-1,l} \right) \le 0 < s_l^{(j)} J_{i,l}^{(j,k)} \right\},$$

for $l = 1$ and 2. By Lemma B.2, for any finite positive integer k and $(\boldsymbol{v}_{(1)}, \cdots, \boldsymbol{v}_{(k)})$ where $\boldsymbol{v}_{(i)} = (\boldsymbol{v}_{(i),1}^{\mathrm{T}}, \boldsymbol{v}_{(i),2}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{d_1+d_2}$ for each $i \in \{1, \cdots, 4\}$, we have

$$\left(D_T(\boldsymbol{v}_{(1)}), \cdots, D_T(\boldsymbol{v}_{(k)}) \right) \xrightarrow{d} \left(D(\boldsymbol{v}_{(1)}), \cdots, D(\boldsymbol{v}_{(k)}) \right), \tag{B.31}$$

namely, $D(\boldsymbol{v})$ is the finite-dimensional limiting distribution of $D_T(\boldsymbol{v})$. The finite-dimensional weak convergence of $W_T(\boldsymbol{u})$ to $W(\boldsymbol{u})$ is implied by Lemma B.5. Therefore, $Q_T(\boldsymbol{u}, \boldsymbol{v})$ weakly converges to $Q(\boldsymbol{u}, \boldsymbol{v})$ in the finite-dimensional sense.

Step 2. Epi-convergence.

Lemma B.3 establishes the stochastic equi-lower-semicontinuouity (s.e-l-sc) of the sequence $\{D_T\}_{T=1}^{\infty}$. From the regular form of $\{W_T\}_{T=1}^{\infty}$, this sequence of random functions converges in distribution to W with respect to the topology of uniform convergence, implying $\{W_T\}_{T=1}^{\infty}$ epi-converge in distribution to W. Then by the finite-dimensional convergence of $\{W_T\}_{T=1}^{\infty}$ implied from Lemma B.5 and Theorem 3 of Knight (1999), $\{W_T\}_{T=1}^{\infty}$ is a sequence of s.e-l-sc random functions. Consequently, $\{Q_T\}_{T=1}^{\infty}$ are s.e-l-sc, which together with the finite-dimensional weak convergence shown in Step 1 implies that $\{Q_T\}_{T=1}^{\infty}$ epi-converges in distribution to Q by Lemma A.9.

For any given $\boldsymbol{v}$, by the separatability of $Q(\boldsymbol{u},\boldsymbol{v}) = W(\boldsymbol{u}) + D(\boldsymbol{v})$, where $W(\boldsymbol{u})$ is quadratic in $\boldsymbol{u}$ as shown in (B.28), we can see that $Q(\boldsymbol{u},\boldsymbol{v})$ is minimized at $U = (U_1,\cdots,U_4)^{\mathrm{T}}$, where for $k \in \{1,\cdots,4\}$,

$$U_k = \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mathbb{1}\left\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0)\right\}\right]^{-1} H_k, \;\; H_k \sim \boldsymbol{N}(\boldsymbol{0},\Sigma_k),$$

and Σ_k is given in Lemma B.5. By Theorem 1 of Knight (1999), we obtain $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = U_T \xrightarrow{d} U$. Let G_D be the argmin set of $D(\boldsymbol{v})$. Since Assumption 3.(ii) implies that neither $\boldsymbol{Z}_{1,t}$ nor $\boldsymbol{Z}_{2,t}$ is multicollinear, following the same arguments as in Yu and Fan (2021), it can be shown G_D is compact almost surely, so that its centroid is well defined. It is worth noting that because the minimizers of $D(\boldsymbol{v})$ are not unique, Theorem 1 of Knight (1999) can not be directly applicable to imply the weak convergence of $\arg\min_{\boldsymbol{v}} D_T(\boldsymbol{v})$ to that of $\arg\min_{\boldsymbol{v}} D(\boldsymbol{v})$. Instead, we consider the centroid of argmin, which can be viewed as a continuous functional of a process, to obtain the desired weak convergence in Theorem 3.3.

Step 3. Continuous mapping for the centroid of the argmin set.

Since $\{D_T(\boldsymbol{v})\}_{T=1}^{\infty}$ and $D(\boldsymbol{v})$ can be endowed into a complete and separable metric space induced by the epi-convergence, we can find a probability space and random elements with $D'_T(\boldsymbol{v}) \overset{d}{=} D_T(\boldsymbol{v})$ for each $T \geq 1$ and $D'(\boldsymbol{v}) \overset{d}{=} D(\boldsymbol{v})$, such that $D'_T(\boldsymbol{v})$ epi-converges to $D'(\boldsymbol{v})$ with probability 1 (van der Vaart and Wellner, 1996). Let $\widehat{\mathcal{G}}'$ and $\mathcal{G}'_D$ be the argmin sets of $D'_T(\boldsymbol{v})$ and $D'(\boldsymbol{v})$, respectively. Condition (i) of Lemma B.4 is ensured by the epi-convergence of $\{D'_T(\boldsymbol{v})\}$ to $D'(\boldsymbol{v})$. Because the point process induced by $\{Tq_{l,t}\}$ is asymptotic Poisson, there are stochastically finite number of jumps in any compact region, and Condition (ii) Lemma B.4 holds with the probability approaching 1. Also, Condition (iii) is ensured by the continuity of the jump size $\xi_i^{(j,k)}$ of $D(\boldsymbol{v})$. Applying Lemma B.4, we have $C(\widehat{\mathcal{G}}') \to C(\mathcal{G}'_D)$, where $C(E)$ denotes the centroid of any bounded set E. Hence, we conclude that $T(\widehat{\boldsymbol{\gamma}}^c - \boldsymbol{\gamma}_0) = C(\widehat{\mathcal{G}}) \xrightarrow{d} C(\mathcal{G}_D) = \boldsymbol{\gamma}_D^c$. Finally, the asymptotic independence between $\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ and $T(\widehat{\boldsymbol{\gamma}}^c - \boldsymbol{\gamma}_0)$ is implied by the independence between $W(\boldsymbol{u})$ and $D(\boldsymbol{v})$ established in Lemma B.5. Because $T(\widehat{\boldsymbol{\gamma}}_1^c - \boldsymbol{\gamma}_{10})$ and $T(\widehat{\boldsymbol{\gamma}}_2^c - \boldsymbol{\gamma}_{20})$ depend asymptotically on $D_1(\boldsymbol{v})$ and $D_2(\boldsymbol{v})$, respectively, which are shown to be independent in Part 3 of the proof of Lemma B.2, the asymptotic independence between $T(\widehat{\boldsymbol{\gamma}}_1^c - \boldsymbol{\gamma}_{10})$ and $T(\widehat{\boldsymbol{\gamma}}_2^c - \boldsymbol{\gamma}_{20})$ follows. □

B.6. Proof of Lemma B.1.

PROOF. First, the left-hand of (B.23) admits the following decomposition:

$$\begin{aligned}
&T\mathbb{E}_T\{m(\boldsymbol{W}_t,\boldsymbol{\beta}_0 + \frac{\boldsymbol{u}}{\sqrt{T}},\boldsymbol{\gamma}_0 + \frac{\boldsymbol{v}}{T}) - m(\boldsymbol{W}_t,\boldsymbol{\beta}_0,\boldsymbol{\gamma}_0)\} \\
=&\sum_{j=1}^{4}\sum_{t=1}^{T}(\boldsymbol{u}_j^{\mathrm{T}}\frac{\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}}{T}\boldsymbol{u}_j - \boldsymbol{u}_j^{\mathrm{T}}\frac{2}{\sqrt{T}}\boldsymbol{X}_t\varepsilon_t)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0 + \frac{\boldsymbol{v}}{T}) \\
&+\sum_{i\neq j}^{4}\sum_{t=1}^{T}\{(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}})^{\mathrm{T}}\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}}) + 2\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}})\varepsilon_t\}\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0 + \frac{\boldsymbol{v}}{T}) \\
=&\sum_{j=1}^{4}H_j(\boldsymbol{h}) + \sum_{i\neq j}^{4}F_{ij}(\boldsymbol{h}), \;\; \text{say.}
\end{aligned} \tag{B.32}$$

For the H_j term, let $R_{j,t} = \boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{u}_j \mathbb{1}_t^{(j)}(\gamma_0)$, by the ULLN in Lemma A.1,

$$(\mathbb{E}_T - \mathbb{E})\{R_{j,t}\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right)\} = o_p(1). \tag{B.33}$$

Note that

$$\begin{aligned}
&\mathbb{E}\{\left|R_{j,t}\mathbb{1}_t^{(j)}(\gamma_0 + \frac{\boldsymbol{v}}{T}) - R_{j,t}\right|\} \\
\leq &\sum_{l=1}^{2} \mathbb{E}\{R_{j,t}|\mathbb{1}_{l,t}(\gamma_{l0}) - \mathbb{1}_{l,t}(\gamma_{l0} + \frac{\boldsymbol{v}_l}{T})|\} \overset{(i)}{\lesssim} \frac{\sum_{l=1}^{2}\|\boldsymbol{v}_l\|}{T} = o(1),
\end{aligned} \tag{B.34}$$

where (i) is implied by Lemma A.2. Then, combining (B.33) and (B.34) yields

$$\begin{aligned}
&\sum_{t=1}^{T} \boldsymbol{u}_j^{\mathrm{T}} \frac{\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}}{T} \boldsymbol{u}_j \mathbb{1}_t^{(j)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right) \\
=&\mathbb{E}_T\left\{R_{j,t}\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right)\right\} \\
=&\mathbb{E}(R_{j,t}) + o_p(1) = \sum_{j=1}^{4} \boldsymbol{u}_j^{\mathrm{T}} \mathbb{E}\left\{\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}\mathbb{1}_t^{(j)}(\gamma_0)\right\}\boldsymbol{u}_j + o_p(1).
\end{aligned} \tag{B.35}$$

For the second part of $H_j(\boldsymbol{h})$, let $S_{j,t} = 2\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{X}_t\varepsilon_t\mathbb{1}_t^{(j)}(\gamma_0)$. Note that

$$\begin{aligned}
&\sqrt{T}\mathbb{E}_T\left[S_{j,t}\left\{\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right) - \mathbb{1}_t^{(j)}(\gamma_0)\right\}\right] \\
\leq &\sum_{l=1}^{2}\sqrt{T}\mathbb{E}_T\left\{|S_{j,t}|\left|\mathbb{1}_{l,t}(\gamma_{l0}) - \mathbb{1}_{l,t}(\gamma_{l0} + \frac{\boldsymbol{v}_l}{T})\right|\right\} = o_p(1),
\end{aligned}$$

according to (A.20) in Lemma A.5. Hence, applying Lemma A.5 gives

$$\sqrt{T}\mathbb{E}_T\left\{S_{j,t}\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right)\right\} = \sqrt{T}\mathbb{E}_T\left\{S_{j,t}\mathbb{1}_t^{(j)}(\gamma_0)\right\} + o_p(1). \tag{B.36}$$

Combining (B.35) and (B.36) and summing across $j = 1, \cdots, 4$ leads to

$$\sum_{j=1}^{4} H_j(\boldsymbol{h}) = W_T(\boldsymbol{u}) + o_p(1). \tag{B.37}$$

For the $F_{ij}(\boldsymbol{h})$ terms ($i \neq j \in \{1, \cdots, 4\}$) in (B.32), we divide them into two cases according to whether there exists $l \in \{1, 2\}$ such that $(i, j) \in \mathcal{S}(\gamma_{l0})$ or not. For those (i, j) that does not have $l \in \{1, 2\}$ such that $(i, j) \in \mathcal{S}(l)$, i.e., $s_1^{(i)} \neq s_1^{(j)}$ and $s_2^{(i)} \neq s_2^{(j)}$,

$$\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0 + \frac{\boldsymbol{v}}{T}\right) \leq \left|\mathbb{1}_{1,t}(\gamma_{l0}) - \mathbb{1}_{1,t}\left(\gamma_{10} + \frac{\boldsymbol{v}_1}{T}\right)\right|\left|\mathbb{1}_{2,t}(\gamma_{h0}) - \mathbb{1}_{2,t}\left(\gamma_{20} + \frac{\boldsymbol{v}_2}{T}\right)\right|.$$

Then, applying (A.21) in Lemma A.5, where we define U_t in Lemma A.5 as

$$\left|(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}})^{\mathrm{T}}\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}}) + 2\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\delta}_{ij,0} - \frac{\boldsymbol{u}_j}{\sqrt{T}})\varepsilon_t\right|,$$

yields that

$$F_{ij}(\boldsymbol{h}) = o_p(1), \ \text{ if } (i, j) \notin \mathcal{S}(l) \text{ for any } l \in \{1, 2\}. \tag{B.38}$$

Otherwise, if there exists $l \in \{1, 2\}$ such that $(i, j) \in \mathcal{S}(l)$,

$$F_{ij}(\boldsymbol{h}) = T\mathbb{E}_T\{\xi_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma_0 + \frac{\boldsymbol{v}}{T})\} + \sqrt{T}\mathbb{E}_T\{T_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma_0 + \frac{\boldsymbol{v}}{T})\}$$

$$+\mathbb{E}_T\{U_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma_0+\frac{\boldsymbol{v}}{T})\}+\sqrt{T}\mathbb{E}_T\{V_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma_0+\frac{\boldsymbol{v}}{T})\}, \tag{B.39}$$

where $\xi_t^{(i,j)}$ is defined in (B.25), and

$$T_t^{(i,j)}=\boldsymbol{\delta}_{ij,0}^{\mathrm{T}}\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{u}_j,\;\; U_t^{(i,j)}=\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{u}_j \text{ and } V_t^{(i,j)}=-2\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{u}_j\varepsilon_t.$$

For the first term on the right-hand side of (B.39), we note that $\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)=1$ means that $\boldsymbol{Z}_t$ is classified into $R_i(\gamma_0)$ under the true γ_0, but is classified into $R_j(\boldsymbol{\gamma})$ under the candidate parameter $\boldsymbol{\gamma}$. Since the i-th and the j-th regions are on the opposite sides of the l-th hyperplane, while are on the same side of the h-th hyperplane for the $h\neq l\in\{1,2\}$, we have the following two implications: (i) $\operatorname{sign}(\boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma_{l0})\neq\operatorname{sign}\left\{\boldsymbol{Z}_{l,t}^{\mathrm{T}}\left(\gamma_{l0}+\frac{\boldsymbol{v}_l}{T}\right)\right\}$, which is equivalent to

$$\mathbb{1}\left\{s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}\left(\gamma_{l0}+\frac{\boldsymbol{v}_l}{T}\right)\le 0<s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma_{l0}\right\}=1;$$

and (ii) $\operatorname{sign}(\boldsymbol{Z}_{h,t}^{\mathrm{T}}\gamma_{h0})=\operatorname{sign}\left\{\boldsymbol{Z}_{h,t}^{\mathrm{T}}\left(\gamma_{h0}+\frac{\boldsymbol{v}_h}{T}\right)\right\}$ for $h\neq l\in\{1,2\}$, which is equivalent to

$$\mathbb{1}\left\{0<\min\{s_h^{(i)}\boldsymbol{Z}_{h,t}^{\mathrm{T}}\gamma_{h0},s_h^{(i)}\boldsymbol{Z}_{h,t}^{\mathrm{T}}\left(\gamma_{h0}+\frac{\boldsymbol{v}_h}{T}\right)\}\right\}=1.$$

For $(i,j)\in\mathcal{S}(l)$, let $\mathbb{1}_t^{(i,j)}(\gamma_0)=\mathbb{1}_t^{(i)}(\gamma_0)+\mathbb{1}_t^{(j)}(\gamma_0)$. It is noted that

$$\begin{aligned}&\left|\mathbb{1}_t^{(i,j)}(\gamma_0)\mathbb{1}_t^{(j)}(\gamma_0+\frac{\boldsymbol{v}}{T})-\mathbb{1}_t^{(i,j)}(\gamma_0)\mathbb{1}_{l,t}\{s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}(\gamma_{l0}+\frac{\boldsymbol{v}_l}{T})\le 0<s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma_{l0}\}\right|\\ \le&\left|\mathbb{1}_{1,t}(\gamma_{10})-\mathbb{1}_{1,t}(\gamma_{10}+\frac{\boldsymbol{v}_1}{T})\right|\left|\mathbb{1}_{2,t}(\gamma_{20})-\mathbb{1}_{2,t}(\gamma_{20}+\frac{\boldsymbol{v}_2}{T})\right|.\end{aligned} \tag{B.40}$$

Applying (A.21) in Lemma A.5, we have

$$T\mathbb{E}_T\left\{|\xi_t^{(i,j)}|\left|\mathbb{1}_{1,t}(\gamma_{10})-\mathbb{1}_{1,t}\left(\gamma_{10}+\frac{\boldsymbol{v}_1}{T}\right)\right|\left|\mathbb{1}_{2,t}(\gamma_{20})-\mathbb{1}_{2,t}\left(\gamma_{20}+\frac{\boldsymbol{v}_2}{T}\right)\right|\right\}=o_p(1),$$

which, together with (B.40), implies that

$$\begin{aligned}&T\mathbb{E}_T\left\{\xi_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)\right\}\\ =&T\mathbb{E}_T\left\{\xi_t^{(i,j)}\mathbb{1}_{l,t}\left\{s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}\left(\gamma_{l0}+\frac{\boldsymbol{v}_l}{T}\right)\le 0<s_l^{(i)}\boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma_{l0}\right\}\right\}+o_p(1)\\ =&T\mathbb{E}_T\left\{\xi_t^{(i,j)}\mathbb{1}_{l,t}\left\{s_l^{(i)}\left(Tq_{l,t}+\boldsymbol{Z}_{l,t}^{\mathrm{T}}\boldsymbol{v}_l\right)\le 0<s_l^{(i)}Tq_{l,t}\right\}\right\}+o_p(1)\\ =&D_T^{(i,j)}(\boldsymbol{v})+o_p(1),\ \ \text{say},\end{aligned} \tag{B.41}$$

where in the second equality $q_{l,t}=\boldsymbol{Z}_{l,t}^{\mathrm{T}}\gamma_0$.

For the second term of (B.39), note that

$$\left|T_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)\right|\le\left|T_t^{(i,j)}\right|\left|\mathbb{1}_{l,t}(\gamma_{l0})-\mathbb{1}_{l,t}(\gamma_{l0}+\frac{\boldsymbol{v}_l}{T})\right|.$$

According to (A.20) in Lemma A.5, it holds that

$$\sqrt{T}\mathbb{E}_T\left\{T_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)\right\}=o_p(1). \tag{B.42}$$

With the same arguments, we have

$$\mathbb{E}_T\left\{U_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)\right\}=o_p(1), \tag{B.43}$$

$$\sqrt{T}\mathbb{E}_T\left\{V_t^{(i,j)}\mathbb{1}_t^{(i)}(\gamma_0)\mathbb{1}_t^{(j)}\left(\gamma_0+\frac{\boldsymbol{v}}{T}\right)\right\}=o_p(1). \tag{B.44}$$

Finally, combining (B.39) and the four parts (B.41)–(B.44) yields

$$F_{ij}(\boldsymbol{h})=D_T^{(i,j)}(\boldsymbol{v})+o_p(1),\ \ \text{if there exists } l\in\{1,2\} \text{ such that } (i,j)\in\mathcal{S}(l). \tag{B.45}$$

Since $Q_T(\boldsymbol{u},\boldsymbol{v})=\sum_{j=1}^4 H_j(\boldsymbol{h})+\sum_{i\neq j}^4 F_{ij}(\boldsymbol{h})$ as shown in (B.32), using (B.37) for the $H_j(\boldsymbol{h})$ terms, and (B.38) and (B.45) for the $F_{ij}(\boldsymbol{h})$ terms, the desired result (B.23) for the decomposition of $Q_T(\boldsymbol{u},\boldsymbol{v})$ is obtained. □

B.7. Proof of Lemma B.2.

PROOF. For notational simplicity, in this proof, we show the marginal weak convergence of $D_T(\boldsymbol{v})$, i.e., $D_T(\boldsymbol{v})\xrightarrow{d}D(\boldsymbol{v})$ for any fixed $\boldsymbol{v}$, since the finite-dimensional weak convergence can be easily extended with the similar argument but more involved notations. Specifically, to show that $\big(D_T(\boldsymbol{v}_{(1)}),\cdots,D_T(\boldsymbol{v}_{(m)})\big)\xrightarrow{d}\big(D(\boldsymbol{v}_{(1)}),\cdots,D(\boldsymbol{v}_{(m)})\big)$ for any finite integer m, it suffices to replace the mapping $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}$ defined in (B.47) associated with the marginal $\boldsymbol{v}=(\boldsymbol{v}_1^{\mathrm{T}},\boldsymbol{v}_2^{\mathrm{T}})^{\mathrm{T}}$ to a m-dimensional mapping $(\mathcal{T}_{l,\boldsymbol{v}_{(1),l}}^{(j,k)},\cdots,\mathcal{T}_{l,\boldsymbol{v}_{(m),l}}^{(j,k)})$ for each $l\in\{1,2\}$ and $(j,k)\in\mathcal{S}(l)$.

The proof is divided to four parts. In Part 1 we express D_T as a functional of point processes. Part 2 first establishes the weak limit of the empirical point process, by verifying Meyer's condition which ensures the asymptotical Poisson for the point process with the mixing sequences. Then we construct an explicit representation of the limiting process. Part 3 shows the asymptotical independence of the point processes associated with different splitting hyperplanes. In Part 4, we employ a continuous mapping theorem for the functional introduced in Part 1 to obtain the weak convergence of $D_T(\boldsymbol{v})$.

Part 1: Transformation into a functional of point processes. In this part, we will express $D_T(\boldsymbol{v})$ as a sum of transformations of point processes.

Recall that

$$D_T(\boldsymbol{v})=\sum_{l=1}^{2}\sum_{t=1}^{T}\sum_{(j,k)\in\mathcal{S}(l)}\xi_t^{(j,k)}\mathbb{1}\left\{s_l^{(j)}\left(Tq_{l,t}+\boldsymbol{Z}_{-1,l,t}^{\mathrm{T}}\boldsymbol{v}_{-1,l}\right)\le 0< s_l^{(j)}Tq_{l,t}\right\},$$

where $\xi_t^{(j,k)}=\left(\boldsymbol{\delta}_{jk,0}^{\mathrm{T}}\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{jk,0}+2\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{jk,0}\varepsilon_t\right)\{\mathbb{1}_t^{(j)}(\gamma_0)+\mathbb{1}_t^{(k)}(\gamma_0)\}.$

We now show that $D_T(\boldsymbol{v})$ can be written as a sum of functionals of some empirical point processes. For each $l\in\{1,2\}$ and $(j,k)\in\mathcal{S}(l)$, we define an empirical point process $\widehat{\mathbf{N}}_{l,T}^{(j,k)}\in M_p(E_l)$, which is the space of Radon point measures defined in Definition A.2, where $E_l=\mathbb{R}_{s_l^{(j)}}\times\mathcal{Z}_{-1,l}\times\mathbb{R}$, as

$$\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F):=\sum_{t=1}^{T}\mathbb{1}\left\{(Tq_{l,t},\boldsymbol{Z}_{-1,l,t},\xi_t^{(j,k)})\in F\right\}\ \ \text{for any} F=(F_1,F_2,F_3)\in E_l, \tag{B.46}$$

where $\mathbb{R}_{s_l^{(j)}}=(0,\infty)$ if $s_l^{(j)}=1$, and $\mathbb{R}_{s_l^{(j)}}=(-\infty,0]$ if $s_l^{(j)}=-1$. The element $\{0\}$ is excluded from the space of $\xi_t^{(j,k)}$ since $\xi_t^{(j,k)}=0$ does not affect $D_T(\boldsymbol{v})$.

For a given $\boldsymbol{v}=(\boldsymbol{v}_1^{\mathrm{T}},\boldsymbol{v}_2^{\mathrm{T}})^{\mathrm{T}}$, for each $l\in\{1,2\}$ and $(j,k)\in\mathcal{S}(l)$, we define a map $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}: M_p(E_l)\to\mathbb{R}$ such that

$$\forall\,\mathbf{N}\in M_p(E):\ \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}(\mathbf{N})=\int_{E_l}g_{l,\boldsymbol{v}_l}^{(j,k)}(x,\boldsymbol{y},z)d\mathbf{N}(x,\boldsymbol{y},z), \tag{B.47}$$

where for each $x \in \mathbb{R}_{s_l^{(j)}}, \boldsymbol{y} \in \mathcal{Z}_{-1,l}$ and $z \in \mathbb{R}$,

$$g_{l,\boldsymbol{v}_l}^{(j,k)}(x,\boldsymbol{y},z) = z \cdot \mathbb{1}\left\{s_l^{(j)}(x+\boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}_{-1,l}) \le 0 < s_l^{(j)}x\right\}.$$

Then, with (B.46) and (B.47) we can write

$$\sum_{t=1}^{T} \xi_t^{(j,k)} \mathbb{1}\left\{s_l^{(j)}\left(Tq_{l,t} + \boldsymbol{Z}_{-1,l,t}^{\mathrm{T}}\boldsymbol{v}_{-1,l}\right) \le 0 < s_l^{(j)} T q_{l,t}\right\} = \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right).$$

Consequently, $D_T(\boldsymbol{v})$ can be expressed as

$$D_T(\boldsymbol{v}) = \sum_{l=1}^{2} \sum_{(j,k)\in\mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right). \tag{B.48}$$

Part 2: Weak limit of $\widehat{\mathbf{N}}_{l,T}^{(j,k)}$. In this part, we derive the weak limit of the empirical point process $\widehat{\mathbf{N}}_{l,T}^{(j,k)}$ for each $l \in L$ and $(j,k) \in \mathcal{S}_l$ in three steps. In step 1, we calculate $\lim_{T\to\infty} \mathbb{E}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F)\right\}$ to obtain the mean measure of the limit process $\mathbf{N}_l^{(j,k)}$ required in (A.26) in Kallenberg's theorem (Lemma A.7). In the next step, we first verify Conditions (a)-(c) of Meyer's theorem (Lemma A.8), and then use it to show (A.27). The above two steps guarantee that the empirical point process $\widehat{\mathbf{N}}_{l,T}^{(j,k)}$ weakly converges to a Poisson process $\mathbf{N}_l^{(j,k)}$. In the final step, we will find an explicit representation of $\mathbf{N}_l^{(j,k)}$.

Step 1: Calculation of the limit of $\mathbb{E}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F)\right\}$.

For any $F = (F_1, F_2, F_3) \in \mathcal{E}_l$, which is the basis of relatively compact open set in E_l, where $F_1 \subset \mathbb{R}_{s_l^{(j)}}$ and $F_2 \subset \mathcal{Z}_{-1,l}, F_3 \subset \mathbb{R}$, we have

$$\begin{aligned}
&\lim_{T\to\infty} \mathbb{E}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F)\right\} \\
&= \lim_{T\to\infty} T\mathbb{P}\left\{\left(Tq_{l,t}, \boldsymbol{Z}_{-1,l,t}, \xi_t^{(j,k)}\right) \in F\right\} \\
&= \lim_{T\to\infty} T \int_{Tq\in F_1, \boldsymbol{z}\in F_2, \xi\in F_3} f_{\xi^{(j,k)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi|q,\boldsymbol{z}) f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) \mathrm{d}q\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
&\overset{(i)}{=} \lim_{T\to\infty} \int_{\tilde{q}\in F_1, \boldsymbol{z}\in F_2, \xi\in F_3} f_{\xi^{(j,k)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid \frac{\tilde{q}}{T}, \boldsymbol{z}) f_{q_l|\boldsymbol{Z}_{-1,l}}(\frac{\tilde{q}}{T} \mid \boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) \mathrm{d}\tilde{q}\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
&\overset{(ii)}{=} \int_{\tilde{q}\in F_1, \boldsymbol{z}\in F_2, \xi\in F_3} f_{\xi^{(j,k)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid 0, \boldsymbol{z}) f_{q_l|\boldsymbol{Z}_{-1,l}}(0 \mid \boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) \mathrm{d}\tilde{q}\mathrm{d}\boldsymbol{z}\mathrm{d}\xi \\
&=: \mu_l^{(j,k)}(F) < \infty,
\end{aligned} \tag{B.49}$$

where (i) is by letting $q = \tilde{q}/T$, (ii) is by the dominating convergence theorem and the continuity of $f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z})$ and $f_{\xi^{(j,k)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid q, \boldsymbol{z})$ at $q=0$, and that $\mu_l^{(j,k)}(F)<\infty$ is because of the uniform boundness of the density functions assumed in Assumption 5 and the compactness of F. The measure $\mu_l^{(j,k)}$ on $E_l = \mathbb{R}_{s_l^{(j)}} \times \mathcal{Z}_{-1,l} \times \mathbb{R}$ is defined as

$$\mu_l^{(j,k)}(\mathrm{d}q, \mathrm{d}\boldsymbol{z}, \mathrm{d}\xi) = f_{\xi^{(j,k)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid 0, \boldsymbol{z}) f_{q_l|\boldsymbol{Z}_{-1,l}}(0 \mid \boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) \mathrm{d}q\mathrm{d}\boldsymbol{z}\mathrm{d}\xi. \tag{B.50}$$

Suppose that $\mu_l^{(j,k)}$ defined above is the mean measure of the point process $\mathbf{N}_l^{(j,k)}$, then (B.49) verifies the condition (A.26) required in Lemma A.7. To verify the other condition (A.27), we use Meyer's theorem, whose requirements are listed in (a)-(c) in Lemma A.8 and are verified as follows.

Step 2: Verification of the conditions of Meyer's theorem.

To show $\lim_{T\to\infty} \mathbb{P}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F)=0\right\} = \mathbb{P}\left\{\widehat{\mathbf{N}}_{l}^{(j,k)}(F)=0\right\}$, we now employ the Meyer's theorem presented in Lemma A.8. The following notations are the same as used in Lemma A.8. For any $F=(F_1,F_2,F_3)\in E_l$ and any sample size $n\geq 1$, define the sequence of "rare" events as

$$A_t^n(F) = \mathbb{1}\left\{(nq_{l,t}, \boldsymbol{Z}_{-1,l,t}, \xi_t^{(j,k)}) \in F\right\},$$

for $1\leq t\leq n (n=1,2,\cdots)$. For any $m>0$, we take $q_m=[Lm]^q$ and $p_m=[Lm]^p$ for some $L\geq 1$ and $p\geq q\geq 1$, where $[x]$ denotes the largest interger not greater than x. Then $t_m = m(q_m+p_m) = m\left([Lm]^q+[Lm]^p\right)$. We illustrate the validity of Conditions (a)-(c) of Lemma A.8 as follows:

It is noted that Condition (a) is ensured by the condition of geometrical decaying α-mixing coefficient imposed in Assumption 1. Furthermore, Condition (b) is valid, since $q_m=[Lm]^q$ and $p_m=[Lm]^p$ for some constants $L\geq 1$ and $p\geq q>1$, leading to $p_{m+1}/p_m\to 1$ and $q_m/p_m\to 0$ as $m\to\infty$. Finally, for Condition (c), we note that

$$\begin{aligned}
t_m^2 I_{p_m} &= t_m^2 \sum_{i=1}^{p_m-i} (p_m-i)\mathbb{P}\left\{A_1^{t_m}(F)\cap A_{i+1}^{t_m}(F)\right\}\\
&\leq t_m^2 p_m \sum_{i=1}^{p_m-i} \mathbb{P}\left\{A_1^{t_m}(F)\cap A_{i+1}^{t_m}(F)\right\}\\
&\leq t_m^2 p_m \sum_{i=1}^{p_m-i} \mathbb{P}\left\{(t_m q_{l,1}\in F_1)\cap(t_m q_{l,i+1}\in F_1)\right\}\\
&\overset{(iii)}{\lesssim} t_m^2 p_m^2 \left\{\mathbb{P}\left(t_m q_{l,1}\in F_1\right)\right\}^2\\
&= t_m^2 p_m^2 \left(\int_{t_m q\in F_1} dF_{q_l}(q)\right)^2\\
&\overset{(iv)}{=} t_m^2 p_m^2 \left(\int_{t_m q\in F_1, \boldsymbol{z}\in\mathcal{Z}_{-1,l}} f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})dqd\boldsymbol{z}\right)^2\\
&\overset{(v)}{=} p_m^2 \left(\int_{q\in F_1, \boldsymbol{z}\in\mathcal{Z}_{-1,l}} f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})dqd\boldsymbol{z} + o(1)\right)^2 \overset{(vi)}{\leq} Cp_m^2
\end{aligned} \tag{B.51}$$

for some positive constant C, where $F_{q_l}(q)$ is the distribution function of $q_l = \boldsymbol{Z}_l^{\mathrm{T}}\gamma_{l0}$. In the above derivation, (iii) is from Assumption 5 (i), (iv) is by conditioning $q_{l,1}$ on $\boldsymbol{Z}_{-1,l}$, (v) is obtained via the same arguments of (i) and (ii) used in deriving (B.49), and (vi) is because $f_{q_l|\boldsymbol{Z}_{-1,l}}$ is bounded with probability 1 by Assumption 5 (ii) and the compactness of F_1. Consequently, (B.51) implies that as $m\to\infty$,

$$I_{p_m} \leq C\frac{p_m^2}{t_m^2} = C\frac{p_m^2}{m^2(p_m+q_m)^2} \leq C\frac{1}{m^2} = o\left(\frac{1}{m}\right),$$

which verifies Condition (c) in Lemma A.8.

With Conditions (a)-(c) verified and $\mathbb{P}(A_t^T(F)) = \mu_l^{(j,k)}(F)/T + o(1/T)$ as shown in deriving (B.49), for any F with $\mu_l^{(j,k)}(F) > 0$, Meyer's theorem implies that

$$\lim_{T\to\infty} \mathbb{P}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F) = 0\right\} = \lim_{T\to\infty} \mathbb{P}\left\{\text{none of } \left\{A_t^T(F)\right\}_{t=1}^T \text{ occurs}\right\}$$
$$= e^{-\mu_l^{(j,k)}(F)} = \mathbb{P}\left\{\mathbf{N}_l^{(j,k)}(F) = 0\right\}, \tag{B.52}$$

where $\mathbf{N}_l^{(j,k)}$ is a Poisson process with mean measure $\mu_l^{(j,k)}$. For F with $\mu_l^{(j,k)}(F) = 0$, (B.52) also holds, since in such case (B.49) implies $\mathbb{E}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F)\right\} \to 0$ as $T \to \infty$, which further implies that $\mathbb{P}\left\{\widehat{\mathbf{N}}_{l,T}(F) = 0\right\} = 1 = e^{-\mu_l^{(j,k)}(F)} = \mathbb{P}\left\{\mathbf{N}_l^{(j,k)}(F) = 0\right\}$. With (B.52) and (B.49), Kallenberg's theorem (Lemma A.7) implies that for each $l \in \{1,2\}$ and $(j,k) \in \mathcal{S}(l)$, $\widehat{\mathbf{N}}_{l,T}^{(j,k)} \Rightarrow \mathbf{N}_l^{(j,k)}$ in $M_p(E_l)$ as $T \to \infty$.

Step 3. Representation of $\mathbf{N}_l^{(j,k)}$.

In this step, we construct a representation of $\mathbf{N}_l^{(j,k)}$ by applying the marking theorem (Proposition 3.8 of Resnick, 2008) twice. First, let $\mathbf{N}_{1,l}^{(j,k)}$ be a canonical Poisson process on $\mathbb{R}_{s_l^{(j)}}$ on points $\{\mathcal{J}_{l,i}^{(j,k)}\}_{i=1}^\infty$ defined as

$$\mathbf{N}_{1,l}^{(j,k)}(\cdot) = \sum_{i=1}^{\infty} \mathbb{1}\left\{\mathcal{J}_{l,i}^{(j,k)} \in \cdot\right\}, \quad \mathcal{J}_{l,i}^{(j,k)} = s_l^{(j)} \sum_{n=1}^{i} \mathcal{E}_{l,n}^{(j,k)}, \tag{B.53}$$

where $\left\{\mathcal{E}_{l,n}^{(j,k)}\right\}_{n=1}^\infty$ is an i.i.d. sequence of unit-exponential variables. Then $\mathbf{N}_{1,l}^{(j,k)}$ has the mean measure $\mu_{1,l}^{(j,k)}(dq) = dq$ on $\mathbb{R}_{s_l^{(j)}}$. Let $\{\boldsymbol{Z}_{l,i}^{(j,k)}\}_{i=1}^\infty$ be an i.i.d. sequence which follows the distribution $F_{\boldsymbol{Z}_{-1,l}}$ and is independent of $\left\{\mathcal{E}_{l,n}^{(j,k)}\right\}_{n=1}^\infty$. Then the marking theorem implies the composed process

$$\mathbf{N}_{2,l}^{(j,k)}(\cdot) = \sum_{i=1}^{\infty} \mathbb{1}\left\{\left(\mathcal{J}_{l,i}^{(j,k)}, \boldsymbol{Z}_{l,i}^{(j,k)}\right) \in \cdot\right\}$$

is a Poisson process with the mean measure $\mu_{2,l}^{(j,k)}(dq, d\boldsymbol{z}) = dq \cdot f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})d\boldsymbol{z}$ on $\mathbb{R}_{s_l^{(j)}} \times \mathcal{Z}_{-1,l}$. Let $\mathcal{T}_l : (q, \boldsymbol{z}) \to (q/f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z}), \boldsymbol{z})$. Then by Proposition 3.7 in Resnick (2008),

$$\mathbf{N}_{3,l}^{(j,k)}(\cdot) = \sum_{i=1}^{\infty} \mathbb{1}\left\{\mathcal{T}_l\left(\mathcal{J}_{l,i}^{(j,k)}, \boldsymbol{Z}_{l,i}^{(j,k)}\right) \in \cdot\right\} = \sum_{i=1}^{\infty} \mathbb{1}\left\{\left(\frac{\mathcal{J}_{l,i}^{(j,k)}}{f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{l,i}^{(j,k)})}, \boldsymbol{Z}_{l,i}^{(j,k)}\right) \in \cdot\right\}$$

is a Poisson process with the mean measure

$$\mu_{3,l}^{(j,k)}(dq, d\boldsymbol{z}) = \mu_{2,l}^{(j,k)} \circ \mathcal{T}_l^{-1}(dq, d\boldsymbol{z}) = f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z})dq \cdot f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})d\boldsymbol{z} \tag{B.54}$$

on $\mathbb{R}_{s_l^{(j)}} \times \mathcal{Z}_{-1,l}$. Finally, let $F_l^{(j,k)}(\cdot|\boldsymbol{z})$ be the conditional distribution function of $\xi^{(j,k)}$ given $q_l = 0$ and $\boldsymbol{Z}_{-1,l} = \boldsymbol{z}$, which makes its density function be $f_{\xi^{(i,j)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid 0, \boldsymbol{z})$. Let $\{\xi_i^{(j,k)}\}_{i=1}^\infty$ be an i.i.d. sequence follows the conditional distribution $F_l^{(j,k)}(\cdot|\boldsymbol{Z}_{l,i}^{(j,k)})$. Then by applying again Proposition 3.7 in Resnick (2008), the composed point process

$$\mathbf{N}_l^{(j,k)}(\cdot)=\sum_{i=1}^{\infty}\mathbb{1}\left\{\left(\frac{\mathcal{J}_{l,i}^{(j,k)}}{f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{l,i}^{(j,k)})},\boldsymbol{Z}_{l,i}^{(j,k)},\xi_i^{(j,k)}\right)\in\cdot\right\} \tag{B.55}$$

is a Poisson process with the mean measure

$$\begin{aligned}\mu_l^{(j,k)}(dq,d\boldsymbol{z},d\xi)&=\mu_{3,l}^{(j,k)}(dq,d\boldsymbol{z})F_l^{(j,k)}(d\xi|\boldsymbol{z})\\&=f_{\xi^{(i,j)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi\mid 0,\boldsymbol{z})\,f_{q_l|\boldsymbol{Z}_{-1,l}}(0\mid\boldsymbol{z})\,f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})\mathrm{d}q\mathrm{d}\boldsymbol{z}\mathrm{d}\xi,\end{aligned}$$

which matches the desired mean measure (B.50).

In summary, through Steps (I)-(III) we derive that for each $l\in\{1,2\}$ and $(j,k)\in\mathcal{S}(l)$, it holds that $\widehat{\mathbf{N}}_{l,T}^{(j,k)}\Rightarrow\mathbf{N}_l^{(j,k)}$ in $M_p(E_l)$ as $T\to\infty$, where $\mathbf{N}_l^{(j,k)}$ is a Poisson point process with the representation (B.55).

Part 3: Asymptotical independence of point processes.

We now show that the empirical point processes $\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)},l\in\{1,2\},(j,k)\in\mathcal{S}(l)\right\}$ are asymptotically independent, that is, for any compact sets $\{F_l^{(j,k)}\in\mathcal{E}_l,l\in\{1,2\},(j,k)\in\mathcal{S}(l)\}$ and non-negative integers $\{k_l^{(j,k)},l\in\{1,2\},(j,k)\in\mathcal{S}(l)\}$, it holds that

$$\begin{aligned}&\mathbb{P}\left\{\bigcap_{(l,j,k)\in\mathcal{I}_s}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F_l^{(j,k)})=k_l^{(j,k)}\right)\right\}\\&\to\prod_{(l,j,k)\in\mathcal{I}_s}\frac{\exp\left(-\mu_l^{(j,k)}(F_l^{(j,k)})\right)\left\{\mu_l^{(j,k)}(F_l^{(j,k)})\right\}^{k_l^{(j,k)}}}{k_l^{(j,k)}!},\end{aligned} \tag{B.56}$$

as $T\to\infty$, where $\mathcal{I}_s$ is any subset of $\mathcal{I}=\{(l,j,k):l\in\{1,2\},(j,k)\in\mathcal{S}(l)\}$.

Suppose that $|\mathcal{I}_s|=n,1\le n\le|\mathcal{I}|$. For notational simplicity, we label the n triples $\left\{\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)},F_l^{(j,k)},k_l^{(j,k)}\right),(l,j,k)\in\mathcal{I}_s\right\}$ as $\left\{(\widehat{\mathbf{N}}_{i,T},F_i,k_i),1\le i\le n\right\}$, and define

$$\widehat{\mathbf{C}}_T=\sum_{i=1}^{n}\widehat{\mathbf{N}}_{i,T}(F_i)=\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbb{1}\left\{(Tq_{i,t},\boldsymbol{Z}_{-1,i,t},\xi_i)\in F_i\right\}=:\sum_{t=1}^{T}\widehat{\mathbf{C}}_t,\ \ \text{say.} \tag{B.57}$$

Let $A_{i,t}^T$ be the event $\{(Tq_{i,t},\boldsymbol{Z}_{-1,i,t},\xi_i)\in F_{i,t}\}$ and $B_t^T=\bigcup_{i=1}^n A_{i,t}^T$, namely B_t^T occurs if and only if at least one of $\{A_{i,t}\}_{i=1}^n$ occurs. The derivation for (B.56) includes two steps. First, we calculate $\lim_{T\to\infty}\mathbb{P}(\widehat{\mathbf{C}}_T=k)$, for which we show $\mathbb{P}(A_{i,t}^T\cap A_{i',t}^T)=O(T^{-2})$ as $T\to\infty$. In the second step, we calculate $\lim_{T\to\infty}\mathbb{P}\left\{\bigcap_{i=1}^n\left(\widehat{\mathbf{N}}_{i,T}(F_i)=k_i\right)\mid\widehat{\mathbf{C}}_T=k\right\}$ with $\sum_{i=1}^n k_i=k$, using the arguments of thinning and blocking.

Step 1. In this step, we first show that for each $1\le t\le T$, the distinct events $A_{i,t}^T$ and $A_{i',t}^T$ cannot happen together asymptotically. Suppose that

$$\begin{aligned}A_{i,t}^T&=\left\{(Tq_{l,t},\boldsymbol{Z}_{-1,l,t},\xi_t^{(j,k)})\in F_i=F_{1,i}\times F_{2,i}\times F_{3,i}\right\}\ \text{ and}\\A_{i',t}^T&=\left\{(Tq_{l',t},\boldsymbol{Z}_{-1,l',t},\xi_t^{(j',k')})\in F_{i'}=F_{1,i'}\times F_{2,i'}\times F_{3,i'}\right\},\end{aligned} \tag{B.58}$$

respectively. First, consider the case that $l = l'$ and $(j,k) \neq (j',k')$. We notice that since both (j,k) and (j',k') belong to $\mathcal{S}(l)$, then either (i) $j = k'$ and $j' = k$ or (ii) $\{j,k\} \cap \{j',k'\} = \emptyset$. Under (i) we have $\mathbb{P}(Tq_{l,t} \in F_{1,i} \cap F_{1,i'}) = 0$, since $F_{1,i} \subset \mathbb{R}_{s_l^j}$ and $F_{1,i'} \subset \mathbb{R}_{s_l^k}$, while $s_l^j = -s_l^k$. Also, since $\xi_t^{(j,k)}\xi_t^{(j',k')} = 0$ under (ii), $\mathbb{P}(\xi_t^{(j,k)} \in F_{3,i}, \xi_t^{(j',k')} \in F_{3,i'}) = 0$. In summary, $\mathbb{P}(A_{i,t}^T \cap A_{i',t}^T) = 0$ if $l = l'$ and $(j,k) \neq (j',k')$.

On the other hand, if $l \neq l'$

$$\begin{aligned}
\mathbb{P}(A_{i,t}^T \cap A_{i',t}^T) =& \mathbb{P}\left(\left\{(Tq_{l,t}, \boldsymbol{Z}_{-1,l,t}, \xi_t^{(j,k)}) \in F_i\right\} \cap \left\{(Tq_{l',t}, \boldsymbol{Z}_{-1,l',t}, \xi_t^{(j',k)}) \in F_{i'}\right\}\right) \\
\leq& \mathbb{P}\left(\{Tq_{l,t} \in F_{1,i}\} \cap \{Tq_{l',t} \in F_{1,i'}\}\right) \\
=& \mathbb{E}_{\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'}}\left\{\int_{Tq\in F_{1,i}, Tq'\in F_{1,i'}} f_{(q_l,q_{l'})|(\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'})}(q,q')dqdq'\right\} \\
=& \frac{1}{T^2}\mathbb{E}_{\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'}}\left\{\int_{\tilde{q}\in F_{1,i}, \tilde{q}'\in F_{1,i'}} f_{(q_l,q_{l'})|(\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'})}\left(\frac{\tilde{q}}{T}, \frac{\tilde{q}'}{T}\right)d\tilde{q}d\tilde{q}'\right\} \\
=& O(T^{-2}) \quad \text{as } T \to \infty. \qquad \text{(B.59)}
\end{aligned}$$

Therefore, we obtain that $\mathbb{P}(A_{i,t}^T \cap A_{i',t}^T) = O(T^{-2})$ as $T \to \infty$ if $i \neq i'$.

Note that by the inclusion-exclusion principle,

$$\begin{aligned}
\mathbb{P}(B_t^T) =& \mathbb{P}\left(\bigcup_{i=1}^n A_{i,t}^T\right) \\
=& \sum_{i=1}^n \mathbb{P}(A_{i,t}^T) + \sum_{k=2}^n (-1)^{k+1} \sum_{1\leq i_1 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1}^T \cap \cdots \cap A_{i_k}^T). \qquad \text{(B.60)}
\end{aligned}$$

Because $\mathbb{P}(A_{i_1}^T \cap \cdots \cap A_{i_k}^T) \leq \mathbb{P}(A_{i_1}^T \cap A_{i_2}^T)$, from (B.59) and (B.60) it yields that

$$\mathbb{P}(B_t^T) = \sum_{i=1}^n \mathbb{P}(A_{i,t}^T) + O(T^{-2}). \qquad \text{(B.61)}$$

From (B.49) we have

$$\mathbb{P}(A_{i,t}^T) = \mu_i(F_i)/T + o(T^{-1}), \qquad \text{(B.62)}$$

which implies that

$$\mathbb{P}(B_t^T) = \sum_{i=1}^n \mu_i(F_i)/T + o(T^{-1}). \qquad \text{(B.63)}$$

With the similar arguments used in Step 2 of Part 2, we can verify the conditions for Meyer's theorem for $\{B_t^T\}_{t=1}^T$, which delivers that for any $0 \leq k \leq T$,

$$\mathbb{P}\left\{\text{exactly } k \text{ of } \{B_t^T\}_{t=1}^T \text{ occur}\right\} \to \frac{\exp\left(-\sum_{i=1}^n \mu_i(F_i)\right)\left\{\sum_{i=1}^n \mu_i(F_i)\right\}^k}{k!},$$

as $T \to \infty$. We notice that

$$\left\{\widehat{\mathbf{C}}_T = k\right\} / \left\{\text{exactly } k \text{ of } \{B_t^T\}_{t=1}^T \text{ occur}\right\} \subset \left\{\text{for some } 1 \leq t \leq T, \widehat{\mathbf{C}}_t \geq 2\right\}$$

and

$$\sum_{t=1}^T \mathbb{P}(\widehat{\mathbf{C}}_t \geq 2) \leq \sum_{t=1}^T \sum_{1\leq i \neq i' \leq n} \mathbb{P}(A_{i_1}^T \cap A_{i_2}^T) = O(nT^{-1}),$$

where n is finite, since it is the cardinality of $\mathcal{I}_s$. Hence, we obtain

$$\mathbb{P}(\widehat{\mathbf{C}}_T = k) = \mathbb{P}\left\{\text{exactly } k \text{ of } \left\{B_t^T\right\}_{t=1}^T \text{ occur}\right\} + o(1)$$
$$\to \frac{\exp\left(-\sum_{i=1}^n \mu_i(F_i)\right)\left\{\sum_{i=1}^n \mu_i(F_i)\right\}^k}{k!}, \quad \text{as } T \to \infty. \tag{B.64}$$

Step 2. Now we turn to calculate $\mathbb{P}\left\{\bigcap_{i=1}^n \left(\widehat{\mathbf{N}}_{i,T}(F_i) = k_i\right)\right\}$. Let $k = \sum_{i=1}^n k_i$. Note that

$$\begin{aligned}
&\mathbb{P}\left\{\bigcap_{i=1}^n \left(\widehat{\mathbf{N}}_{i,T}(F_i) = k_i\right)\right\} \\
=&\mathbb{P}\left(\widehat{\mathbf{C}}_T = k\right)\mathbb{P}\left\{\bigcap_{i=1}^n \left(\widehat{\mathbf{N}}_{i,T}(F_i) = k_i\right) \mid \widehat{\mathbf{C}}_T = k\right\} \\
=&\mathbb{P}\left(\widehat{\mathbf{C}}_T = k\right)\left[\mathbb{P}\left\{\bigcap_{i=1}^n \left(k_i \text{ of } \left\{A_{i,t}^T\right\}_{t=1}^T \text{ are assigned}\right) \mid k \text{ of } \left\{B_t^T\right\}_{t=1}^T \text{ occur}\right\} + o(1)\right] \\
=:&P_{1,T} \times P_{2,T} + o(1), \ \text{ say.}
\end{aligned}$$

For $P_{1,T}$, by (B.64) we have

$$P_{1,T} \to \frac{\exp\left(-\sum_{i=1}^n \mu_i(F_i)\right)\left\{\sum_{i=1}^n \mu_i(F_i)\right\}^{\sum_{i=1}^n k_i}}{\left(\sum_{i=1}^n k_i\right)!}, \tag{B.65}$$

as $T \to \infty$. We now proceed to obtain the limits of $P_{2,T}$ by the blocking argument as in Meyer (1973).

Specifically, for any positive integer m, partition the observation indices into consecutive blocks of p_m and q_m alternately, where p_m and q_m are the same as those in Step (2) of Part 2, beginning with the initial block $\{1, \cdots, p_m\}$. Let P_m and Q_m denote those indices falling into size p_m and q_m blocks, respectively, and $t_m = m(p_m + q_m)$. Let $I_t^{i,t_m} = \left\{A_{i,t}^{t_m} \text{ happens if } B_t^{t_m} \text{ happens}\right\}$. According to (B.62) and (B.63),

$$\begin{aligned}
\mathbb{P}(I_t^{i,t_m}) =&\mathbb{P}(A_{i,t}^{t_m} | B_t^{t_m}) = \frac{\mathbb{P}(A_{i,t}^{t_m} \cap B_t^{t_m})}{\mathbb{P}(B_t^{t_m})} = \frac{\mathbb{P}(A_{i,t}^{t_m})}{\mathbb{P}(B_t^{t_m})} \\
=&\frac{\mu_i(F_i)}{\sum_{i=1}^n \mu_i(F_i)} + o(1) =: p_i + o(1), \ \text{ say,}
\end{aligned}$$

as $m \to \infty$.

Let $\mathcal{G}_k = \{G_k = \{j_s\}_{s=1}^k : 1 \le j_1 \le \cdots \le j_k \le t_m\}$ be the collection of the subsets of $\{1, \cdots, t_m\}$ with the cardinality k. Then,

$$\left\{k \text{ of } \left\{B_t^{t_m}\right\}_{t=1}^T \text{ occur}\right\} = \cup_{G_k \in \mathcal{G}_k}\{B_t^{t_m} \text{ occur iff } t \in G_k\}, \tag{B.66}$$

where "iff" is short for "if and only if". For each $G_k = \{j_s\}_{s=1}^k \in \mathcal{G}_k$, let

$$H(G_k) = \{(H_i = \{j_{i_s}\}_{s=1}^{k_i})_{i=1}^n : \cup_{i=1}^n H_i = G_k \text{ and } H_i \cap H_{i'} = \emptyset \text{ if } i \ne i'\}$$

be the collection of all possible n-partitions of G_k with each segment H_i containing k_i indices of G_k. Then we note that $|H(G_k)| = k!/(\prod_{i=1}^n k_i!)$.

$$\mathbb{P}\{\cap_{i=1}^n (k_i \text{ of } \{A_{i,t}^{t_m}\}_{t=1}^{t_m} \text{ are assigned}) \mid B_t^{t_m} \text{ occur iff } t \in G_k\}$$

$$= \sum_{(H_i)_{i=1}^n \in H(G_k)} \mathbb{P}(\cap_{i=1}^n \cap_{s=1}^{k_i} I_{j_{i_s}}^{i,t_m}). \tag{B.67}$$

By inspecting the proof of Theorem 1 of Meyer (1973), we find that if k of $\{B_t^{t_m}\}_{t=1}^{t_m}$ happens, then asymptotically all the k indices lie in separate blocks in P_m, implying that any $|j-j'| \geq q_m$ for any $j \neq j' \in G_k$. Therefore, for large enough m, we have

$$\left| \mathbb{P}(\cap_{i=1}^n \cap_{s=1}^{k_i} I_{j_{i_s}}^{i,t_m}) - \prod_{i=1}^n \prod_{s=1}^{k_i} \mathbb{P}(I_{j_{i_s}}^{i,t_m}) \right| \leq k\alpha_{t_m}(q_m),$$

by applying the definition of the α-mixing coefficients repeatedly for k times. Since $|H(G_k)| = k!/(\prod_{i=1}^n k_i!)$ and $\mathbb{P}(I_{j_{i_s}}^{i,t_m}) = p_i + o(1)$, we obtain

$$\left| \sum_{\mathcal{C}_*} \mathbb{P}(\bigcap_{i\in[n], t\in[k_i]} I_{j_t}^{i,t_m}) - \frac{k!}{\prod_{i=1}^n k_i!} \prod_{i=1}^n p_i^{k_i} + o(1) \right| \leq k \frac{k!}{\prod_{i=1}^n k_i!} \alpha_{t_m}(q_m) = o(1), \tag{B.68}$$

where the last equality is due to that k is a given integer and $\alpha_{t_m}(q_m) \to 0$ as $m \to \infty$. Combining (B.67) and (B.68) leads to

$$\mathbb{P}\{\cap_{i=1}^n (k_i \text{ of } \{A_{i,t}^{t_m}\}_{t=1}^{t_m} \text{ are assigned}) \mid B_t^{t_m} \text{ occur iff } t \in G_k\} = \frac{k!}{\prod_{i=1}^n k_i!} \prod_{i=1}^n p_i^{k_i} + o(1),$$

for each $G_k \in \mathcal{G}_k$. This together with (B.66) yields that

$$P_{2,t_m} = \frac{k!}{\prod_{i=1}^n k_i!} \prod_{i=1}^n p_i^{k_i} + o(1), \quad \text{as } m \to \infty,$$

where $P_{2,t_m} = \mathbb{P}\{\cap_{i=1}^n (k_i \text{ of } \{A_{i,t}^{t_m}\}_{t=1}^{t_m} \text{ are assigned}) \mid k \text{ of } \{B_t^{t_m}\}_{t=1}^{t_m} \text{ occur}\}$. Since for any T, there exists a m such that $T \in [t_m, t_{m+1})$, the above result implies that

$$P_{2,T} \to \frac{k!}{\prod_{i=1}^n k_i!} \prod_{i=1}^n p_i^{k_i} + o(1) = \frac{(\sum_{i=1}^n k_i)!}{\prod_{i=1}^n k_i!} \prod_{i=1}^n \left\{ \frac{\mu_i(F_i)}{\sum_{i=1}^n \mu_i(F_i)} \right\}^{k_i}, \quad \text{as } T \to \infty, \tag{B.69}$$

since $k = \sum_{i=1}^n k_i$ and $p_i = \mu_i(F_i)/(\sum_{i=1}^n \mu_i(F_i))$. Combining (B.65) with (B.69) yields that

$$\mathbb{P}\left\{ \bigcap_{i=1}^n \left(\widehat{\mathbf{N}}_{i,T}(F_i) = k_i \right) \right\} = P_{1,T} P_{2,T} + o(1) = \prod_{i=1}^n \frac{\exp\left(-\mu_i(F_i)\right) \{\mu_i(F_i)\}^{k_i}}{k_i!} + o(1),$$

which proves (B.56) and implies that $\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}, l \in \{1,2\}, (j,k) \in \mathcal{S}(l)\right)$ are asymptotically independent. These together with Part 2 conclude that $\widehat{\mathbf{N}}_{l,T}^{(j,k)} \Rightarrow \mathbf{N}_l^{(j,k)}$ in $M_p(E_l)$ as $T \to \infty$, where $\left(\mathbf{N}_l^{(j,k)}, l \in \{1,2\}, (j,k) \in \mathcal{S}(l)\right)$ are independent Poisson point processes with the representation (B.55).

Part 4: Continuous mapping.

In this part, we show that $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right) \xrightarrow{d} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_l^{(j,k)}\right)$ as $T \to \infty$. If $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}(\cdot)$ is a continuous functional in $M_p(E_l)$, then it follows by the continuous mapping thoerem. To show that $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}(\cdot)$ is continuous mapping from $M_p(E_l)$ to $\mathbb{R}$, we use Proposition 3.13 in

Resnick (2008), which requires $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}(\cdot)$ has a compact support. Therefore, we use a truncation argument. Recall that for any $\mathbf{N} \in M_p(E)$,

$$\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}(\mathbf{N}) = \int_{E_l} g_{l,\boldsymbol{v}_l}^{(j,k)}(x, \boldsymbol{y}, z) d\mathbf{N}(x, \boldsymbol{y}, z),$$

where $x \in \mathbb{R}_{s_l^{(j)}}, \boldsymbol{y} \in \mathcal{Z}_{-1,l}, z \in \mathbb{R}$, and

$$g_{l,\boldsymbol{v}_l}^{(j,k)}(x, \boldsymbol{y}, z) = z \cdot \mathbb{1}\left\{s_l^{(j)}(x + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}_{-1,l}) \le 0 < s_l^{(j)}x\right\}.$$

Therefore, the support of $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}$ is $\mathcal{Q}_l^{(j)} \times \mathcal{Z}_{-1,l} \times \mathbb{R}$, where $\mathcal{Q}_l^{(j)} = \{q : s_l^{(j)}(q + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}_{-1,l}) \le 0 < s_l^{(j)}q$ for some $\boldsymbol{y} \in \mathcal{Z}_{-1,l}\}$, which is compact since $\mathcal{Z}_{-1,l}$ is compact. For any $M > 0$, we let $E_{l,M} = \left\{(x, \boldsymbol{y}, z) : x \in \mathbb{R}_{s_l^{(j)}}, \boldsymbol{y} \in \mathcal{Z}_{-1,l}, |z| < M\right\}$, which is a compact set. Let

$$R_T = \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right) = \int_{E_l} g_l^{(j,k)}(x, \boldsymbol{y}, z) d\widehat{\mathbf{N}}_{l,T}^{(j,k)}(x, \boldsymbol{y}, z),$$

$$R_{T,M} = \int_{E_{l,M}} g_l^{(j,k)}(x, \boldsymbol{y}, z) d\widehat{\mathbf{N}}_{l,T}^{(j,k)}(x, \boldsymbol{y}, z),$$

$$R_{0,M} = \int_{E_{l,M}} g_l^{(j,k)}(x, \boldsymbol{y}, z) d\mathbf{N}_{l,T}^{(j,k)}(x, \boldsymbol{y}, z) \text{ and}$$

$$R_0 = \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right) = \int_{E_l} g_l^{(j,k)}(x, \boldsymbol{y}, z) d\mathbf{N}_{l,T}^{(j,k)}(x, \boldsymbol{y}, z).$$

In the following, we show in three steps that (i) $R_{T,M} \xrightarrow{d} R_{0,M}$ for any fixed $M > 0$ as $T \to \infty$ by the continuous mapping theorem, (ii) $\lim_{M\to\infty} \limsup_{T\to\infty} \mathbb{P}\{|R_T - R_{T,M}| > \varepsilon\} \to 0$ for any $\varepsilon > 0$, and (iii) $R_{0,M} \xrightarrow{d} R_0$ as $M \to \infty$. Then by Theorem 4.2 of Billingsley (1968), $R_T \xrightarrow{d} R_0$ as $T \to \infty$.

Step (1). For any fixed $M > 0$, let $\mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)}(\mathbf{N}) = \int_{E_{l,M}} g_{l,\boldsymbol{v}_l}^{(j,k)}(x, \boldsymbol{y}, z) d\mathbf{N}(x, \boldsymbol{y}, z)$ for any $\mathbf{N} \in M_p(E)$. By Proposition 3.13 in Resnick (2008), if any sequence $\mathbf{N}_n \Rightarrow \mathbf{N}$, then the points of $\mathbf{N}_n$ locating in $E_{l,m}$ converge to that of $\mathbf{N}$ locating in $E_{l,m}$. Since restricted on $E_{l,M}$, the function $g_l^{(j,k)}(x, \boldsymbol{y}, z)$ has a compact support but is discontinuous at $x = 0$ or $x + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}_{-1,l} = 0$, the functional $\mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)}$ is continuous except on

$$\mathcal{D}(\mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)}) = \{\mathbf{N} \in M_p(E) : x_i^{\mathbf{N}} = 0 \text{ or } x_i^{\mathbf{N}} + (\boldsymbol{y}_i^{\mathbf{N}})^{\mathrm{T}}\boldsymbol{v}_{-1,l} = 0 \text{ for some } i \ge 1\},$$

where $(x_i^{\mathbf{N}}, \boldsymbol{y}_i^{\mathbf{N}}, z_i^{\mathbf{N}}, i \ge 1)$ denote the points of $\mathbf{N}$. Since

$$\mathbb{P}\left\{\mathbf{N}_l^{(j,k)} \in \mathcal{D}(\mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)})\right\} = \mathbb{P}\left\{\exists\, i, J_{l,i}^{(j,k)} = 0 \text{ or } J_{l,i}^{(j,k)} + (\boldsymbol{Z}_{l,i}^{(j,k)})^{\mathrm{T}}\boldsymbol{v}_{-1,l} = 0\right\} = 0 \quad \text{(B.70)}$$

and $J_{l,i}^{(j,k)}$ is absolutely continuous, we have

$$R_{T,M} = \mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right) \xrightarrow{d} \mathcal{M}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_l^{(j,k)}\right) = R_{0,M}, \quad \text{(B.71)}$$

for any fixed $M > 0$ as $T \to \infty$, by the continuous mapping theorem.

Step (2). Next, we show that

$$\lim_{M\to\infty} \limsup_{T\to\infty} \mathbb{P}\{|R_T - R_{T,M}| > \varepsilon\} \to 0, \quad \text{(B.72)}$$

for any $\varepsilon > 0$. For notational simplicity, we denote $\xi_t = \xi_t^{(j,k)}$, $q_t = q_{l,t}$, $\boldsymbol{Z}_{-1,t} = \boldsymbol{Z}_{-1,l,t}$, and suppose $s_l^{(j)} = 1$ without loss of generality. Then, for any $M > 0$,

$$
\begin{aligned}
|R_T - R_{T,M}| \leq & \sum_{t=1}^{T} \left\{ |\xi_t| \mathbb{1}\left(|\xi_t| \geq M\right) \mathbb{1}\left(Tq_t + \boldsymbol{Z}_{-1,t}^{\mathrm{T}} \boldsymbol{v}_{-1,l} \leq 0 < Tq_t\right) \right\} \\
=: & \sum_{t=1}^{T} G_t(M), \quad \text{say.}
\end{aligned}
\tag{B.73}
$$

Since

$$
\begin{aligned}
\mathbb{E}\left\{ |\xi_t| \mathbb{1}\left(|\xi_t| \geq M\right) \mid \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0 \right\} \leq & \{\mathbb{E}(|\xi_t|^2 | \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0)\}^{1/2} \left\{ \mathbb{P}(|\xi_t| > M | \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0) \right\}^{1/2} \\
\leq & \{\mathbb{E}(|\xi_t|^2 | \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0)\}^{1/2} \frac{\{\mathbb{E}(|\xi_t|^2 | \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0)\}^{1/2}}{M} \\
= & O_p\left(M^{-1}\right)
\end{aligned}
\tag{B.74}
$$

almost surely, where the first inequality is via Cauchy-Schwarz inequality and the second is by Markov inequality, provided $\mathbb{E}(|\xi_t|^2 | \boldsymbol{Z}_{l,t}^{\mathrm{T}} \boldsymbol{\gamma} = 0) < \infty$ for $\boldsymbol{\gamma}$ in a neighborhood of $\boldsymbol{\gamma}_{l,0}$, which is ensured by Assumption 4 (iv). Using (B.74) and with the similar arguments as in the proof of Lemma A.2 (i), we can show that $\mathbb{E}\left\{G_t(M)\right\} = O\left((MT)^{-1}\right)$. Therefore,

$$
\mathbb{E}|R_T - R_{T,M}| \leq \sum_{t=1}^{T} \mathbb{E}\left\{G_t(M)\right\} = O\left(M^{-1}\right),
$$

for any T and M, which implies (B.72) by Markov inequality.

Step (3). Next, we show that $R_0 = R_{0,M} + o_p(1)$. We notice that

$$
R_0 - R_{0,M} = \sum_{i=1}^{\infty} \left[\xi_i^{(j,k)} \mathbb{1}\left(|\xi_i^{(j,k)}| > M \right) \mathbb{1}\left\{ J_{l,i}^{(j,k)} + \left(\boldsymbol{Z}_{l,i}^{(j,k)} \right)^{\mathrm{T}} \boldsymbol{v}_{-1,l} \leq 0 < J_{l,i}^{(j,k)} \right\} \right].
$$

Let $Z_{\max} = \max\left\{ -\left(\boldsymbol{Z}_{l,i}^{(j,k)} \right)^{\mathrm{T}} \boldsymbol{v}_{-1,l}, \boldsymbol{Z}_{l,i}^{(j,k)} \in \mathcal{Z}_{-1,l} \right\}$, which is bounded since both $\mathcal{Z}_{-1,l}$ and the space of $\boldsymbol{v}_{-1,l}$ are compact. This means that $Z_{\max} < \infty$. Since $f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{-1,l})$ is uniformly bounded by some constant, say F_l, by Assumption 5 (ii), the event

$$
\frac{\mathcal{J}_{l,i}^{(j,k)}}{f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{l,i}^{(j,k)})} + \left(\boldsymbol{Z}_{l,i}^{(j,k)} \right)^{\mathrm{T}} \boldsymbol{v}_{-1,l} \leq 0 < \frac{\mathcal{J}_{l,i}^{(j,k)}}{f_{q|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{l,i}^{(j,k)})} \equiv J_{l,i}^{(j,k)}
$$

implies $0 \leq \mathcal{J}_{l,i}^{(j,k)} \leq F_l Z_{\max}$. Therefore,

$$
|R_0 - R_{0,M}| \leq \sum_{i=1}^{\infty} \left\{ \xi_i^{(j,k)} \mathbb{1}\left(|\xi_i^{(j,k)}| > M \right) \mathbb{1}(0 \leq \mathcal{J}_{l,i}^{(j,k)} \leq F_l Z_{\max}) \right\}.
\tag{B.75}
$$

Note that $\mathcal{J}_{l,i}^{(j,k)} = \sum_{n=1}^{i} \mathcal{E}_{l,n}^{(j,k)}$ where $\{\mathcal{E}_{l,n}^{(j,k)}\}_{n=1}^{\infty}$ is an i.i.d. sequence of unit-exponential variables, and $\{\xi_i^{(j,k)}\}_{i=1}^{\infty}$ be an i.i.d. sequence follows the conditional distribution $F_l^{(j,k)}(\cdot|\boldsymbol{Z}_{l,i}^{(j,k)})$, that is independent to $\{\mathcal{E}_{l,n}^{(j,k)}\}_{n=1}^{\infty}$. Hence, $\mathcal{P}(t) = \sum_{i=1}^{N(t)} \xi_i^{(j,k)} \mathbb{1}(|\xi_i^{(j,k)}| > M)$ for $t \geq 0$

is compound Poisson process with the jump size $\xi_i^{(j,k)}\mathbb{1}(|\xi_i^{(j,k)}| > M)$, where $N(t) = \sum_{i=1}^{\infty} \mathbb{1}(\mathcal{J}_{l,i}^{(j,k)} \le t)$ is a homogeneous Poisson process with rate 1. Therefore, we have

$$
\begin{aligned}
\mathbb{E}\{|R_0 - R_{0,M}|\} &\overset{(i)}{\le} F_l Z_{\max}\mathbb{E}\left\{\xi_i^{(j,k)}\mathbb{1}\left(|\xi_i^{(j,k)}| > M\right)\right\}, \\
&\overset{(ii)}{\le} F_l Z_{\max}\sqrt{\mathbb{E}\left\{\left(\xi_i^{(j,k)}\right)^2\right\}}\sqrt{\mathbb{P}\left(|\xi_i^{(j,k)}| > M\right)} \\
&\overset{(iii)}{\le} F_l Z_{\max}\mathbb{E}\left\{\left(\xi_i^{(j,k)}\right)^2\right\}/M \to 0, \text{ as } M \to \infty,
\end{aligned}
\tag{B.76}
$$

where (i) is from Wald's identity (Wald, 1944), (ii) is from Cauchy-Schwartz's inequality and (iii) is from Markov's inequality. Because of the above result, we obtain $R_{0,M} = R_0 + o_p(1)$, which further implies that $R_{0,M} \overset{d}{\to} R_0$.

Through the three steps we have shown that (i) $R_{T,M} \overset{d}{\to} R_{0,M}$ for any fixed $M > 0$ as $T \to \infty$ by the continuous mapping theorem, (ii) $\lim_{M\to\infty}\limsup_{T\to\infty}\mathbb{P}\{|R_T - R_{T,M}| > \varepsilon\} \to 0$ for any $\varepsilon > 0$, and (iii) $R_{0,M} \overset{d}{\to} R_0$ as $M \to \infty$. Therefore, by applying Theorem 4.2 of Billingsley (1968), $R_T \overset{d}{\to} R_0$ as $T \to \infty$, i.e., $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right) \overset{d}{\to} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right)$. Because in Part 3 it is shown that $\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}, l \in \{1,2\}, (j,k) \in \mathcal{S}(l)\right)$ are asymptotically independent, we conclude that

$$
\sum_{l=1}^{L}\sum_{(j,k)\in\mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}\right) \overset{d}{\to} \sum_{l=1}^{L}\sum_{(j,k)\in\mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right), \tag{B.77}
$$

as $T \to \infty$, which concludes the proof. □

B.8. Proof of Lemma B.3.

PROOF. The proof for this lemma adapts that in Chernozhukov and Hong (2004). First, we decompose $D_T(\boldsymbol{v}) = \sum_{l=1}^{2}\sum_{(j,k)\in\mathcal{S}(l)} D_T^{(j,k)}(\boldsymbol{v}_l)$, where $\boldsymbol{v}_l \in \mathbb{R}^{d_l}$ for $l = 1, 2$, and

$$
D_T^{(j,k)}(\boldsymbol{v}_l) = \sum_{t=1}^{T} \xi_t^{(j,k)}\mathbb{1}\left\{s_l^{(j)}\left(Tq_{l,t} + \boldsymbol{Z}_{-1,l,t}^{\mathrm{T}}\boldsymbol{v}_{-1,l}\right) \le 0 < s_l^{(j)}Tq_{l,t}\right\}.
$$

It is sufficient to show that $D_T^{(j,k)}(\boldsymbol{v}_l)$ is stochastic equi-lower-semicontinuous for each $l \in \{1,2\}$ and $(j,k) \in \mathcal{S}(l)$. Without loss of generality, we take $l = 1, j = 1, k = 2$, since the other cases can be proved in the same way. To simplify notations, let $\tilde{\boldsymbol{v}} = \boldsymbol{v}_1, \tilde{q}_t = q_{1,t}, \tilde{\boldsymbol{Z}}_t = \boldsymbol{Z}_{-1,l,t}, \tilde{\xi}_t = \xi_t^{(j,k)}$ and $\tilde{D}_T(\tilde{\boldsymbol{v}}) = D_T^{(1,2)}(\boldsymbol{v}_1)$. With the above notations,

$$
\tilde{D}_T(\tilde{\boldsymbol{v}}) = \sum_{t=1}^{T} \tilde{\xi}_t\mathbb{1}\left\{\left(T\tilde{q}_t + \tilde{\boldsymbol{Z}}_t^{\mathrm{T}}\tilde{\boldsymbol{v}}_{-1}\right) \le 0 < T\tilde{q}_t\right\}.
$$

Because $\tilde{D}_T(\tilde{\boldsymbol{v}})$ is a piece-wise constant function, which implies that $\tilde{D}_T(\tilde{\boldsymbol{v}})$ takes discrete values in each compact open set, it suffices to show that for any compact set $B \subset \mathbb{R}^{d_1}$ and any $\delta > 0$, there are open neighborhoods $V(\tilde{\boldsymbol{v}}_1), \cdots, V(\tilde{\boldsymbol{v}}_k)$ of some $\tilde{\boldsymbol{v}}_1, \cdots, \tilde{\boldsymbol{v}}_k$ such that $B \subset \cup_{j=1}^{k} V(\tilde{\boldsymbol{v}}_j)$ and

$$
\mathbb{P}\left(\cup_{j=1}^{k}\left\{\inf_{\boldsymbol{v}\in V(\tilde{\boldsymbol{v}}_j)} \tilde{D}_T(\boldsymbol{v}) \le \tilde{D}_T(\tilde{\boldsymbol{v}}_j)\right\}\right) < \delta, \tag{B.78}
$$

for sufficiently large T.

Let $\{\mathcal{Z}_\phi(\tilde{\boldsymbol{z}}_j), j \leq J(\phi)\}$ be $J(\phi)$ closed equal-sized cubes with the side-length ϕ such that $\mathcal{Z}_{-1,1}$, the support of the distribution of $\boldsymbol{Z}_{-1,1}$, can be covered by the union of $\{\mathcal{Z}_\phi(\tilde{\boldsymbol{z}}_j), j \leq J(\phi)\}$, and the center of the cube $\mathcal{Z}_\phi(\tilde{\boldsymbol{z}}_j)$ is denoted as $\tilde{\boldsymbol{z}}_j$. Construct $(2m+1)J(\phi)$ sets $\{V_{kj}, l = -m, \cdots, m, j \leq J(\phi)\} \subset \mathbb{R}^{d_1}$ as

$$V_{kj} = \{\tilde{\boldsymbol{v}} \in R^{d_1} : \nu_k - \psi < \tilde{\boldsymbol{z}}^{\mathrm{T}} \tilde{\boldsymbol{v}}_{-1} < \nu_k + \psi, \forall \tilde{\boldsymbol{z}} \in \mathcal{Z}_\phi(\tilde{\boldsymbol{z}}_j)\},$$

where $\psi > 0$ and $\nu_k = k\psi$ for $k \in \{-m, \cdots, 0, \cdots, m\}$. Since $\mathcal{Z}_{-1,1}$ is a compact set, which implies that the range of $\tilde{\boldsymbol{z}}^{\mathrm{T}} \tilde{\boldsymbol{v}}_{-1}$ is compact for any compact B, the union of $\{V_{kj}\}$ can cover B by selecting sufficiently large m.

Because $\tilde{D}_T(\tilde{\boldsymbol{v}})$ is piece-wise constant, a discontinuity of $\tilde{D}_T(\tilde{\boldsymbol{v}})$ can potentially occur in $\cup_j V_{kj}$ only if there exist $\boldsymbol{v}_* \in \cup_j V_{kj}$ and $(T\tilde{q}_{t_*}, \tilde{\boldsymbol{Z}}_{t_*})$ for some $t_* \in \{1, \cdots, T\}$ such that $T\tilde{q}_{t_*} = \tilde{\boldsymbol{Z}}_{t_*}^{\mathrm{T}} \boldsymbol{v}_*$, satisfying $\nu_k - \psi \leq T\tilde{q}_{t_*} \leq \nu_k + \psi$. If there is such $(T\tilde{q}_{t_*}, \tilde{\boldsymbol{Z}}_{t_*})$, we say $\tilde{D}_T(\tilde{\boldsymbol{v}})$ has a breakpoint in $\cup_j V_{kj}$. Define $\mathcal{B}_T = |\{t : 0 < T\tilde{q}_t < \bar{Z}\}|$, where $\bar{Z} = \sup_{\boldsymbol{z} \in \mathcal{Z}_{-1,1}, \boldsymbol{v} \in B} \boldsymbol{z}^{\mathrm{T}} \boldsymbol{v}$, as an upper bound on the number of breakpoint of $\tilde{D}_T(\tilde{\boldsymbol{v}})$ in B, and let $\mathcal{B} = |\{i : \mathcal{J}_i < \bar{Z}\}|$, where $\mathcal{J}_i = \sum_{m=1}^{i} \mathcal{E}_i$ with $\{\mathcal{E}_i\}_{i=1}^{\infty}$ being i.i.d. unit exponentially distributed variables. Because the point process induced by $\{T\tilde{q}_t, t \in \{1, \cdots, T\} : \tilde{q}_t > 0\}$ weakly converges to the point process induced by $\{\mathcal{J}_i\}_{i=1}^{\infty}$ as shown in the proof of Lemma B.2, by the continuous mapping theorem, we have $\mathcal{B}_T \xrightarrow{d} \mathcal{B}$. Therefore, the number of breakpoints $\mathcal{B}_T = O_p(1)$.

We now show the breakpoints are separated, namely, no more than one breakpoint can happen in $\cup_j V_{kj}$ with probability arbitrarily close to one if ψ is sufficiently small. Let A_k to be the event that $\tilde{D}_T(\tilde{\boldsymbol{v}})$ has more than one breakpoint in $\cup_j V_{kj}$. Relabelling $\{T\tilde{q}_t, t \in \{1, \cdots, T\} : \tilde{q}_t > 0\}$ as $\{\mathcal{J}_{iT}\}$ such that $0 < \mathcal{J}_{1T} \leq \mathcal{J}_{2T} \leq \cdots$. Then, because the point process corresponding to $\{\tilde{q}_t, t \in \{1, \cdots, T\} : \tilde{q}_t > 0\}$ converges weakly to that corresponding to $\{J_i\}_{i=1}^{\infty}$, according to continuous mapping theorem, for any finite $k \leq T$,

$$(\mathcal{J}_{1T}, \cdots, \mathcal{J}_{kT}) \xrightarrow{d} (\mathcal{J}_1, \cdots, \mathcal{J}_k). \tag{B.79}$$

Define A_k to be the event that $\tilde{D}_T(\tilde{\boldsymbol{v}})$ has more than two break-points in $\cup_j V_{kj}$. Since $\cup_k A_k$ happens if at least one pair $(\mathcal{J}_{(i-1)T}, \mathcal{J}_{iT})$ for some $i \leq \mathcal{B}_T$ satisfying $\mathcal{J}_{iT} - \mathcal{J}_{(i-1)T} < 2\psi$, we have

$$\begin{aligned}
\limsup_{T\to\infty} \mathbb{P}(\cup_k A_k) \leq& \limsup_{T\to\infty} \mathbb{P}\left\{\min_{2\leq i\leq \mathcal{N}_T} (\mathcal{J}_{iT} - \mathcal{J}_{(i-1)T}) < 2\psi\right\} \\
&\leq \limsup_{T\to\infty} \mathbb{P}\left\{\min_{2\leq i\leq K} (\mathcal{J}_{iT} - \mathcal{J}_{(i-1)T}) < 2\psi\right\} + \mathbb{P}(\mathcal{B}_T > K) \\
&\overset{(i)}{\leq} \mathbb{P}\left\{\min_{2\leq i\leq K} (\mathcal{J}_{i} - \mathcal{J}_{(i-1)}) < 2\psi\right\} + \mathbb{P}(\mathcal{B} > K) \\
&\overset{(ii)}{\leq} \delta/2,
\end{aligned} \tag{B.80}$$

where (i) is by (B.79), (ii) is by taking K sufficiently large such that $\mathbb{P}(\mathcal{B} > K) < \delta/4$, and taking ψ sufficiently small such that $\mathbb{P}\left\{\min_{2\leq i\leq K} (\mathcal{J}_{i} - \mathcal{J}_{(i-1)}) < 2\psi\right\} < \delta/4$. The latter is possible since by definition $\mathcal{J}_i - \mathcal{J}_{(i-1)} = \mathcal{E}_{i-1}$ has independent unit exponential distribution. Hence

$$\mathbb{P}\left\{\min_{2\leq i\leq K} (\mathcal{J}_{i} - \mathcal{J}_{(i-1)}) < 2\psi\right\} = \mathbb{P}\left\{\min_{2\leq i\leq K} \mathcal{E}_{i-1} < 2\psi\right\} = 1 - e^{-2\psi(K-1)},$$

which converges to 0 as $\psi(K-1) \to 0$.

We construct centers $\tilde{\boldsymbol{v}}_{kj}$ in V_{kj} such that

$$\nu_k - \psi < \tilde{\boldsymbol{z}}^{\mathrm{T}} \tilde{\boldsymbol{v}}_{-1,kj} < \nu_k - \psi + \eta, \ \ \forall \tilde{\boldsymbol{z}} \in \mathcal{Z}_\phi(\tilde{\boldsymbol{z}}_j),$$

where η will be set sufficiently small in the next step. Depending on η, we will set ϕ sufficiently small as well to satisfy the above constraints. Note that the left-side hand of (B.78) can be decomposed as

$$\mathbb{P}\left(\cup_{j,k} \left\{ \inf_{\boldsymbol{v} \in V_{kj}(\tilde{\boldsymbol{v}}_{kj})} \tilde{D}(\boldsymbol{v}) \le \tilde{D}_T(\tilde{\boldsymbol{v}}_{kj}) \right\} \right) < \limsup_{T\to\infty} \mathbb{P}\{B(\eta)\} + \limsup_{T\to\infty} \mathbb{P}(\cup_k A_k), \quad \text{(B.81)}$$

where $B(\eta)$ is the event that $\{\mathcal{J}_{iT}, i \le K\}$ are separated, and at least one of $\mathcal{J}_{iT} \in [\nu_{k_i} - \psi, \nu_{k_i} - \psi + \eta]$ for some $k_i \in \{1, \cdots, K\}$. The bound (B.81) holds because $\tilde{D}(\boldsymbol{v})$ can only jump if $\mathcal{J}_{iT}$ increases, implying that

$$\cup_{j,k} \left\{ \inf_{\boldsymbol{v} \in V_{kj}(\tilde{\boldsymbol{v}}_{kj})} \tilde{D}(\boldsymbol{v}) \le \tilde{D}_T(\tilde{\boldsymbol{v}}_{kj}) \right\} \cap (\cup_k A_k)^c = B(\eta).$$

Due to (B.79) and the fact that $\{\mathcal{J}_i\}$ have a bounded density, we have

$$\limsup_{T\to\infty} \mathbb{P}\{B(\eta)\} = O(K\eta) < \delta/2, \quad \text{(B.82)}$$

by choosing η sufficiently small. Combining (B.80)–(B.82) completes the proof for Lemma B.3. □

B.9. Proof of Lemma B.4.

PROOF. Let $f_n(\boldsymbol{v}) = \sum_{i=0}^{\infty} a_{ni} \mathbb{1}(\boldsymbol{v} \in F_{ni})$, where $\{a_{ni} \in \mathbb{R}\}_{i=1}^{\infty}$ are jump sizes and $\{F_{ni} \in \mathbb{R}^d\}_{i=0}^{\infty}$ are non-overlapping level sets. Let $\tilde{f}_n(\boldsymbol{v}) = \sum_{i=0}^{\infty} i \mathbb{1}(\boldsymbol{v} \in F_{ni})$ be the associated jump process. Note $\tilde{f}_n$ has a jump with size 1 at the boundary of each level set F_{ni}. Let the limiting piece-wise constant function be $f_0(\boldsymbol{v}) = \sum_{i=0}^{\infty} a_{0i} \mathbb{1}(\boldsymbol{v} \in F_{0i})$, whose associated jump process be $\tilde{f}_0(\boldsymbol{v}) = \sum_{i=0}^{\infty} i \mathbb{1}(\boldsymbol{v} \in F_{0i})$. For any compact set E, we define $I_n(E) = \{i : F_{n,i} \cap E \notin \emptyset\}$ and $I_0(E) = \{i : F_{0,i} \cap E \notin \emptyset\}$ be the index sets for the level sets of f_n and f_0 that have intersections with E, respectively. Let the argmin sets of f_n and f_0 on the compact set E be G_n and G_0, respectively.

Step 1. Convergence of level sets.

We first show the convergence of the level sets $\{F_{ni}, i \in I_n(E)\}$ to $\{F_{0i}, i \in I_n(E)\}$, using the epi-convergence of the jump processes $\{\tilde{f}_n\}$. For any interior point $\boldsymbol{v}_i$ in $F_{0,i}$, which is a continuous point of f_0 and $\tilde{f}_0$, let $\varepsilon_0 > 0$ be any sufficiently small constant such that $\mathcal{N}(\boldsymbol{v}_i; \varepsilon_0) \subset F_{0,i}$. By a similar argument to that used in the proof of Lemma B.3, there exists some $\boldsymbol{v}_i' \in \mathcal{N}(\boldsymbol{v}_i; \varepsilon_0)$ such that f_n and $\tilde{f}_n$ are asymptotically equi-lower semicontinuous at $\boldsymbol{v}_i'$. Since we have $\{\tilde{f}_n\}$ epi-converge to $\tilde{f}_0$, by applying Theorem 7.10 of Rockafellar and Wets (1998), we have the pointwise convergence $\tilde{f}_n(\boldsymbol{v}_i') \to \tilde{f}_0(\boldsymbol{v}_i')$ as $n \to \infty$. Let $\partial F_{0,i} = \bar{F}_{0,i} \setminus F_{0,i}^{\circ}$ be the boundary of $F_{0,i}$ for each $i \in I_0(E)$, where sets $\bar{F}_{0,i}$ and $F_{0,i}^{\circ}$ are the closure and interior of $F_{0,i}$, respectively. Then we can find infinitely many $\boldsymbol{v} \in F_{0,i}^{\circ}$ with $\inf_{\boldsymbol{v}' \in \partial F_{0,i}} \|\boldsymbol{v} - \boldsymbol{v}'\| = \varepsilon_0$, such that $\tilde{f}_n(\boldsymbol{v}) \to \tilde{f}_0(\boldsymbol{v}) = i$. This together with the connectness of $F_{n,i}$ implies that $F_{0,i}^{\varepsilon_0 -} \subset F_{n,i}$, where $F_{0,i}^{\varepsilon_0 -} = \bar{F}_{0,i} \setminus \left\{\boldsymbol{v} \in \bar{F}_{0,i}, \inf_{\boldsymbol{v}' \in \partial F_{0,i}} \|\boldsymbol{v} - \boldsymbol{v}'\| \le \varepsilon_0\right\}$. Similarly we can find infinitely many $\boldsymbol{v} \in E \setminus \bar{F}_{0,i}$ with $\inf_{\boldsymbol{v}' \in \partial F_{0,i}} \|\boldsymbol{v} - \boldsymbol{v}'\| = \varepsilon_0$, such that $\tilde{f}_n(\boldsymbol{v}) \to \tilde{f}_0(\boldsymbol{v}) \ne i$. It means that for each sufficiently large n, there is a jump of $\tilde{f}_n$ in the region $\{\boldsymbol{v} \in E : \inf_{\boldsymbol{v}' \in \partial F_{0,i}} \|\boldsymbol{v} - \boldsymbol{v}'\| < \varepsilon_0\}$ around the boundary of $F_{0,i}$ for each $i \in I_0(E)$. Therefore, we obtain $|I_n(E)| \to |I_0(E)|$ as $n \to \infty$. Also, since ε_0 can be taken arbitrarily close to 0 and

$\mu(\partial F_{0,i}) = 0$, where μ is the Lebesgue measure, for each $1 \le i \le N_0$ and each sufficiently large n, it holds that $|\mathbb{1}(\boldsymbol{v} \in F_{n,i}) - \mathbb{1}(\boldsymbol{v} \in F_{0,i})| \to 0$ almost surely under the Lebesgue measure.

Step 2. Convergence of the argmin level set.

We now show the minimzed set of f_n converges to that of f_0. Let F_{0,i_*} be the level set on which f_0 attains its minimum. By the condition that $\xi_{0,i} \ne \xi_{0,j}$ if $i \ne j$, such i_* is unique. Hence $F_{0,i_*} = G_0$. Note that unlike the proof of Theorem 1 of Knight (1999), applying Theorem 7.33 of Rockafellar and Wets (1998) can only ensure $G_n \subset G_0$ asymptotically. However, such result can be strengthened by utilizing the piece-wise constant property of f_n and f_0. From the above paragraph, it has been shown that each level set $F_{n,i}$ of f_n converges to $F_{0,i}$ of f_0. As argued in the previous paragraph, for each $i \in I_0(E)$ we can find $\boldsymbol{v}_i \in F_{0,i}$, such that f_0 is continuous at $\boldsymbol{v}_i$ and $\{f_n\}$ are asymptotically equi-lower semicontinuous at $\boldsymbol{v}_i$. Hence the epi-convergence of $\{f_n\}$ to f_0 implies the pointwise convergence of $f_n(\boldsymbol{v}_i) \to f_0(\boldsymbol{v}_i) = a_{i0}$, meaning that $a_{n,i} \to a_{0,i}$ for each $i \in I_0(E)$. Because $\{a_{0,i}, i \in I_0(E)\}$ is uniquely minimized at $i = i_*$, for any $\epsilon > 0$ such that for any sufficiently large n, we have $a_{n,i_*} < a_{n,i} + \epsilon$, which means that the minimizer level set G_n of f_n is unique and equals to $F_{n,i*}$. This together with the result in the previous paragraph implies $|\mathbb{1}(\boldsymbol{v} \in G_n) - \mathbb{1}(\boldsymbol{v} \in G_0)| \to 0$ for almost surely $\boldsymbol{v}$. The desired result (B.27) in Lemma B.4 then follows by applying the dominated convergence theorem. □

B.10. Proof of Lemma B.5.

PROOF. By (B.24), $W_T(\boldsymbol{u})$ can be written as $W_T(\boldsymbol{u}) = \sum_{i=1}^4 (\boldsymbol{u}_i^\top \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top \mathbb{1}\{\boldsymbol{Z} \in R_i(\gamma_0)\}]\boldsymbol{u}_i - 2\boldsymbol{u}_i^\top H_{i,T})$, where

$$H_{i,T} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \boldsymbol{X}_t \varepsilon_t \mathbb{1}\{\boldsymbol{Z}_t \in R_i(\gamma_0)\}.$$

Let $\boldsymbol{a} \in \mathbb{R}^p$ with $\|\boldsymbol{a}\| = 1$. Also let $H_{\boldsymbol{a},i,T} = \boldsymbol{a}^\top H_{i,T}$ and $\sigma^2_{\boldsymbol{a},i} = \boldsymbol{a}^\top \Sigma_i \boldsymbol{a}^\top$, where $\Sigma_i = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top \varepsilon^2 \mathbb{1}\{\boldsymbol{Z} \in R_i(\gamma_0)\}]$. Then with Assumptions 1.(ii) and 3.(i), by the martingale central limit theorem (Hall and Heyde, 1980), it holds that $\sigma^{-1}_{\boldsymbol{a},i} H_{\boldsymbol{a},i,T} \xrightarrow{d} N(0,1)$. Hence, by the Cramer-Wold device, we obtain $H_{i,T} \xrightarrow{d} N(0, \Sigma_i)$, which implies that $W_T(\boldsymbol{u}) \xrightarrow{d} W(\boldsymbol{u})$. Since the stochastic component of $W_T(\boldsymbol{u})$ is linear in $\boldsymbol{u}$, the stochastic equicontinuity of $W_T(\boldsymbol{u})$ can be trivially proved. Hence $W_T \xrightarrow{d} W$ in $\ell^\infty(\mathbb{B})$.

We now show the asymptotic independence between $W_T(\boldsymbol{u})$ and $D_T(\boldsymbol{v})$. For independence observations, it can be readily proved by the characteristic function approach used in Yu (2012), which however may not be suitable for the dependence case. In this proof, we employ the device established in Hsing (1995), which can be used to show the asymptotic independence between the extreme type and sum type statistics for the mixing sequences. We notice that while the original results in that paper were for univariate random variables, they can be extended to multivariate cases with essentially the same proof.

As in Part 1 of the proof of Lemma B.2, we write $D_T(\boldsymbol{v}) = \sum_{l=1}^2 \sum_{(j,k)\in\mathcal{S}(l)} \mathcal{T}^{(j,k)}_{l,\boldsymbol{v}_l}\left(\widehat{\mathbf{N}}^{(j,k)}_{l,T}\right)$, where $\widehat{\mathbf{N}}^{(j,k)}_{l,T}$ is a point process defined in (B.46) and $\mathcal{T}^{(j,k)}_{l,\boldsymbol{v}_l}$ is a continuous functional. Therefore, it suffices to show the asymptotic independence between $\widehat{\mathbf{N}}^{(j,k)}_{l,T}$ and $H_{\boldsymbol{a},i,T}$ for any $\boldsymbol{a} \in \mathbb{R}^p$ with $\|\boldsymbol{a}\| = 1$, $l \in \{1, 2\}$, $(j,k) \in \mathcal{S}(l)$ and $i \in \{1, \cdots, 4\}$. If one has

$$\mathbb{P}\left\{H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t} \le x, \widehat{\mathbf{N}}^{(j,k)}_{l,T}(F_i) = k_i, 1 \le i \le s\right\}$$

$$=\mathbb{P}\left(H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t}\le x\right)\mathbb{P}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F_i)=k_i, 1\le i\le s\right\}+o(1), \tag{B.83}$$

for any $x\in\mathbb{R}$, positive integer s, non-negative integers $\{k_i\}_{i=1}^s$, and non-overlapping sets $\{F_i=(F_{1i},F_{2i},F_{3i})\}_{i=1}^s\in\mathcal{E}(l)$, where $\mathcal{E}(l)$ is the basis of relatively compact open set in E_l as used in Part 2 of Section B.7, then $\widehat{\mathbf{N}}_{l,T}^{(j,k)}$ is independent with $H_{\boldsymbol{a},i,T}$.

First, we verify Conditions (2.1) and (2.2) of Hsing (1995). Let $\zeta_t=(Tq_{l,t},\boldsymbol{Z}_{-1,l,t},\zeta_t^{(j,k)})$ and $B_{T,i}=F_i$ for $1\le i\le s$ and $B_{T,s+1}=\cap_{i=1}^s B_{T,i}^c$. Then,

$$\begin{aligned}\limsup_{T\to\infty}T\mathbb{P}(\zeta_t\notin B_{T,s+1})\le&\limsup_{T\to\infty}\sum_{i=1}^s T\mathbb{P}(\zeta_t\in B_{T,i})\\ =&\sum_{i=1}^s\mu_l^{(j,k)}(B_{T,i})<\infty,\end{aligned} \tag{B.84}$$

where the equality is due to (B.49). Hence, Condition (2.1) of Hsing (1995) is ensured. In addition, Condition (2.2) of the same paper also holds, since

$$\begin{aligned}&\lim_{l\to\infty}\limsup_{T\to\infty}\mathbb{P}\left\{\cup_{t=l}^T(\zeta_t\notin B_{T,s+1})|\zeta_1\notin B_{T,s+1}\right\}\\ =&\lim_{l\to\infty}\limsup_{T\to\infty}\frac{\mathbb{P}\left\{\cup_{t=l}^T(\zeta_t\notin B_{T,s+1})\cap(\zeta_1\notin B_{T,s+1})\right\}}{\mathbb{P}(\zeta_1\notin B_{T,s+1})}\\ =&\lim_{l\to\infty}\limsup_{T\to\infty}\frac{O(T^{-2})}{O(T^{-1})}=0,\end{aligned}$$

where the denominator part is from (B.84) and the numerator is derived in the same way as in (B.51).

We now show the desired (B.83) with similar arguments as in Theorem 2.2 of Hsing (1995). Let $\tilde{\zeta}_T=(\zeta_1,\cdots,\zeta_T)^\intercal$. For any $\tilde{A}=(A_1,\cdots,A_T)$, the notation $\tilde{\zeta}_T\in\tilde{A}$ stands for $\zeta_t\in A_t$ for each $1\le t\le T$, and $\tilde{\zeta}_T\notin\tilde{A}$ otherwise. Let $\tilde{\mathcal{B}}_T=\{\tilde{B}=(B_1,\cdots,B_T)\}$, where each $B_t\in\{B_{T,1},\cdots,B_{T,s+1}\}$ for each $1\le t\le T$. Also we let

$$\tilde{\mathcal{B}}_T'=\left\{\tilde{B}\in\tilde{\mathcal{B}}_T:\sum_{t=1}^T\mathbb{1}(B_t=B_{T,i})=k_i,\ \text{ for } 1\le i\le s\right\}.$$

By such constructions, we have

$$\begin{aligned}&\left\{\cup_{\tilde{B}\in\tilde{\mathcal{B}}}(\tilde{\zeta}_T\in\tilde{B})\right\}\cap\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(B_{T,i})=k_i,1\le i\le s\right\}\\ \overset{(i)}{=}&\cup_{\tilde{B}\in\tilde{\mathcal{B}}'}(\tilde{\zeta}_T\in\tilde{B})\overset{(ii)}{=}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(B_{T,i})=k_i,1\le i\le s\right\}.\end{aligned}$$

Also, we note that (i) implies that

$$\begin{aligned}0\le&\mathbb{P}\left\{H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t}\le x,\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F_i)=k_i,1\le i\le s\right\}\\ &-\mathbb{P}\left\{H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t}\le x,\cup_{\tilde{B}\in\tilde{\mathcal{B}}'}(\tilde{\zeta}_T\in\tilde{B})\right\}\\ \le&\mathbb{P}\left\{\cap_{\tilde{B}\in\tilde{\mathcal{B}}}(\tilde{\zeta}_T\notin\tilde{B})\right\}.\end{aligned} \tag{B.85}$$

With the fact that the events $\{(\tilde{\zeta}_T\notin\tilde{B})\}_{\tilde{B}\in\tilde{\mathcal{B}}'}$ are disjoint, repeatedly applying Theorem 2.1 of Hsing (1995) leads to

$$\mathbb{P}\left\{H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t}\le x,\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F_i)=k_i,1\le i\le s\right\}$$

$$\stackrel{(iii)}{=} \sum_{\tilde{B}\in\tilde{\mathcal{B}}'} \mathbb{P}\left\{H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t} \le x, \tilde{\zeta}_T \in \tilde{B}\right\} + o(1)$$

$$\stackrel{(iv)}{=} \mathbb{P}(H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t} \le x) \sum_{\tilde{B}\in\tilde{\mathcal{B}}'} \mathbb{P}\left(\tilde{\zeta}_T \in \tilde{B}\right) + o(1)$$

$$= \mathbb{P}(H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t} \le x)\mathbb{P}\left\{\cup_{\tilde{B}\in\tilde{\mathcal{B}}'}(\tilde{\zeta}_T \in \tilde{B})\right\} + o(1)$$

$$\stackrel{(v)}{=} \mathbb{P}(H_{\boldsymbol{a},i,T}/\sigma_{\boldsymbol{a},i,t} \le x)\mathbb{P}\left\{\widehat{\mathbf{N}}_{l,T}^{(j,k)}(F_i) = k_i, 1 \le i \le s\right\} + o(1),$$

where (iii) is because of (B.85) and (2.3) in Theorem 2.1 of Hsing (1995), (iv) is implied by (2.4) in the same theorem, and (v) is due to the equivalence relationship (ii). Hence, (B.83) is now verified. Since the above derivations hold for any $\boldsymbol{a} \in \mathbb{R}^p$ with $\|\boldsymbol{a}\| = 1$, $l \in \{1, 2\}, (j, k) \in \mathcal{S}(l)$ and $i \in \{1, \cdots, 4\}$, we complete the proof for the asymptotic independence between $W_T(\boldsymbol{u})$ and $D_T(\boldsymbol{v})$. □

APPENDIX C: PROOF FOR SECTION 4 AND ADDITIONAL ALGORITHMS

C.1. Proof of Theorem 4.1. The following proof is for Theorem 4.1 on the validity of the MIQP.

PROOF. Let the criterion function of the MIQP be

$$\mathbb{V}_T(\boldsymbol{\ell}) = \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \sum_{k=1}^{4}\sum_{i=1}^{p} X_{t,i}\ell_{k,i,t}\right)^2,$$

where $\boldsymbol{\ell} = \{\ell_{k,i,t} : k = 1, \cdots, 4, i = 1, \cdots, p, t = 1, \cdots, T\}$. The constraints of the MIQP are

1. $\boldsymbol{\beta}_k \in \mathcal{B}, \quad \boldsymbol{\gamma}_j \in \Gamma,$
2. $g_{j,t} \in \{0, 1\}, \quad I_{k,t} \in \{0, 1\},$
3. $L_i \le \beta_{k,i} \le U_i,$
4. $(g_{j,t} - 1)(M_{j,t} + \epsilon) < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\boldsymbol{\gamma}_j \le g_{j,t}M_{j,t},$

5.(i). $I_{k,t}L_i \le \ell_{k,i,t} \le I_{k,t}U_i,$

5.(ii). $L_i(1 - I_{k,t}) \le \beta_{k,i} - \ell_{k,i,t} \le U_i(1 - I_{k,t}),$

6. $I_{k,t} \le s_j^{(k)}g_{j,t} + (1 - s_j^{(k)})/2, \ \ I_{k,t} \ge \sum_{j=1}^{2}\left\{s_j^{(k)}g_{j,t} + (1 - s_j^{(k)})/2\right\} - 1,$

for $k = 1, \cdots, 4, j = 1, 2, i = 1, \cdots, p$ and $t = 1, \cdots, T$. Define $\boldsymbol{g} = \{g_{j,t} : j = 1, 2, t = 1, \cdots, T\}$, $\boldsymbol{\mathcal{I}} = \{I_{k,t} : k = 1, \cdots, 4, t = 1, \cdots, T\}$. The solution of the MIQP is denoted as $(\bar{\boldsymbol{\beta}}, \bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{\mathcal{I}}}, \bar{\boldsymbol{\ell}}) = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{g},\boldsymbol{\mathcal{I}},\boldsymbol{\ell}} \mathbb{V}_T(\boldsymbol{\ell})$.

To prove the theorem, it suffices to show that (i) $\mathbb{M}_T(\bar{\boldsymbol{\theta}}) = \mathbb{V}_T(\bar{\boldsymbol{\ell}})$, where $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\gamma}}^{\mathrm{T}}, \bar{\boldsymbol{\beta}}^{\mathrm{T}})^{\mathrm{T}}$; (ii) $\mathbb{V}_T(\bar{\boldsymbol{\ell}}) \ge \mathbb{M}_T(\widehat{\boldsymbol{\theta}})$; and (iii) $\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) \ge \mathbb{V}_T(\bar{\boldsymbol{\ell}})$.

Proof of (i): It is sufficient to show that

$$\left(Y_t - \sum_{k=1}^{4}\boldsymbol{X}_t^{\mathrm{T}}\bar{\boldsymbol{\beta}}_k\mathbb{1}_j(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\bar{\boldsymbol{\gamma}}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\bar{\boldsymbol{\gamma}}_2)\right)^2 = \left(Y_t - \sum_{k=1}^{4}\sum_{i=1}^{p} X_{t,i}\bar{\ell}_{k,i,t}\right)^2. \tag{C.1}$$

We show that $\bar{\ell}_{k,i,t} = \bar{\beta}_{k,i}\mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\bar{\gamma}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\bar{\gamma}_2)$. If $\mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\bar{\gamma}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\bar{\gamma}_2) = 1$, then by Constraints 2 and 6, we have $I_{k,t} = 1$, which implies that $\bar{\ell}_{k,i,t} = \bar{\beta}_{k,i}$. If $\mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\bar{\gamma}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\bar{\gamma}_2) = 0$, then by Constraints 2 and 6 we have $I_{k,t} = 0$, which implies that $\bar{\ell}_{k,i,t} = 0$. Combining the two cases verifies $\bar{\ell}_{k,i,t} = \bar{\beta}_{k,i}\mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\bar{\gamma}_1, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\bar{\gamma}_2)$ for each k, i, t, which implies (C.1).

Proof of (ii): Note that

$$V_T(\bar{\boldsymbol{\ell}}) = \mathbb{M}_T(\bar{\boldsymbol{\theta}}) \geq \min_{\boldsymbol{\beta}\in\mathcal{A},\boldsymbol{\gamma}\in\mathcal{G}} \mathbb{M}_T(\boldsymbol{\theta}) = \mathbb{M}_T(\widehat{\boldsymbol{\theta}}),$$

where the first equality is by (i) and the last equality is by the definition of $\widehat{\boldsymbol{\theta}}$.

Proof of (iii): Define $\widehat{\ell}_{k,i,t} = \widehat{\beta}_{k,i}\widehat{I}_{k,t}$, where $\widehat{I}_{k,t} = \prod_{j=1}^{2} s_j^{(k)}\widehat{g}_{j,t}$ and $\widehat{g}_{j,t} = \mathbb{1}\{\boldsymbol{Z}_t^{\mathrm{T}}\widehat{\gamma}_j > 0\}$. Then by definition $\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) = \mathbb{V}_T(\widehat{\boldsymbol{\ell}})$, where $\widehat{\boldsymbol{\ell}} = \{\widehat{\ell}_{k,i,t}\}$. If $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{d}}, \widehat{\boldsymbol{\ell}})$ satisfy Constraints 1-6 above, then by the definition of $\bar{\boldsymbol{\ell}}$, we have $\mathbb{V}_T(\widehat{\boldsymbol{\ell}}) \geq \mathbb{V}_T(\bar{\boldsymbol{\ell}})$ and hence, (iii) can be verified. Constraints 1-3 are ensured by the definitions. For Constraint 4, note that if $\boldsymbol{Z}_{j,t}^{\mathrm{T}}\widehat{\gamma}_j > 0$, then by definition $\widehat{g}_{j,t} = I(\boldsymbol{Z}_{j,t}^{\mathrm{T}}\widehat{\gamma}_j > 0) = 1$. Constraint 4 becomes $0 < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\widehat{\gamma}_j \leq M_{j,T} = \sup_{\boldsymbol{\gamma}\in\Gamma_j}|\boldsymbol{Z}_{j,t}^{\mathrm{T}}\gamma|$, which is satisfied. When $\boldsymbol{Z}_{j,t}^{\mathrm{T}}\widehat{\gamma}_j \leq 0$, then $\widehat{g}_{j,t} = 0$. Condition 4 becomes $-M_{j,t} - \epsilon < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\widehat{\gamma}_j \leq 0$, which holds for any $\epsilon > 0$. Hence, Condition 4 is verified. For Condition 5, note that if $\widehat{I}_{k,t} = 1$, then $\widehat{\ell}_{k,i,t} = \widehat{\beta}_{k,i}$ by its definition, which meet the requirement in Constraint 5 (i) and (ii). Otherwise, if $\widehat{I}_{k,t} = 0$, then $\widehat{\ell}_{k,i,t} = 0$, and Constraints 5 (i) and (ii) are satisfied. For Constraint 6, it is ready to verify that

$$\sum_{j=1}^{2}\left\{s_j^{(k)}\widehat{g}_{j,t} + (1 - s_j^{(k)})/2\right\} - 1 \leq \prod_{j=1}^{2} s_j^{(k)}\widehat{g}_{j,t} \leq s_j^{(k)}\widehat{g}_{j,t} + (1 - s_j^{(k)})/2,$$

for any $\widehat{g}_{1,t}, \widehat{g}_{2,t} \in \{0, 1\}$ and $s_1^{(k)}, s_2^{(k)} \in \{-1, 1\}$. In summary, $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{d}}, \widehat{\boldsymbol{\ell}})$ satisfies Constraints 1-6, implying that

$$\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) = \mathbb{V}_T(\widehat{\boldsymbol{\ell}}) \geq \mathbb{V}_T(\bar{\boldsymbol{\ell}}),$$

which proves (iii). Combining parts (i), (ii) and (iii), we obtain $\mathbb{M}_T(\widehat{\boldsymbol{\theta}}) = \mathbb{M}_T(\bar{\boldsymbol{\theta}})$, which completes the proof of Theorem 4.1. □

C.2. Block coordinate descent. The MIQP presented in Section 4 of the main paper may be slow when the dimension of $\boldsymbol{X}_t$ and the sample size T are large. As an alternative, we present a block coordinate descent (BCD) algorithm.

Iterate the following two steps until $\max_{1\leq k\leq 4}\|\widehat{\boldsymbol{\beta}}_k^{s+1} - \widehat{\boldsymbol{\beta}}_k^{s}\| < \eta$.
Step 1. For each given $\widehat{\boldsymbol{\beta}}^s$, solve the following mixed integer linear programming (MILP) problem:

$$\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{g},\boldsymbol{\mathcal{I}},\boldsymbol{\ell}} \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{4}\left\{(\boldsymbol{X}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k^s)^2 - 2Y_t\boldsymbol{X}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k^s\right\} I_{k,t} \tag{C.2}$$

$$\text{subject to} \begin{cases} \gamma_j \in \Gamma_j,\ g_{j,t} \in \{0,1\},\ I_{k,t} \in \{0,1\}; \\ (g_{j,t} - 1)(M_{j,t} + \epsilon) < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\gamma_k \leq g_{j,t}M_{j,t},\ I_{k,t}L_i \leq \ell_{k,i,t} \leq I_{k,t}U_i; \\ I_{k,t} \leq s_l^{(k)}g_{l,t} + \dfrac{1 - s_l^{(k)}}{2},\ I_{k,t} \geq \displaystyle\sum_{l=1}^{2}\left(s_l^{(k)}g_{l,t} + \frac{1 - s_l^{(k)}}{2}\right) + 1 - L, \end{cases} \tag{C.3}$$

for $k = 1, \cdots, 4, j = 1, 2, i = 1, \cdots, d_x$ and $t = 1, \cdots, T$. Let the solution be $\widehat{\gamma}^{s+1}$.

Step 2. For the given $\widehat{\gamma}^{s+1}$, obtain

$$\widehat{\boldsymbol{\beta}}_k^{s+1} = [\mathbb{E}_T\{\boldsymbol{X}_t\boldsymbol{X}_t^{\mathrm{T}}\mathbb{1}(\boldsymbol{Z}_t \in R_k(\widehat{\gamma}^{s+1})\}]^{-1}\mathbb{E}_T\{Y_t\boldsymbol{X}_t\mathbb{1}(\boldsymbol{Z}_t \in R_k(\widehat{\gamma}^{s+1})\}.$$

REMARK C.1. The advantages of the BCD compared with the MIQP are that the optimization with respect to $\boldsymbol{\gamma}$ in each iteration is a linear programming instead of quadratic programming, and that for $\boldsymbol{\beta}_k$ in each iteration has a close form solution. Therefore, the BCD can significantly reduce computation cost. However, unlike the MIQP presented in the main paper, there is no theoretical guarantee for the global optimality of the solutions of the BCD. For the BCD, the specification of the initial value $\widehat{\boldsymbol{\theta}}^0$ is important. In practice, it can be obtained from a grid search procedure, or we can use the output of the MIQP after several iterations as the initial value for the BCD.

The following Table S1 reports the comparison between the joint MIQP algorithm proposed in Section 4 of the main paper and the block coordinate descent algorithm presented in Section C.2. The sample was generated according to

$$Y_t = \sum_{k=1}^{4} \boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\beta}_{k0}\mathbb{1}_k(\boldsymbol{Z}_{1,t}^{\mathrm{T}}\gamma_{10}, \boldsymbol{Z}_{2,t}^{\mathrm{T}}\gamma_{20}) + \varepsilon_t \quad t = 1, \cdots, T$$

where $\boldsymbol{X}_t = (\tilde{\boldsymbol{X}}_t^{\mathrm{T}}, 1)^{\mathrm{T}}$ with $\tilde{\boldsymbol{X}}_t = (X_{1,t}, \cdots, X_{p-1,t})^{\mathrm{T}}$ and $\boldsymbol{Z}_{j,t} = (\tilde{\boldsymbol{Z}}_{j,t}^{\mathrm{T}}, 1)^{\mathrm{T}}$ with $\tilde{\boldsymbol{Z}}_{j,t} = (Z_{j,1,t}, \cdots, Z_{j,d-1,t})^{\mathrm{T}}$ for $j = 1, 2$, and the residuals $\varepsilon_t = \sigma(\boldsymbol{X}_t, \boldsymbol{Z}_t)e_t$ with $\sigma(\boldsymbol{X}_t, \boldsymbol{Z}_t) = 1 + 0.1X_{1,t}^2 + 0.1Z_{1,1,t}^2$ and $\{e_t\}_{t=1}^T$ being generated independently from the standard normal distribution and being independent of $\{\boldsymbol{X}_t, \boldsymbol{Z}_t\}_{t=1}^T$. Let $\boldsymbol{V}_t = (\tilde{\boldsymbol{X}}_t^{\mathrm{T}}, \tilde{\boldsymbol{Z}}_{1,t}^{\mathrm{T}}, \tilde{\boldsymbol{Z}}_{2,t}^{\mathrm{T}})^{\mathrm{T}}$. We generated $\{\boldsymbol{V}_t\}_{t=1}^T \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \Sigma_V)$, where $\Sigma_V = (\sigma_{ij})_{i,j=1,\cdots,7}$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.1$ if $i \neq j$. We considered two sets of dimensions for $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$. For $p = 4$ and $d = 3$, the regression coefficients of the four regimes were $\boldsymbol{\beta}_{10} = (1, 1, 1, 1)^{\mathrm{T}}, \boldsymbol{\beta}_{20} = (-3, -2, -1, 0), \boldsymbol{\beta}_{30} = (0, 1, 3, -1)^{\mathrm{T}}$ and $\boldsymbol{\beta}_{40} = (2, -1, 0, 2)^{\mathrm{T}}$, and the two boundary coefficients $\boldsymbol{\gamma}_{10} = (1, -1, 0)^{\mathrm{T}}$ and $\boldsymbol{\gamma}_{20} = (1, 1, 0)^{\mathrm{T}}$, respectively. For $p = 10$ and $d = 6$, the regression coefficients of the four regimes were $\boldsymbol{\beta}_{10} = (1, 1, 1, 1, 1, 0, \cdots, 0)^{\mathrm{T}}, \boldsymbol{\beta}_{20} = (-3, -2, -1, 1, 0, \cdots, 0), \boldsymbol{\beta}_{30} = (0, 1, 3, -1, 1, 0, \cdots, 0)^{\mathrm{T}}$ and $\boldsymbol{\beta}_{40} = (2, -1, 0, 2, 1, 0, \cdots, 0)^{\mathrm{T}}$, and the two boundary coefficients $\boldsymbol{\gamma}_{10} = (1, -1, -1, -1, 0, 0)^{\mathrm{T}}$ and $\boldsymbol{\gamma}_{20} = (1, 1, 1, 1, 0, 0)^{\mathrm{T}}$, respectively. The simulation experimented four sample sizes: $\{200, 400, 800, 1600\}$, and the experiments were repeated 500 times for each sample size. The initial values for the BCD were set as the outputs of the MIQP after $5\log(T)$ iterations. The stopping criterion parameter was specified as $\eta = 10^{-4}$.

Table S1 shows that the estimation errors of both $\boldsymbol{\gamma}_0$ and $\boldsymbol{\beta}_0$ obtained with the BCD were slightly larger than those with the MIQP, while their discrepancies were shrinking as T increased. The running time of the BCD, on the other hand, was significantly shorter than that of the joint MIQP, especially when the dimensions and sample sizes were large, because of the reasons we discussed above and in the main paper. Therefore, it is advocated to use the iterative BCD for large dimensions and sample sizes. However, it should also be noted that it is crucial to choose a good initial value for the BCD for its success. In the above simulations, we used the outputs of the MIQP after several iterations to ensure the quality of the initial values, as poor initial values can lead to large estimation errors.

C.3. MIQP for the three-regime models. The MIQP algorithms are not only suitable for solving the LS problem of the four-regime segmented regression but can also be extended to other segmented regressions. In Section 7.2 of the main paper we have reported simulation

TABLE S1
Empirical average estimation errors $\|\gamma_0 - \widehat{\gamma}\|_2$ and $\|\beta_0 - \widehat{\beta}\|_2$ (multiplied by 10), and running time (in second) obtained with the joint MIQP algorithm and the block coordinate descent (BCD) algorithm. The numbers inside the parentheses are the standard errors of the simulated averages.

T	$p=4,d=3$						$p=10,d=6$					
	MIQP			BCD			MIQP			BCD		
	$\widehat{\gamma}$	$\widehat{\beta}$	Time	$\widehat{\gamma}$	$\widehat{\beta}$	Time	$\widehat{\gamma}$	$\widehat{\beta}$	Time	$\widehat{\gamma}$	$\widehat{\beta}$	Time
200	1.00	6.02	65.3	1.13	6.03	7.7	3.15	10.93	149.1	3.31	11.17	10.0
	(0.59)	(1.15)	(11.3)	(0.62)	(1.14)	(1.2)	(1.83)	(2.40)	(15.3)	(1.88)	(2.49)	(1.9)
400	0.51	4.08	364.1	0.59	4.11	14.3	1.59	7.78	583.6	1.66	7.85	23.9
	(0.31)	(0.76)	(27.9)	(0.28)	(0.79)	(3.1)	(0.92)	(1.51)	(40.1)	(0.98)	(1.63)	(6.4)
800	0.25	2.84	1157.7	0.27	2.85	48.9	0.78	5.20	1817.3	0.82	5.31	69.0
	(0.15)	(0.49)	(53.2)	(0.14)	(0.47)	(7.0)	(0.41)	(0.82)	(89.4)	(0.43)	(0.84)	(11.4)
1600	0.13	2.00	2792.2	0.14	2.00	162.4	0.38	3.61	4502.9	0.40	3.67	208.1
	(0.07)	(0.37)	(162.0)	(0.07)	(0.38)	(12.1)	(0.19)	(0.57)	(217.5)	(0.18)	(0.58)	(21.4)

results under segmented models with less than four regimes to compare the four-regime estimation under misspecifications with the estimation with correctly specified models, where the corresponding MIQPs for less than four regimes models were applied. In this subsection, we present MIQP formulations for the three-regime models with and without intersections. The MIQP for the two-regime model was proposed in Lee et al. (2021).

(i) MIQP for the three-regime model with non-intersected boundaries.

Let $\boldsymbol{g} = \{g_{j,t} : j = 1,2, t = 1,\cdots,T\}$, $\boldsymbol{\mathcal{I}} = \{I_{k,t} : k = 1,2,3, t = 1,\cdots,T\}$ and $\boldsymbol{\ell} = \{\ell_{k,i,t} : k = 1,2,3, i = 1,\cdots,p, t = 1,\cdots,T\}$. Consider solving the following problem

$$\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{g},\boldsymbol{\mathcal{I}},\boldsymbol{\ell}} \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \sum_{k=1}^{3}\sum_{i=1}^{p} X_{i,t}\ell_{k,i,t}\right)^2,$$

$$\text{subject to} \begin{cases} \boldsymbol{\beta}_k \in \mathcal{B}, \quad \boldsymbol{\gamma}_j \in \Gamma, \quad g_{j,t} \in \{0,1\}, \quad I_{k,t} \in \{0,1\}, \quad L_i \le \beta_{k,i} \le U_i; \\ (g_{j,t}-1)(M_{j,t}+\epsilon) < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\boldsymbol{\gamma}_j \le g_{j,t}M_{j,t}; \\ g_{j,t}L_i \le \ell_{j,i,t} \le g_{j,t}U_i, \quad I_{2,t}L_i \le \ell_{2,i,t} \le I_{2,t}U_i; \\ L_i(1-g_{j,t}) \le \beta_{k,i} - \ell_{j,i,t} \le U_i(1-g_{j,t}); \\ L_i(1-I_{2,t}) \le \beta_{2,i} - \ell_{2,i,t} \le U_i(1-I_{2,t}); \\ I_{2,t} \le g_{1,t}, \quad I_{2,t} \le 1-g_{2,t}, \quad I_{2,t} \ge g_{1,t} - g_{2,t}, \end{cases}$$

for $k = 1,2,3, j = 1,2, i = 1,\cdots,p$ and $t = 1,\cdots,T$.

(ii) MIQP for the three-regime model with intersected boundaries.

Let $\boldsymbol{g} = \{g_{j,t} : j = 1,2, t = 1,\cdots,T\}$, $\boldsymbol{\mathcal{I}} = \{I_{k,t} : k = 1,2,3, t = 1,\cdots,T\}$ and $\boldsymbol{\ell} = \{\ell_{k,i,t} : k = 1,\cdots,3, i = 1,\cdots,p, t = 1,\cdots,T\}$. Solve the following problem:

$$\min_{\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{g},\boldsymbol{\mathcal{I}},\boldsymbol{\ell}} \frac{1}{T}\sum_{t=1}^{T}\left(Y_t - \sum_{k=1}^{3}\sum_{i=1}^{p} X_{i,t}\ell_{k,i,t}\right)^2, \tag{C.4}$$

$$\text{subject to} \begin{cases} \boldsymbol{\beta}_k \in \mathcal{B}, \quad \boldsymbol{\gamma}_j \in \Gamma; \\ g_{j,t} \in \{0,1\}, \quad I_{k,t} \in \{0,1\}; \\ L_i \le \beta_{k,i} \le U_i \\ (g_{j,t}-1)(M_{j,t}+\epsilon) < \boldsymbol{Z}_{j,t}^{\mathrm{T}}\boldsymbol{\gamma}_j \le g_{j,t}M_{j,t}; \\ g_{1,t}L_i \le \ell_{1,i,t} \le g_{1,t}U_i, \quad I_{k,t}L_i \le \ell_{k,i,t} \le I_{k,t}U_i \\ L_i(1-g_{1,t}) \le \beta_{k,i} - \ell_{1,i,t} \le U_i(1-g_{1,t}); \\ L_i(1-I_{k,t}) \le \beta_{k,i} - \ell_{k,i,t} \le U_i(1-I_{k,t}); \\ I_{2,t} \le 1-g_{1,t}, \quad I_{2,t} \le 1-g_{2,t}, \quad I_{2,t} \ge 1-g_{1,t}-g_{2,t}; \\ I_{3,t} \le g_{1,t}, \quad I_{3,t} \le 1-g_{2,t}, \quad I_{3,t} \ge g_{1,t}-g_{2,t}, \end{cases}$$

for $k = 1,2,3, j = 1,2, i = 1,\cdots,p$ and $t = 1,\cdots,T$.

APPENDIX D: PROOFS FOR SECTION 5

In this section, we analyze the validity of the proposed smoothed regression bootstrap for the inference of the boundary coefficient $\boldsymbol{\gamma}_0$ and the regression coefficient $\boldsymbol{\beta}_0$. Our proofs include two parts. In Section D.1, we presents some conditions for a general bootstrap population, under which the consistency of the bootstrap is shown. In Section D.2, we verify the proposed smoothed regression bootstrap satisfies these conditions, and hence establish its consistency.

D.1. Sufficient conditions for a consistent bootstrap for the segmented regressions. Given a sample $\mathcal{D}_T$ from the model of segmented regression (2.1) of the main paper, suppose the LSE for $\boldsymbol{\beta}_0$ obtained with $\mathcal{D}_T$ is $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1^{\mathrm{T}},\cdots,\widehat{\boldsymbol{\beta}}_4^{\mathrm{T}})^{\mathrm{T}}$, and the centriod of the LSEs for $\boldsymbol{\gamma}_0$ is $\widehat{\boldsymbol{\gamma}}^c$. To simplify notations, in this section we use $\widehat{\boldsymbol{\gamma}}$ for $\widehat{\boldsymbol{\gamma}}^c$. Let $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\gamma}},\widehat{\boldsymbol{\beta}})$. The model to generate the bootstrap resamples is

$$Y = \sum_{k=1}^{4} \boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k \mathbb{1}\{\boldsymbol{Z} \in R_k(\widehat{\boldsymbol{\gamma}}^c)\} + \varepsilon, \tag{D.1}$$

where $(\boldsymbol{X},\boldsymbol{Z},\varepsilon) \sim \widehat{\mathbb{Q}}_h$, which generate the bootstrap population that mirrors the population distribution $\mathbb{P}_0$ that generates the original sample $\mathcal{D}_T$. Let $\{Y_i^*,\boldsymbol{X}_i^*,\boldsymbol{Z}_i^*\}_{i=1}^{m_T}$ be a bootstrap resample from (D.1), we denote by $\widehat{\mathbb{Q}}_h^*$ as its empirical measure. The LSEs obtained with the bootstrap resample are $\widehat{\boldsymbol{\theta}}^* = (\widehat{\boldsymbol{\gamma}}^*,\widehat{\boldsymbol{\beta}}^*)$ such that

$$\begin{aligned} \widehat{\mathbb{Q}}_h^*\{m(\boldsymbol{W}^*,\widehat{\boldsymbol{\theta}}^*)\} &= \min_{\boldsymbol{\theta}\in\Theta} \widehat{\mathbb{Q}}_h^*\{m(\boldsymbol{W}^*,\boldsymbol{\theta})\} \\ &= \min_{\boldsymbol{\theta}\in\Theta} \frac{1}{m_T}\sum_{i=1}^{m_T}[Y_i^* - \{\sum_{k=1}^{4}\boldsymbol{X}_i^*\boldsymbol{\beta}_k\mathbb{1}(\boldsymbol{Z}_i^* \in R_k(\boldsymbol{\gamma})\}]^2. \end{aligned} \tag{D.2}$$

Let the bootstrap LSE set for $\boldsymbol{\gamma}$ be $\widehat{G}^*$, whose centriod is denoted as $\widehat{\boldsymbol{\gamma}}^{*c}$.

The sufficient conditions for a consistent boostrap for the segmented regressions are listed as follows.

(C1) [Consistency] $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0$.
(C2) [Moment conditions] $\limsup_{T\to\infty}\widehat{\mathbb{Q}}_h(\|\boldsymbol{X}\|^4) < \infty$ and $\limsup_{T\to\infty}\widehat{\mathbb{Q}}_h(\varepsilon^4) < \infty$.

(C3) $\widehat{\mathbb{Q}}_h(\varepsilon_Q|\boldsymbol{X},\boldsymbol{Z})=0$ and $\widehat{\mathbb{Q}}_h(\varepsilon_Q^2)\to\mathbb{P}_0(\varepsilon^2)$.

(C4) Suppose that $U(\boldsymbol{X},\boldsymbol{Z})$ is a function of $(\boldsymbol{X},\boldsymbol{Z})$ with $\mathbb{P}_0\{|U(\boldsymbol{X},\boldsymbol{Z})|\}<\infty$, then

$$\sup_{R\subset\mathcal{Z}}\left|\widehat{\mathbb{Q}}_h\{U(\boldsymbol{X},\boldsymbol{Z})\mathbb{1}(\boldsymbol{Z}\in R)\}-\mathbb{P}_0\{U(\boldsymbol{X},\boldsymbol{Z})\mathbb{1}(\boldsymbol{Z}\in R)\}\right|\to 0. \tag{D.3}$$

(C5) There exist some constants δ_1 and $c_1>0$ such that for each $l=1$ and 2 and any $\epsilon\in(0,\delta_1)$, it holds that $\widehat{\mathbb{Q}}_h\{\mathbb{1}(|q_l|<\epsilon)|\boldsymbol{Z}_{-1,l}\}>c_1\epsilon$ almost surely.

(C6) There exists some constant $r>8$ such that for each $l=1$ and 2, there exists a neighborhood $\mathcal{N}_l$ for $\boldsymbol{\gamma}_{l0}$ such that $\sup_{\boldsymbol{\gamma}\in\mathcal{N}_l}\widehat{\mathbb{Q}}_h(\|\boldsymbol{X}\|^r|\boldsymbol{Z}_l^{\mathrm{T}}\boldsymbol{\gamma}=0)<\infty$, $\inf_{\boldsymbol{\gamma}\in\mathcal{N}_l}\widehat{\mathbb{Q}}_h(\|\boldsymbol{X}\||\boldsymbol{Z}_l^{\mathrm{T}}\boldsymbol{\gamma}=0)>0$ and $\sup_{\boldsymbol{\gamma}\in\gamma_l}\widehat{\mathbb{Q}}_h(\varepsilon^r|\boldsymbol{Z}_l^{\mathrm{T}}\boldsymbol{\gamma}=0)<\infty$.

(C7) For each $l\in\{1,2\}$, as $T\to\infty$ the following hold.

(i) Let $\widetilde{f}_{\boldsymbol{Z}_l}$ be the density function of $\boldsymbol{Z}_l$ under $\widehat{\mathbb{Q}}_h$ and $f_{\boldsymbol{Z}_l}$ be the density function of $\boldsymbol{Z}_l$ under $\mathbb{P}_0$, then $\|\widetilde{f}_{\boldsymbol{Z}_l}-f_{\boldsymbol{Z}_l}\|_\infty\to 0$;

(ii) For each $(j,k)\in\mathcal{S}(l)$,

$$\widehat{\mathbb{Q}}_h\left\{e^{it\xi_Q^{(j,k)}}|q_{l,Q}=0,\boldsymbol{Z}_{-1,l}\right\}\to\mathbb{P}_0\left\{e^{it\xi^{(j,k)}}|q_l=0,\boldsymbol{Z}_{-1,l}\right\}\ \text{ almost surely.} \tag{D.4}$$

(iii) Under $\widehat{\mathbb{Q}}_h$, the conditional density $\widetilde{f}_{\xi_Q^{(j,k)}|(q_{l,Q},\boldsymbol{Z}_{-1,l})}(\xi|q,\boldsymbol{z})$ and $\widetilde{f}_{q_{l,Q}|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z})$ are continuous at $q=0$ and bounded by some $0<F<\infty$ for any $\xi\in\mathbb{R}$ and $\boldsymbol{z}\in\mathcal{Z}_{-1,l}$;

The following Lemmas D.1–D.5 will establish that under Conditions *(C1)–(C7)*, the asymptotic distributions of the bootstrap estimators are the same as that of the estimators obtained with the sample $\mathcal{D}_T$. The proofs essentially mimics that in Section B, while require careful verification for the validity of replacing $(\mathbb{P}_0,\boldsymbol{\theta}_0)$ with its bootstrap counterpart $(\widehat{\mathbb{Q}}_h,\widehat{\boldsymbol{\theta}})$.

LEMMA D.1. *Assume that Assumptions 1-5 and Conditions (C1)–(C4) hold, then* $\widehat{\boldsymbol{\theta}}^*\xrightarrow{P}\boldsymbol{\theta}_0$.

PROOF. First, we show $\sup_{\boldsymbol{\theta}\in\Theta}\left|\widehat{\mathbb{Q}}_h\{m(\boldsymbol{W},\boldsymbol{\theta})\}-\mathbb{P}_0\{m(\boldsymbol{W},\boldsymbol{\theta})\}\right|\to 0$. For any $\boldsymbol{\theta}$, under $\widehat{\mathbb{Q}}_h$ where $Y=\sum_{k=1}^4\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\mathbb{1}\{\boldsymbol{Z}\in R(\widehat{\boldsymbol{\gamma}})\}+\varepsilon$,

$$\begin{aligned}
&\widehat{\mathbb{Q}}_h\{m(\boldsymbol{W},\boldsymbol{\theta})\}=\widehat{\mathbb{Q}}_h(\varepsilon_Q^2)\\
&+\sum_{k=1}^4\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k)\}^2\mathbb{1}^{(k)}(\widehat{\boldsymbol{\gamma}})\mathbb{1}^{(k)}(\boldsymbol{\gamma})]+\sum_{k\neq j}\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_j-\widehat{\boldsymbol{\beta}}_k)\}^2\mathbb{1}^{(k)}(\widehat{\boldsymbol{\gamma}})\mathbb{1}^{(j)}(\boldsymbol{\gamma})]\\
&+2\sum_{k=1}^4\widehat{\mathbb{Q}}_h[\varepsilon\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k)\mathbb{1}^{(k)}(\widehat{\boldsymbol{\gamma}})\mathbb{1}^{(k)}(\boldsymbol{\gamma})]+2\sum_{k\neq j}\widehat{\mathbb{Q}}_h[\varepsilon\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_j-\widehat{\boldsymbol{\beta}}_k)\mathbb{1}^{(k)}(\widehat{\boldsymbol{\gamma}})\mathbb{1}^{(j)}(\boldsymbol{\gamma})]\\
&=:A_Q+B_{1,Q}(\boldsymbol{\theta})+B_{2,Q}(\boldsymbol{\theta})+C_{1,Q}(\boldsymbol{\theta})+C_{2,Q}(\boldsymbol{\theta}),\ \text{ say.}
\end{aligned}$$

Similarly, under P_0 where $Y=\sum_{k=1}^4\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_0\mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)+\varepsilon$,

$$\begin{aligned}
&\mathbb{P}_0\{m(\boldsymbol{W},\boldsymbol{\theta})\}=\mathbb{P}_0(\varepsilon^2)\\
&+\sum_{k=1}^4\mathbb{P}_0[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)\mathbb{1}^{(k)}(\boldsymbol{\gamma})]+\sum_{k\neq j}\mathbb{P}_0[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_j-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\boldsymbol{\gamma}_0)\mathbb{1}^{(j)}(\boldsymbol{\gamma})]
\end{aligned}$$

$$+2\sum_{k=1}^{4}\mathbb{P}_0[\varepsilon\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)]+2\sum_{k\neq j}\mathbb{P}_0[\varepsilon\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_j-\boldsymbol{\beta}_{0,k})\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(j)}(\gamma)]$$

$$=:A_P+B_{1,P}(\boldsymbol{\theta})+B_{2,P}(\boldsymbol{\theta})+C_{1,P}(\boldsymbol{\theta})+C_{2,P}(\boldsymbol{\theta}),\ \text{say.}$$

Therefore, it suffices to show that $A_Q\to A_P, \sup_{\boldsymbol{\theta}\in\Theta}|B_{i,Q}(\boldsymbol{\theta})-B_{i,P}(\boldsymbol{\theta})|\to 0$ and $\sup_{\boldsymbol{\theta}\in\Theta}|C_{i,Q}(\boldsymbol{\theta})-C_{i,P}(\boldsymbol{\theta})|\to 0$ for $i=1,2$. The first part $A_Q\to A_P$ is because of Condition *(C3)*. Denote $B_{1,Q}(\boldsymbol{\theta})=\sum_{k=1}^{4}B_{1,k,Q}$ and $B_{1,P}(\boldsymbol{\theta})=\sum_{k=1}^{4}B_{1,k,P}$. Then for each k, using the triangle inequality, we obtain

$$|B_{1,k,Q}(\boldsymbol{\theta})-B_{1,k,P}(\boldsymbol{\theta})|\le D_1(\boldsymbol{\theta})+D_2(\boldsymbol{\theta})+D_3(\boldsymbol{\theta}), \tag{D.5}$$

where

$$D_1(\boldsymbol{\theta})=\left|\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k)\}^2\mathbb{1}^{(k)}(\widehat{\gamma})\mathbb{1}^{(k)}(\gamma)]-\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\widehat{\gamma})\mathbb{1}^{(k)}(\gamma)]\right|,$$

$$D_2(\boldsymbol{\theta})=\left|\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\widehat{\gamma})\mathbb{1}^{(k)}(\gamma)]-\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)]\right|,$$

$$D_3(\boldsymbol{\theta})=\left|\widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)]-\mathbb{P}_0[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)]\right|.$$

For $D_1(\boldsymbol{\theta})$, it can be bounded by

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta}D_1(\boldsymbol{\theta})&\le\sup_{\boldsymbol{\theta}\in\Theta}\widehat{\mathbb{Q}}_h\left|\left\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k)\right\}^2-\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\right|\\&=\sup_{\boldsymbol{\theta}\in\Theta}\widehat{\mathbb{Q}}_h\left|\left\{\boldsymbol{X}^{\mathrm{T}}\left(2\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_{0,k}\right)\right\}\left\{\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{\beta}_{0,k}-\widehat{\boldsymbol{\beta}}_k\right)\right\}\right|\\&\le\sqrt{\widehat{\mathbb{Q}}_h\left\{\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{\beta}_{0,k}-\widehat{\boldsymbol{\beta}}_k\right)\right\}^2}\sup_{\boldsymbol{\theta}\in\Theta}\sqrt{\widehat{\mathbb{Q}}_h\left\{\boldsymbol{X}^{\mathrm{T}}\left(2\boldsymbol{\beta}_k-\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}_{0,k}\right)\right\}^2},\end{aligned}\tag{D.6}$$

where (D.6) converges to 0 is because its first term converges to 0 by $\boldsymbol{\beta}_{0,k}\to\widehat{\boldsymbol{\beta}}_k$ and Cauchy-Schwartz inequality, and its second term is uniformly bounded since $\limsup_T\widehat{\mathbb{Q}}_h\left\{\|\boldsymbol{X}\|^4\right\}<\infty$ and Θ is compact. For $D_2(\boldsymbol{\theta})$,

$$\begin{aligned}D_2(\boldsymbol{\theta})\le&\widehat{\mathbb{Q}}_h\left|\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^2\mathbb{1}^{(k)}(\gamma)\left\{\mathbb{1}^{(k)}(\widehat{\gamma})-\mathbb{1}^{(k)}(\gamma_0)\right\}\right|\\&\sqrt{\widehat{\mathbb{Q}}_h\left[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k})\}^4\right]}\sqrt{\widehat{\mathbb{Q}}_h\left\{\left|\mathbb{1}^{(k)}(\widehat{\gamma})-\mathbb{1}^{(k)}(\gamma_0)\right|\right\}},\end{aligned}\tag{D.7}$$

where the first term on the right-hand side is uniformly bounded and the second term converges to zero by the dominated convergence theorem and $\widehat{\gamma}\to\gamma_0$ in *(C1)*, we have $\sup_{\boldsymbol{\theta}\in\Theta}D_2(\boldsymbol{\theta})\to 0$. For $D_3(\boldsymbol{\theta})$, let δ_i be the i-th element of $\boldsymbol{\beta}_k-\boldsymbol{\beta}_{0,k}$, then

$$D_3(\boldsymbol{\theta})\le\sum_{i,j\in[p]}\delta_i\delta_j\left|\widehat{\mathbb{Q}}_h\left\{X_iX_j\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)\right\}-\mathbb{P}_0\left\{X_iX_j\mathbb{1}^{(k)}(\gamma_0)\mathbb{1}^{(k)}(\gamma)\right\}\right|.$$

By the compactness of Θ, $\delta_i\delta_j$ is uniformly bounded. Then from (D.3) in *(C4)*, we obtain $\sup_{\boldsymbol{\theta}\in\Theta}D_3(\boldsymbol{\theta})\to 0$. By (D.5) and triangle inequality, $\sup_{\boldsymbol{\theta}\in\Theta}|B_{1,k,Q}(\boldsymbol{\theta})-B_{1,k,P}(\boldsymbol{\theta})|\to 0$. Summing accross k results in $\sup_{\boldsymbol{\theta}\in\Theta}|B_{1,Q}(\boldsymbol{\theta})-B_{1,P}(\boldsymbol{\theta})|\to 0$.

With the same argument as above except for replacing $\boldsymbol{\beta}_k$ by $\boldsymbol{\beta}_j$ and $\mathbb{1}^{(k)}(\gamma)$ by $\mathbb{1}^{(j)}(\gamma)$, we can show $\sup_{\boldsymbol{\theta}\in\Theta}|B_{2,Q}(\boldsymbol{\theta})-B_{2,P}(\boldsymbol{\theta})|\to 0$. Similarly, using the above decomposition argument and with Conditions *(C2)*, *(C4)* and *(C1)*, it can be readily shown that $\sup_{\boldsymbol{\theta}\in\Theta}|C_{i,Q}(\boldsymbol{\theta})-C_{i,P}(\boldsymbol{\theta})|\to 0$ for $i=1,2$. Combining the above pieces gives

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\widehat{\mathbb{Q}}_h\{m(\boldsymbol{W},\boldsymbol{\theta})\}-\mathbb{P}_0\{m(\boldsymbol{W},\boldsymbol{\theta})\}\right|\to 0. \tag{D.8}$$

Because (i) $\widehat{\mathbb{Q}}_h^*$ is the empirical measure of $\widehat{\mathbb{Q}}_h$, (ii) $\widehat{\mathbb{Q}}_h \{\sup_{\boldsymbol{\theta}\in\Theta} m(\boldsymbol{W},\boldsymbol{\theta})\} < \infty$ by the condition *(C2)* and the compactness of Θ, and (iii) $\mathcal{F} = \{m(\mathbf{w},\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$ has a finite VC-dimension, the Glivenko-Cantelli theorem implies that

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \widehat{\mathbb{Q}}_h^* \{m(\boldsymbol{W}^*,\boldsymbol{\theta})\} - \widehat{\mathbb{Q}}_h \{m(\boldsymbol{W},\boldsymbol{\theta})\} \right| \xrightarrow{P} 0. \tag{D.9}$$

Consequently, from (D.8) and (D.9) we have

$$\sup_{\boldsymbol{\theta}\in\Theta} \left| \widehat{\mathbb{Q}}_h^* \{m(\boldsymbol{W}^*,\boldsymbol{\theta})\} - \mathbb{P}_0 \{m(\boldsymbol{W},\boldsymbol{\theta})\} \right| \xrightarrow{P} 0. \tag{D.10}$$

Because of (D.10) together with the facts that $\boldsymbol{\theta} \mapsto \mathbb{P}_0 \{m(\boldsymbol{W},\boldsymbol{\theta})\}$ is continuous and $\boldsymbol{\theta}_0$ is the unique minimizer of $\mathbb{P}_0 \{m(\boldsymbol{W},\boldsymbol{\theta})\}$ as established in Appendix B, it follows that $\widehat{\boldsymbol{\theta}}^* \xrightarrow{P} \boldsymbol{\theta}_0$ using the similar arguments as in Section B.2. $\square$

LEMMA D.2. *Assume that Assumptions 1-5 and Conditions (C1)–(C6) hold. Then* $\sqrt{m_T}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = O_p(1)$ *and* $m_T(\widehat{\gamma}^* - \gamma) = O_p(1)$.

Proof. From Lemma D.1 we know that $\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}} = o_p(1)$ and $\widehat{\gamma}^* - \gamma = o_p(1)$. The proof of the convergence rate of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\gamma}^*$ is analogous to the proof of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\gamma}$ in Appendix B.

First, because of the conditional zero mean condition of ε in *(C3)*, we can decompose $\widehat{\mathbb{Q}}_h \{m(\boldsymbol{W},\boldsymbol{\theta}) - m(\boldsymbol{W},\boldsymbol{\theta}_Q)\}$ as

$$\begin{aligned}
\widehat{\mathbb{Q}}_h \{m(\boldsymbol{W},\boldsymbol{\theta}) - m(\boldsymbol{W},\boldsymbol{\theta}_Q)\} &= \sum_{j=1}^{4} \widehat{\mathbb{Q}}_h[\{\boldsymbol{X}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)\}^2 \mathbb{1}^{(j)}(\widehat{\gamma})\mathbb{1}^{(j)}(\gamma)] \\
&\quad + \sum_{i=1}^{4}\sum_{k\neq i}^{4} \widehat{\mathbb{Q}}_h[(\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_{Q,i} - \boldsymbol{\beta}_k))^2 \mathbb{1}^{(i)}(\widehat{\gamma})\mathbb{1}^{(k)}(\gamma)] \\
&=: \sum_{j=1}^{4} J_j^Q(\boldsymbol{\theta}) + \sum_{i=1}^{4}\sum_{k\neq i}^{4} G_{ik}^Q(\boldsymbol{\theta}), \quad \text{say.}
\end{aligned} \tag{D.11}$$

Because $\widehat{\gamma} \to \gamma_0$ and (D.3), it can be shown that

$$\sup_{i,j\in[p]} \sup_{\gamma\in\Gamma} \left| \widehat{\mathbb{Q}}_h \left\{ X_i X_j \mathbb{1}^{(j)}(\widehat{\gamma})\mathbb{1}^{(j)}(\gamma) \right\} - \mathbb{P}_0 \left\{ X_i X_j \mathbb{1}^{(j)}(\gamma_0)\mathbb{1}^{(j)}(\gamma) \right\} \right| \to 0, \tag{D.12}$$

following similar arguments as for $D_2(\boldsymbol{\theta})$ and $D_3(\boldsymbol{\theta})$ in the previous lemma. Since the smallest eigenvalue of $\mathbb{P}_0 \left\{ \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mathbb{1}^{(j)}(\gamma_0)\mathbb{1}^{(j)}(\gamma) \right\}$ is uniformly bounded away from 0, (D.12) implies that the smallest eigenvalue of $\widehat{\mathbb{Q}}_h \left\{ \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mathbb{1}^{(j)}(\widehat{\gamma})\mathbb{1}^{(j)}(\gamma) \right\}$ is uniformly bounded away from 0 if γ is in some neighborhood of γ_0, for $T \geq T_0$ with some $T_0 > 0$, because the entry-wise convergence of matrices can imply the convergence of eigenvalues, which can be easily seen from the perspective of characteristic polynomials. This implies that

$$J_j^Q(\boldsymbol{\theta}) \geq \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|^2,$$

for $j \in \{1, \cdots, 4\}$ and $T \geq T_0$.

With Conditions *(C4)* and *(C5)*, which imply that Assumptions 3.(ii), 4.(i) and 4.(iii) hold when replacing $\mathbb{P}$ with $\widehat{\mathbb{Q}}_h$, the moment inequalies in Lemma A.2 hold under the bootstrap population $\widehat{\mathbb{Q}}_h$. Then, with the same argument as in Step 1 of the proof of Theorem 3.1, it can be shown that for any γ in some neighborhood of $\widehat{\gamma}$,

$$\frac{J_{k_l}^Q(\boldsymbol{\theta}) + J_{i_l}^Q(\boldsymbol{\theta})}{2} + G_{i_l k_l}^Q(\boldsymbol{\theta}) + G_{k_l i_l}^Q(\boldsymbol{\theta}) \gtrsim \left(\|\boldsymbol{\beta}_{Q,i_l} - \boldsymbol{\beta}_{i_l}\|^2 + \|\boldsymbol{\beta}_{Q,k_l} - \boldsymbol{\beta}_{k_l}\|^2 + \|\widehat{\gamma}_l - \gamma_l\| \right),$$

where k_l and i_l are defined the same as in Appendix B, which further implies

$$\widehat{\mathbb{Q}}_h \{m(\boldsymbol{W},\boldsymbol{\theta}) - m(\boldsymbol{W},\boldsymbol{\theta}_Q)\} \gtrsim \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 + \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|, \tag{D.13}$$

for $\boldsymbol{\gamma}$ in some neighborhood of $\widehat{\boldsymbol{\gamma}}$.

Let $\mathbb{G}_T^* = \sqrt{m_T}(\widehat{\mathbb{Q}}_h^* - \widehat{\mathbb{Q}}_h)$. By inspecting the proofs of Lemmas A.4–A.6, we notice that these lemmas can be established once we have the moment inequalities outlined in Lemma A.3, whose conditions hold if we replace the population $\mathbb{P}_0$ by $\widehat{\mathbb{Q}}_h^*$. Therefore, we can replace $\mathbb{G}_T$ in Lemma A.6 by $\mathbb{G}_T^*$, under Conditions *(C4)–(C6)*. Then, with the same arguments as in Step 2 in Section B.4, we obtain

$$\|\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\|^2 + \|\widehat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}\| \lesssim \|\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\|^2 o_p(1) + 4\lambda\|\widehat{\boldsymbol{\gamma}}^* - \widehat{\boldsymbol{\gamma}}\| + O_p(m_T^{-1}),$$

for any $\lambda \in (0,1)$, which implies that $\|\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\|^2 = O_p(m_T^{-1})$, and thus $\|\widehat{\boldsymbol{\gamma}}^* - \widehat{\boldsymbol{\gamma}}\| = O_p(m_T^{-1})$. □

We now proceed to derive a result similar to Lemma B.1.

LEMMA D.3. *Assume that Assumptions 1-5 and Conditions (C1)–(C6) hold. Then uniformly for $\boldsymbol{h} = (\boldsymbol{u}^\top, \boldsymbol{v}^\top)^\top$ in any compact set in $\mathbb{R}^{4p+d_1+d_2}$,*

$$\begin{aligned} & m_T\widehat{\mathbb{Q}}_h^* \left\{ m(\boldsymbol{W}^*, \widehat{\boldsymbol{\beta}} + \frac{\boldsymbol{u}}{\sqrt{m_T}}, \widehat{\boldsymbol{\gamma}} + \frac{\boldsymbol{v}}{m_T}) - m(\boldsymbol{W}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}) \right\} \\ =& D_T^*(\boldsymbol{v}) - 2W_T^*(\boldsymbol{u}) + o_P(1), \end{aligned} \tag{D.14}$$

where

$$W_T^*(\boldsymbol{u}) = \sum_{j=1}^{4} \left[\sqrt{m_T}\widehat{\mathbb{Q}}_h^* \left\{ \boldsymbol{u}_j^\top \boldsymbol{X} \varepsilon_Q \mathbb{1}^{(j)}(\widehat{\boldsymbol{\gamma}}) \right\} + \boldsymbol{u}_j^\top \widehat{\mathbb{Q}}_h \left\{ \boldsymbol{X}\boldsymbol{X}^\top \mathbb{1}^{(j)}(\widehat{\boldsymbol{\gamma}}) \right\} \boldsymbol{u}_j \right],$$

and

$$D_T^*(\boldsymbol{v}) = \sum_{l=1}^{2} \sum_{(j,k)\in\mathcal{S}(l)} m_T\widehat{\mathbb{Q}}_h^* \left[\xi_Q^{(j,k)} \mathbb{1}\left\{ s_l^{(j)} \left(m_T q_{l,Q} + \boldsymbol{Z}_{-1,l}^\top \boldsymbol{v}_{-1,l} \right) \le 0 < s_l^{(j)} m_T q_{l,Q} \right\} \right],$$

with $\xi_Q^{(j,k)} = \left(\widehat{\boldsymbol{\delta}}_{jk}^\top \boldsymbol{X}\boldsymbol{X}^\top \widehat{\boldsymbol{\delta}}_{jk} + 2\boldsymbol{X}^\top \widehat{\boldsymbol{\delta}}_{jk}\varepsilon_Q \right) \left\{ \mathbb{1}^{(j)}(\widehat{\boldsymbol{\gamma}}) + \mathbb{1}^{(k)}(\widehat{\boldsymbol{\gamma}}) \right\},$

where $\widehat{\boldsymbol{\delta}}_{jk} = \widehat{\boldsymbol{\beta}}_j - \widehat{\boldsymbol{\beta}}_k$, $q_{l,Q} = \boldsymbol{Z}_l^\top \widehat{\gamma}_l$, $\mathcal{S}(l)$ is the set of indices of adjacent regions split by the l-th hyperplane as defined in (3), and $s_l^{(j)} = \operatorname{sign}(\boldsymbol{z}^\top \gamma_{l0})$ for $\boldsymbol{z} \in D^{(j)}(\gamma_0)$ as defined in (2).

Proof. The left-hand side of (D.14) can be decomposed in the same way as (B.32) in the proof of Lemma B.1. It is noted that Lemma B.1 is established by showing the decomposed terms in (B.32) besides $D_T(\boldsymbol{v})$ and $W_T(\boldsymbol{u})$ all converge to 0 in probability with the application of Lemma A.5. With Conditions *(C4)–(C6)*, Lemma A.4 holds with $\mathbb{G}_T$ replaced by $\mathbb{G}_T^*$. It can be derived with similar lines of the proof of Lemma A.5 that

$$\begin{aligned} &\sup_{\|\boldsymbol{\gamma}_l - \boldsymbol{\gamma}_{Q,l}\| \le m_T^{-1}} \sqrt{m_T}\widehat{\mathbb{Q}}_h^* \left\{ U \left| \mathbb{1}_j(\boldsymbol{\gamma}_j) - \mathbb{1}_j(\boldsymbol{\gamma}_{Q,j}) \right| \left| \mathbb{1}_l(\boldsymbol{\gamma}_l) - \mathbb{1}_l(\widehat{\boldsymbol{\gamma}}_l) \right| \right\} = o_p(1), \\ &\sup_{\substack{\|\boldsymbol{\gamma}_j - \boldsymbol{\gamma}_{Q,j}\| \le m_T^{-1} \\ \|\boldsymbol{\gamma}_l - \boldsymbol{\gamma}_{Q,l}\| \le m_T^{-1}}} m_T\widehat{\mathbb{Q}}_h^* \left\{ U \left| \mathbb{1}_j(\boldsymbol{\gamma}_j) - \mathbb{1}_j(\boldsymbol{\gamma}_{Q,j}) \right| \left| \mathbb{1}_l(\boldsymbol{\gamma}_l) - \mathbb{1}_l(\widehat{\boldsymbol{\gamma}}_l) \right| \right\} = o_p(1), \end{aligned} \tag{D.15}$$

for $U = \|\boldsymbol{X}\|^2$ and $U = |\varepsilon_Q|\|\boldsymbol{X}\|$. Then, the above lemma can be proved by following the same arguments as in Section B.6. □

LEMMA D.4. *Assume that Conditions (C1)–(C7) hold. Then the finite-dimensional weak limit of $D_T^*(\boldsymbol{v})$ is the same as $D(\boldsymbol{v})$ as presented in Lemma B.2.*

Proof. The derivation of the finite-dimensional weak limit of $D_T^*(\boldsymbol{v})$ is in parallel to that of $D_T(\boldsymbol{v})$ in the proof of Lemma B.2.

First, as (B.48) in Part 1, $D_T^*(\boldsymbol{v})$ can be expressed as a sum of functionals of some empirical point processes. For each $l \in \{1,2\}$ and $(j,k) \in \mathcal{S}(l)$, we define an empirical point process $\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}(\cdot) \in M_p(E_l)$, where $E_l = \mathbb{R}_{s_l^{(j)}} \times \mathcal{Z}_{-1,l} \times \mathbb{R}$ as:

$$\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}(F) := m_T \widehat{\mathbb{Q}}_h^* \left[\mathbb{1}\left\{ (m_T q_{l,Q}, \boldsymbol{Z}_{-1,l}, \xi_Q^{(j,k)}) \in F \right\} \right], \tag{D.16}$$

for each $F = (F_1, F_2, F_3) \in E_l$. Then $D_T^*(\boldsymbol{v})$ can be expressed as

$$D_T^*(\boldsymbol{v}) = \sum_{l=1}^{2} \sum_{(j,k)\in\mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)} \left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)} \right). \tag{D.17}$$

where the functional $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}$ is defined in (B.47).

Second, we derive the weak limit of $\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}$ as in Part 2 of the proof of Lemma B.2. The two ingredients are the calculation of the limit of $\widehat{\mathbb{Q}}_h \left\{ \widehat{\mathbf{N}}_{Q,l,T}^{(j,k)} \right\}$, which is required in Kallenberg's theorem, and the application of Meyer's theorem. First, for any $F = (F_1, F_2, F_3) \in \mathcal{E}_l$, the basis of relatively compact open set in E_l, we claim that:

$$\lim_{T, m_T \to \infty} \widehat{\mathbb{Q}}_h \left\{ \widehat{\mathbf{N}}_{Q,l,T}^{(j,k)} \right\} = \mu_l^{(j,k)}(F), \tag{D.18}$$

where the mean measure $\mu_l^{(j,k)}$ is defined in (B.50). This can be shown as below. Note that

$$\begin{aligned}
\widehat{\mathbb{Q}}_h \left\{ \widehat{\mathbf{N}}_{Q,l,T}^{(j,k)} \right\} &= m_T \widehat{\mathbb{Q}}_h \left[\mathbb{1}\left\{ (m_T q_{l,Q}, \boldsymbol{Z}_{-1,l}, \xi_Q^{(j,k)}) \in F \right\} \right] \\
&= m_T \int_{m_T q \in F_1, \boldsymbol{z} \in F_2, \xi \in F_3} \widetilde{f}_Q^{(i,j)}(q, \boldsymbol{z}, \xi) dq d\boldsymbol{z} d\xi \\
&= \int_{\tilde{q} \in F_1, \boldsymbol{z} \in F_2, \xi \in F_3} \widetilde{f}_Q^{(i,j)} \left(\frac{\tilde{q}}{m_T}, \boldsymbol{z}, \xi \right) d\tilde{q} d\boldsymbol{z} d\xi,
\end{aligned}$$

where $\widetilde{f}_Q^{(i,j)}(q, \boldsymbol{z}, \xi)$ is the joint density function of $(q_{l,Q}, \boldsymbol{Z}_{-1,l}, \xi_Q^{(i,j)})$ under $\widehat{\mathbb{Q}}_h$. The claim (D.18) can be verified as follows.

$$\begin{aligned}
&\int_{\tilde{q} \in F_1, \boldsymbol{z} \in F_2, \xi \in F_3} \widetilde{f}_Q^{(i,j)} \left(\frac{\tilde{q}}{m_T}, \boldsymbol{z}, \xi \right) d\tilde{q} d\boldsymbol{z} d\xi \\
&= \int_{\tilde{q} \in F_1, \boldsymbol{z} \in F_2, \xi \in F_3} \widetilde{f}_{\xi_Q^{(j,k)} | (q_{l,Q}, \boldsymbol{Z}_{-1,l})} \left(\xi | \frac{\tilde{q}}{m_T}, \boldsymbol{z} \right) \widetilde{f}_{q_{l,Q} | \boldsymbol{Z}_{-1,l}} \left(\frac{\tilde{q}}{m_T} | \boldsymbol{z} \right) \widetilde{f}_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) d\tilde{q} d\boldsymbol{z} d\xi \\
&\overset{(i)}{\to} \int_{\tilde{q} \in F_1, \boldsymbol{z} \in F_2, \xi \in F_3} \widetilde{f}_{\xi_Q^{(j,k)} | (q_{l,Q}, \boldsymbol{Z}_{-1,l})} (\xi | 0, \boldsymbol{z}) \widetilde{f}_{q_{l,Q} | \boldsymbol{Z}_{-1,l}} (0 | \boldsymbol{z}) \widetilde{f}_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) d\tilde{q} d\boldsymbol{z} d\xi \quad (\text{as } m_T \to \infty) \\
&= \int_{\tilde{q} \in F_1, \boldsymbol{z} \in F_2} \widehat{\mathbb{Q}}_h \left\{ \xi_Q^{(j,k)} \in F_3 | q_{l,Q} = 0, \boldsymbol{Z}_{-1,l} = \boldsymbol{z} \right\} \widetilde{f}_{q_{l,Q} | \boldsymbol{Z}_{-1,l}} (0 | \boldsymbol{z}) \widetilde{f}_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) d\tilde{q} d\boldsymbol{z}
\end{aligned}$$

$$\overset{(ii)}{\to} \int_{\tilde{q}\in F_1,\boldsymbol{z}\in F_2} \mathbb{P}_0 \left\{\xi^{(j,k)} \in F_3 | q_l = 0, \boldsymbol{Z}_{-1,l} = \boldsymbol{z}\right\} f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) d\tilde{q} d\boldsymbol{z} \quad (\text{as } T\to\infty)$$

$$= \int_{\tilde{q}\in F_1,\boldsymbol{z}\in F_2,\xi\in F_3} f_{\xi^{(i,j)}|(q_l,\boldsymbol{Z}_{-1,l})}(\xi \mid 0,\boldsymbol{z}) f_{q_l|\boldsymbol{Z}_{-1,l}}(0\mid\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) \mathrm{d}\tilde{q}\mathrm{d}\boldsymbol{z}\mathrm{d}\xi$$

$$= \mu_l^{(j,k)}(F),$$

where (i) is implied by the dominated convergence theorem because of the continuity and boundness of $\widetilde{f}_{\xi_Q^{(j,k)}|(q_{l,Q},\boldsymbol{Z}_{-1,l})}(\xi|q,\boldsymbol{z})$ and $\widetilde{f}_{q_{l,Q}|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z})$ at $q=0$ as assumed in *(C7)*. For (ii), since the characteristic function of $\xi_Q^{(i,j)}|q_{l,Q},\boldsymbol{Z}_{-1,l}$ under $\widehat{\mathbb{Q}}_h$ converges to that of $\xi^{(i,j)}|q_l,\boldsymbol{Z}_{-1,l}$ under $\mathbb{P}_0$, then $\widehat{\mathbb{Q}}_h\left\{\xi_Q^{(j,k)}\in F_3|q_{l,Q}=0,\boldsymbol{Z}_{-1,l}=\boldsymbol{z}\right\}\to \mathbb{P}_0\left\{\xi^{(j,k)}\in F_3|q_l=0,\boldsymbol{Z}_{-1,l}=\boldsymbol{z}\right\}$. In addition, it is easy to see that

$$\sup_{\boldsymbol{z}\in\mathcal{Z}_{-1,l}} \left|\widetilde{f}_{q_{l,Q}|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z})\widetilde{f}_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z}) - f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{z}) f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})\right| \to 0$$

as $T\to\infty$, due to $\|\widetilde{f}_{\boldsymbol{Z}_l} - f_{\boldsymbol{Z}_l}\|_\infty \to 0$ assumed in *(C7)*. Then (ii) follows from the dominated convergence theorem.

Since observations under $\widehat{\mathbb{Q}}_h^*$ are i.i.d., for any F with $\mu_l^{(j,k)}(F)>0$, Meyer's theorem implies that

$$\lim_{m_T\to\infty} \widehat{\mathbb{Q}}_h\left\{\mathbb{1}\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}=0\right)\right\} = e^{-\mu_l^{(j,k)}(F)}. \quad \text{(D.19)}$$

For F with $\mu_l^{(j,k)}(F)=0$, (D.19) also holds, since in such the case (D.18) implies $\widehat{\mathbb{Q}}_h\left\{\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}(F)\right\}\to 0$ as $T\to\infty$, which further implies that $\widehat{\mathbb{Q}}_h\left\{\mathbb{1}\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}=0\right)\right\}=1=e^{-\mu_l^{(j,k)}(F)}$. Since $\mu_l^{(j,k)}$ is the mean measure of $\mathbf{N}_l^{(j,k)}$ introduced in Part 2 of the proof of Lemma B.2, with the statements (D.18) and (D.19), Kallenberg's theorem (Lemma A.7) implies that for each $l\in\{1,2\}$ and $(j,k)\in\mathcal{S}(l)$, we have $\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}\Rightarrow\mathbf{N}_l^{(j,k)}$ in $M_p(E_l)$ as $m_T,T\to\infty$. Therefore, $\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}$ has the same weak limit as $\widehat{\mathbf{N}}_{l,T}^{(j,k)}$.

As derived in Part 3 of the proof of Lemma B.2, the point process $\mathbf{N}_l^{(j,k)}$ has the representation (B.55). By inspecting Part 4 of the proof of Lemma B.2 which shows the asymptotical independence of $\left(\widehat{\mathbf{N}}_{l,T}^{(j,k)}, l\in\{1,2\},(j,k)\in\mathcal{S}(l)\right)$, we find that to show the asymptotical independence of $\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}, l\in\{1,2\},(j,k)\in\mathcal{S}(l)\right)$, it suffices to show that (B.59) holds if $\mathbb{P}$ is replaced by $\widehat{\mathbb{Q}}_h$, which is indeed true since $\|\widetilde{f}_{\boldsymbol{Z}}-f_{\boldsymbol{Z}}\|_\infty\to 0$ and the uniform boundness of $f_{(q_l,q_{l'})|(\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'})}(q,q')$ at a neighborhood of $(0,0)$ implies the uniform boundness of $\widetilde{f}_{(q_{l,Q},q_{l',Q})|(\boldsymbol{Z}_{-1,l},\boldsymbol{Z}_{-1,l'})}(q,q')$ at the neighborhood, which ensures (B.59) holds when replacing $\mathbb{P}$ is replaced by $\widehat{\mathbb{Q}}_h$. The rest arguments in Part 3 of the proof of Lemma B.2 obviously hold under $\widehat{\mathbb{Q}}_h$ and $\widehat{\mathbb{Q}}_h^*$, since the observations under $\mathbb{P}_T$ are weakly dependent and the observations under $\widehat{\mathbb{Q}}_h^*$ are i.i.d. Therefore, the asymptotical independence of $\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}, l\in\{1,2\},(j,k)\in\mathcal{S}(l)\right)$ can be established.

As for adapting Part 4 of the proof of Lemma B.2, it is sufficient to verify that (I)–(III) therein hold under $\widehat{\mathbb{Q}}_h$. Let

$$R_{Q,T} = \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}\right) \text{ and } R_{Q,T,M} = \int_{E_{l,M}} g_l^{(j,k)}(x,\boldsymbol{y},z) d\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}(x,\boldsymbol{y},z).$$

For (I), the arguments, which is mainly the continuous mapping theorem, for showing $R_{T,M} \Rightarrow R_{0,M}$ also implies $R_{Q,T,M} \Rightarrow R_{0,M}$, since the probability of discontinuities is evaluated under the distribution of the limiting process $\mathbf{N}_l^{(j,k)}$. For (II), with the notations in (II) in Part 4 of the proof of Lemma B.2, we first have

$$\begin{aligned}|R_{Q,T} - R_{Q,T,M}| =& m_T \widehat{\mathbb{Q}}_h^* \left\{ |\xi| \mathbb{1}\left(|\xi| \geq M\right) \mathbb{1}\left(m_T q_{Q,l} + \boldsymbol{Z}_{-1}^{\mathrm{T}} \boldsymbol{v}_{-1,l} \leq 0 < m_T q_{Q,l}\right) \right\} \\ =:& m_T \widehat{\mathbb{Q}}_h^* (G_Q(M)), \text{ say.}\end{aligned} \tag{D.20}$$

With Condition *(C4)* we have $\widehat{\mathbb{Q}}_h \left\{ \left| \xi_Q^{(j,k)} \right|^4 | \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0 \right\} < C$ for some $C < \infty$ if $\boldsymbol{\gamma}$ is in some neighborhood of $\widehat{\gamma}_l$ and each $l \in \{1,2\}$. As in (B.74) it can be readily shown that $\widehat{\mathbb{Q}}_h$

$$\widehat{\mathbb{Q}}_h \left\{ |\xi| \mathbb{1}\left(|\xi| \geq M\right) \mid \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0 \right\} = O(M^{-1}). \tag{D.21}$$

Using (D.21) and with the similar arguments as in the proof of Lemma A.4 (i), we can show that which implies $\widehat{\mathbb{Q}}_h \left\{ |G_t(M)| \right\} = O((M m_T)^{-1})$, which implies $\widehat{\mathbb{Q}}_h \left\{ |R_{Q,T} - R_{Q,T,M}| \right\} = O(M^{-1})$ due to (D.21). Then

$$\lim_{M \to \infty} \limsup_{T \to \infty} \widehat{\mathbb{Q}}_h \left\{ |R_{Q,T} - R_{Q,T,M}| > \varepsilon \right\} \to 0,$$

for any $\varepsilon > 0$ according to the Markov inequality, which verifies (II). Since (III) in Part 4 of the proof of Lemma B.2 is for the truncation error of $R_0 = \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right)$, which is regardless of $\mathbb{P}_0$ or $\widehat{\mathbb{Q}}_h$, it also holds under the bootstrap scenario. Therefore, with (I)–(III) and by applying Theorem 4.2 of Billingsley (1968), $R_{T,Q} \Rightarrow R_0$ as $m_T, \to \infty$, i.e., $\mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}\right) \Rightarrow \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right)$. Because it has been shown that $\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}, l \in \{1,2\}, (j,k) \in \mathcal{S}(l)\right)$ are asymptotically independent, we conclude that

$$D_T^*(\boldsymbol{v}) = \sum_{l=1}^{L} \sum_{(j,k) \in \mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\widehat{\mathbf{N}}_{Q,l,T}^{(j,k)}\right) \Rightarrow \sum_{l=1}^{L} \sum_{(j,k) \in \mathcal{S}(l)} \mathcal{T}_{l,\boldsymbol{v}_l}^{(j,k)}\left(\mathbf{N}_{l,T}^{(j,k)}\right), \tag{D.22}$$

as $m_T, T \to \infty$, where the right-hand side of (D.22) is identical to that of (B.77), which is the weak limit of $D_T(\boldsymbol{v})$, the proof is completed. □

Let $\widehat{\boldsymbol{\gamma}}^{*c} = C(\widehat{G}^*)$ be the centriod of the LSEs $\widehat{\boldsymbol{\gamma}}^*$ obtained with the bootstrap resample. Let $\mathcal{L}_T^*$ be the distribution of $\{m_T(\widehat{\boldsymbol{\gamma}}^{*c} - \widehat{\boldsymbol{\gamma}}^c), \sqrt{m_T}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\}$ and $\mathcal{L}_T$ be the distribution of $\{T(\widehat{\boldsymbol{\gamma}}^c - \boldsymbol{\gamma}_0), \sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}$. The s.e-l-sc of $\{D_T^*\}$ can be obtained with the same proof as for Lemma B.3. With the same arguments as the proof for Theorem 3.3, we can establish that $\mathcal{L}_T^*$ has the same limiting distribution as that of $\mathcal{L}_T$, which implies the following result.

LEMMA D.5. *Assume that Conditions (C1)–(C7) hold, then $\rho(\mathcal{L}_T^*, \mathcal{L}_T) \to 0$ as $T, m_T \to \infty$, for any metric ρ that metrizes weak convergence of distributions.*

D.2. Proof of Theorem 5.1.

PROOF. To show the validity of the smoothed regression bootstrap, we just need to verify Conditions *(C1)–(C7)* hold with the probability approaching 1, conditionally on the data $\{\boldsymbol{W}_t = (Y_t, \boldsymbol{X}_t, \boldsymbol{Z}_t)\}_{t=1}^T$, where under the bootstrap distribution $\widehat{\mathbb{Q}}_h$, the bootstrap sample $(\boldsymbol{X}^*, \boldsymbol{Z}^*) \sim \tilde{F}(\mathbf{x}, \boldsymbol{z})$, whose density function is the nonparametric density estimator $\tilde{f}(\mathbf{x}, \boldsymbol{z})$. First, under Assumptions 6.(i) and (iii), we have $\|\tilde{f}(\mathbf{x}, \boldsymbol{z}) - f(\mathbf{x}, \boldsymbol{z})\|_\infty = o_p(1)$, as a standard result in kernel regression estimation (Györfi et al., 1989 and Hansen, 2008). Conditioning on $(\boldsymbol{X}^*, \boldsymbol{Z}^*)$, the noise $\varepsilon^* \sim \tilde{\sigma}(\boldsymbol{X}^*, \boldsymbol{Z}^*)e^*$, where $e^* \sim \widehat{F}_e$ which is independent of $\tilde{f}(\mathbf{x}, \boldsymbol{z})$. The bootstrap response is generated from

$$Y^* = \sum_{k=1}^{4} (\boldsymbol{X}^*)^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_k \mathbb{1}\{\boldsymbol{Z}^* \in R_k(\widehat{\boldsymbol{\gamma}})\} + \varepsilon_Q^*. \tag{D.23}$$

Condition *(C1)* is a direct consequence of Theorem 3.1. Let $\tilde{f}(\mathbf{x}) = \int \tilde{f}(\mathbf{x}, \boldsymbol{z}) d\boldsymbol{z}$ and $f(\mathbf{x})$ be the density of $\boldsymbol{X}$ under $\mathbb{P}_0$. Then we have $\|\tilde{f}(\mathbf{x}) - f(\mathbf{x})\|_\infty$ converges to 0 in probability, which is implied by $\|\tilde{f}(\mathbf{x}, \boldsymbol{z}) - f(\mathbf{x}, \boldsymbol{z})\|_\infty \xrightarrow{P} 0$ and the dominidated convergence theorem. Therefore, $\widehat{\mathbb{Q}}_h(\|\boldsymbol{X}\|^4) = \int \|\mathbf{x}\|^4 \tilde{f}(\mathbf{x}) d\mathbf{x}$ converges to $\mathbb{P}_0(\|\boldsymbol{X}\|^4) < \infty$ by the dominidated convergence theorem, which verifies the first condition in *(C2)*. For the second condition of the boundness of $\widehat{\mathbb{Q}}_h(\varepsilon^4)$, we notice that by the independence of $\widehat{F}_e$ and $\tilde{f}(\mathbf{x}, \boldsymbol{z})$,

$$\widehat{\mathbb{Q}}_h(\varepsilon^4) = \int \tilde{\sigma}^4(\boldsymbol{X}, \boldsymbol{Z}) e^4 \tilde{f}(\mathbf{x}, \boldsymbol{z}) d\mathbf{x} d\boldsymbol{z} d\widehat{F}_e(e), \tag{D.24}$$

which is $O_p(1)$ because of (v) of Lemma D.6, the uniform boundness of $\tilde{\sigma}(\mathbf{x}, \boldsymbol{z})$ and $\tilde{f}(\mathbf{x}, \boldsymbol{z})$, which are also compactly supported. Therefore, we conclude that *(C2)* holds in probability approaching 1. Because $\varepsilon = \tilde{\sigma}(\boldsymbol{X}, \boldsymbol{Z})e$, where $e \sim \widehat{F}_e$ is independent of $(\boldsymbol{X}, \boldsymbol{Z})$ and has a zero mean, it holds that $\widehat{\mathbb{Q}}_h(\varepsilon | \boldsymbol{X}, \boldsymbol{Z}) = 0$. As a standard result in local linear regression, Assumptions 6. (i) and (ii) imply $\|\tilde{\sigma}(\mathbf{x}, \boldsymbol{z}) - \sigma(\mathbf{x}, \boldsymbol{z})\|_\infty \xrightarrow{P} 0$, which together with (iv) of Lemma D.6 leads to $\widehat{\mathbb{Q}}_h(\varepsilon_Q^2) \xrightarrow{P} \mathbb{P}_0(\varepsilon^2)$. Therefore, Condition *(C3)* holds in probability. Because $(\boldsymbol{X}, \boldsymbol{Z})$ has a compact support and $\|\tilde{f}(\mathbf{x}, \boldsymbol{z}) - f(\mathbf{x}, \boldsymbol{z})\|_\infty \xrightarrow{P} 0$, applying the dominated convergence therorem yields that (D.3) holds in probability. Therefore, *(C4)* is ensured.

To show *(C5)*, we first note that for any $l \in \{1, 2\}$,

$$\left| \tilde{f}_{q_{l,Q}|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z}) - f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z}) \right| = \left| \frac{\tilde{f}_{q_{l,Q},\boldsymbol{Z}_{-1,l}}(q, \boldsymbol{z})}{\tilde{f}_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})} - \frac{f_{q_l,\boldsymbol{Z}_{-1,l}}(q, \boldsymbol{z})}{f_{\boldsymbol{Z}_{-1,l}}(\boldsymbol{z})} \right| \xrightarrow{P} 0, \tag{D.25}$$

for q and $\boldsymbol{z}$ uniformly. Since $f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z})$ is bounded for each $\boldsymbol{z} \in \mathcal{Z}_{-1,l}$ and q_l in the neighborhood of 0 as required in Assumption 5. (ii), (D.25) implies that $\tilde{f}_{q_{l,Q}|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z})$ is bounded in probability. Then using the dominated convergence theorem, Condition *(C5)* can be shown. Assumption 6. (i) requires that $\mathcal{X} \times \mathcal{Z}$ is compact and implies that $f_{\boldsymbol{X}|\boldsymbol{Z}}$ is bounded. Hence, for any finite r,

$$\begin{aligned} \widehat{\mathbb{Q}}_h\left(\|\boldsymbol{X}\|^r | \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0\right) &= \int_{\mathcal{X} \times \mathcal{Z}} \|\mathbf{x}\|^r \mathbb{1}(\boldsymbol{z}^{\mathrm{T}} \boldsymbol{\gamma} = 0) \frac{\tilde{f}_{\boldsymbol{X}, \boldsymbol{Z}_l}(\mathbf{x}, \boldsymbol{z})}{\tilde{f}_{\boldsymbol{Z}_l}(\boldsymbol{z})} d\mathbf{x} d\boldsymbol{z} \\ &\xrightarrow{P} \mathbb{P}_0\left(\|\boldsymbol{X}\|^r | \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0\right), \end{aligned} \tag{D.26}$$

by the dominated convergence theorem. With the consistency of $\widehat{\boldsymbol{\gamma}}$ and Assumption 4, (D.26) implies the first two conditions in Condition *(C6)*. Since

$$\widehat{\mathbb{Q}}_h\left(\varepsilon^r | \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0\right) = \int_{\mathbb{R}} x^r d\widehat{F}_e(x) \int_{\mathcal{X} \times \mathcal{Z}} \tilde{\sigma}(\mathbf{x}, \boldsymbol{z}) \mathbb{1}(\boldsymbol{z}^{\mathrm{T}} \boldsymbol{\gamma} = 0) \frac{\tilde{f}_{\boldsymbol{X}, \boldsymbol{Z}_l}(\mathbf{x}, \boldsymbol{z})}{\tilde{f}_{\boldsymbol{Z}_l}(\boldsymbol{z})} d\mathbf{x} d\boldsymbol{z},$$

using Lemma D.6 (v) and Assumption 6. (ii) ensures that $\widehat{\mathbb{Q}}_h\left(\varepsilon^r | \boldsymbol{Z}_l^{\mathrm{T}} \boldsymbol{\gamma} = 0\right) < \infty$ for the r specified in Assumption 4 (iv). Hence, Condition *(C6)* is verified.

For *(C7)*, (i) is a direct consequence of $\|\tilde{f}(\mathbf{x},\boldsymbol{z}) - f(\mathbf{x},\boldsymbol{z})\|_\infty \xrightarrow{P} 0$. For (ii), recall that $\xi_Q^{(j,k)} = \left(\widehat{\boldsymbol{\delta}}_{jk}^{\mathrm{T}} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\delta}}_{jk} + 2\boldsymbol{X}^{\mathrm{T}}\widehat{\boldsymbol{\delta}}_{jk}\tilde{\sigma}(\boldsymbol{X},\boldsymbol{Z})e_Q\right)\{\mathbb{1}^{(j)}(\widehat{\gamma}) + \mathbb{1}^{(k)}(\widehat{\gamma})\}$, to emphasis it is a function of $(\boldsymbol{X},\boldsymbol{Z},e,\widehat{\boldsymbol{\theta}})$, we write $\xi_Q^{(j,k)} = \xi(\boldsymbol{X},\boldsymbol{Z},e,\widehat{\boldsymbol{\theta}})$. Then,

$$
\begin{aligned}
\widehat{\mathbb{Q}}_h\left\{e^{it\xi_Q^{(j,k)}}|q_{l,Q}=0,\boldsymbol{Z}_{-1,l}\right\} &= \int e^{it\xi(\mathbf{x},\boldsymbol{z},e,\widehat{\boldsymbol{\theta}})}\mathbb{1}(\boldsymbol{z}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_l=0)\frac{\tilde{f}_{\boldsymbol{X},\boldsymbol{Z}_l}(\mathbf{x},\boldsymbol{z})}{\tilde{f}_{\boldsymbol{Z}_l}(\boldsymbol{z})}d\mathbf{x}d\boldsymbol{z}d\widehat{F}_e(e)\\
&\xrightarrow{P}\int e^{it\xi(\mathbf{x},\boldsymbol{z},e,\boldsymbol{\theta}_0)}\mathbb{1}(\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\gamma}_{l0}=0)\frac{f_{\boldsymbol{X},\boldsymbol{Z}_l}(\mathbf{x},\boldsymbol{z})}{f_{\boldsymbol{Z}_l}(\boldsymbol{z})}d\mathbf{x}d\boldsymbol{z}dF_e(e)\\
&=\mathbb{P}_0\left\{e^{it\xi^{(j,k)}}|q_l=0,\boldsymbol{Z}_{-1,l}\right\},
\end{aligned}
\tag{D.27}
$$

by Lemma D.6 (i) and the dominated convergence theorem. Therefore, *(C7)* (ii) holds in probability. Finally, for *(C7)* (iii) we note that for each $l\in\{1,2\}$,$\boldsymbol{z}_{-1,i}\in\mathcal{Z}_{-1,l}, q\in\mathbb{R}$ and $\varepsilon>0$, there exists $\delta>0$ such that if $|q|<\delta$,

$$
\begin{aligned}
&|\tilde{f}_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{Z}_{-1,l}) - \tilde{f}_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{-1,l})|\\
\le&\sum_{i=1}^{2}|\tilde{f}_{q_l|\boldsymbol{Z}_{-1,l}}(q_i|\boldsymbol{Z}_{-1,l}) - f_{q_l|\boldsymbol{Z}_{-1,i}}(q_i|\boldsymbol{Z}_{-1,l})| + |f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{Z}_{-1,l}) - f_{q_l|\boldsymbol{Z}_{-1,l}}(0|\boldsymbol{Z}_{-1,l})|,
\end{aligned}
$$

where $q_1=q$ and $q_2=0$. With (D.25), which shows the first term of the right-hand side of the above inequality is $o_p(1)$, and Assumption 5. (ii), which implies that for any $\varepsilon>0$, there exists $\delta>0$ such that the second term is less than ϵ provided that $|q|<\delta$, it can be shown that $\tilde{f}_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{Z}_{-1,l})$ is continuous at 0 for each $\boldsymbol{z}_{-1,l}$ in probability. Similarly, the continuity of $\widetilde{f}_{\xi_Q^{(j,k)}|(q_{l,Q},\boldsymbol{Z}_{-1,l})}(\xi|q,\boldsymbol{z})$ can be shown. Hence, Condition *(C7)* holds with the probability approaching 1. Finally, with Conditions *(C1)*–*(C7)* verified, Theorem 5.1 follows by applying Lemma D.4. □

LEMMA D.6. *Let F_e and φ_e be the distribution function and characteristic function of e, respectively. Then under Assumptions 1-6,*

(i) for any $\eta>0$, $\sup_{|\xi|\le\eta}\left\{\left|\int\exp(i\xi x)d\widehat{F}_e - \varphi_e(\xi)\right|\right\}\xrightarrow{P}0$;

(ii) $\|\widehat{F}_e - F_e\|_\infty\xrightarrow{P}0$;

(iii) $\int|x|d\widehat{F}_e(x)\xrightarrow{P}\mathbb{P}_0(|e|)$;

(iv) $\int x^2 d\widehat{F}_e(x)\xrightarrow{P}1$;

(v) $\int x^r d\widehat{F}_e(x) = O_p(1)$, where r is specified in Assumption 4.

PROOF. (i) Let $F_{T,e}$ be the empirical distribution function of $\{e_t\}_{t=1}^T$. Note that

$$
\int\exp(i\xi x)d\widehat{F}_e(x) = \exp(-it\bar{e}_T)\,\mathbb{P}_T\{\exp(i\xi\widehat{e}_t)\}.
$$

Hence, for any $|\xi|\le\eta$ with $\eta>0$, we have

$$
\begin{aligned}
&\left|\int\exp(i\xi x)d\widehat{F}_e - \exp(-i\xi\bar{e}_T)\int\exp(i\xi x)dF_{T,e}(x)\right|\\
=&|\mathbb{P}_T\{\exp(i\xi\widehat{e}_t)\} - \mathbb{P}_T\{\exp(i\xi e_t)\}| \le |\eta|\mathbb{P}_T(|\widehat{e}_t - e_t|).
\end{aligned}
\tag{D.28}
$$

We claim that

$$
\mathbb{P}_T(|\widehat{e}_t - e_t|)\xrightarrow{P}0,
\tag{D.29}
$$

which will be shown later. Then (D.28) implies that

$$\sup_{|\xi|\le\eta}\left\{\left|\int \exp(i\xi x)d\widehat{F}_e - \exp\left(-i\xi\bar{e}_T\right)\int \exp(i\xi x)dF_{T,e}(x)\right|\right\} \xrightarrow{P} 0, \tag{D.30}$$

and Lemma D.6 (i) follows from the facts that $\bar{e}_T = \mathbb{P}_T(\widehat{e}_t) \xrightarrow{P} 0$, and

$$\sup_{|\xi|\le\eta}\left|\mathbb{P}_T\left\{\exp(i\xi e_t)\right\} - \mathbb{P}_0\left\{\exp(i\xi e_t)\right\}\right| \xrightarrow{P} 0$$

by the ULLN under mixing sequences.

It remains to verify the claim (D.29). For notational simplicity, we denote $\widehat{\sigma}_t := \tilde{\sigma}(\boldsymbol{X}_t, \boldsymbol{Z}_t)$ and $\sigma_t := \sigma(\boldsymbol{X}_t, \boldsymbol{Z}_t)$. Then $\sup_{1\le t\le T}|\sigma_t - \widehat{\sigma}_t| = o_p(1)$ by Assumption 5.(ii). Note that

$$\begin{aligned}
\widehat{e}_t =& \frac{Y_t - \sum_{k=1}^4 \boldsymbol{X}_t^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_k \mathbb{1}_t^{(k)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t} \\
=& \frac{\sum_{j=1}^4 \boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j0}\right)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t} + \frac{\sum_{j=1}^4\sum_{i\ne j}^K \boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{j0}\right)\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t} \\
&+ \frac{\sigma_t - \widehat{\sigma}_t}{\widehat{\sigma}_t} e_t + e_t =: E_{1,t} + E_{2,t} + E_{3,t} + e_t, \quad \text{say.}
\end{aligned} \tag{D.31}$$

Denote $\hat{E}_{k,T} = \mathbb{P}_T(|E_{k,t}|)$ for $k = 1, 2, 3$. Then to show (D.29), it suffices to show $\hat{E}_{k,T} \xrightarrow{P} 0$ as $T \to \infty$. For the first term $E_{1,T}$, we have

$$\begin{aligned}
\hat{E}_{1,T} \le& \sum_{j=1}^4 \mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j0}\right)\mathbb{1}_t^{(j)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t}\right|\right\} \\
\le& \sum_{j=1}^4 \mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j0}\right)}{\sigma_t + o_p(1)}\right|\right\} \le \sum_{j=1}^4 \mathbb{P}_T\left\{\left|\frac{\|\boldsymbol{X}_t\|\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j0}\|}{\sigma_t + o_p(1)}\right|\right\} \\
=& O_p(T^{-1/2}),
\end{aligned} \tag{D.32}$$

since $\sigma_t > \underline{\sigma} > 0$ and $\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j0}\| = O_p(T^{-1/2})$. For the second term $E_{2,T}$, it is $o_p(1)$ if for each $i \ne j \in \{1, \cdots, 4\}$, $\mathbb{P}_T\left\{\left|\boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{j0}\right)\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})/\widehat{\sigma}_t\right|\right\} = o_p(1)$, which can be shown as

$$\begin{aligned}
&\mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}\left(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{j0}\right)\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t}\right|\right\} \\
\le& \mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{i0})}{\widehat{\sigma}_t}\right|\right\} + \mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{ij,0}\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t}\right|\right\},
\end{aligned}$$

where the first term is $O_p(T^{-1/2})$ from the same reason as for (D.32). For the second term,

$$\mathbb{P}_T\left\{\left|\frac{\boldsymbol{X}_t^{\mathrm{T}}\boldsymbol{\delta}_{ij,0}\mathbb{1}_t^{(i)}(\boldsymbol{\gamma}_0)\mathbb{1}_t^{(j)}(\widehat{\boldsymbol{\gamma}})}{\widehat{\sigma}_t}\right|\right\} \le \sum_{l=1}^L \mathbb{P}_T\left\{\frac{\|\boldsymbol{X}_t\|\|\boldsymbol{\delta}_{ij,0}\|}{\sigma_t + o_p(1)}\left|\mathbb{1}_{l,t}(\gamma_{l0}) - \mathbb{1}_{l,t}(\widehat{\gamma}_l)\right|\right\},$$

which is $O_p(T^{-1})$ because of (A.20) in Lemma A.4. Therefore, we obtain $\hat{E}_{2,T} = O_p(T^{-1/2})$. For the third term $\hat{E}_{3,T}$, it holds that

$$\hat{E}_{3,T} = \mathbb{P}_T\left(\left|\frac{\sigma_t - \widehat{\sigma}_t}{\widehat{\sigma}_t}e_t\right|\right) \le \sqrt{\mathbb{P}_T\left(\left|\frac{\sigma_t - \widehat{\sigma}_t}{\widehat{\sigma}_t}\right|^2\right)}\sqrt{\mathbb{P}_T(e_t^2)}.$$

Since $\mathbb{P}_T(e_t^2) = O_p(1)$, and $|\sigma_t - \widehat{\sigma}_t| = o_p(1), \sigma_t < \underline{\sigma}$ uniformly for $t \in \{1, \cdots, T\}$, it yields that $E_{3,T} = o_p(1)$. Combining with (D.31), it yields that for any $t \in \{1, \cdots, T\}$.

$$\mathbb{P}_T(|\widehat{e}_t - e_t|) \le \hat{E}_{1,T} + \hat{E}_{2,T} + \hat{E}_{3,T} = o_p(1), \tag{D.33}$$

which verifies the claim (D.29), and thus completes the proof for (i).

(ii) By Levy-Cramer continuity theorem, (i) implies that $\widehat{F}_e(x) = F_e(x) + o_p(1)$ for any $x \in \mathbb{R}$. Then (ii) follows from the continuity of F_e and Polya's theorem.

(iii) Note that

$$\begin{aligned}\left|\int |x| d\widehat{F}_e(x) - \mathbb{P}_T(|e_t|)\right| =& |\mathbb{P}_T(|\widehat{e}_t - \bar{e}_T| - |e_t|)| \\ \le& \mathbb{P}_T(|\widehat{e}_t - e_t|) + |\bar{e}_T| \xrightarrow{P} 0,\end{aligned}$$

implied by (D.29) and $\bar{e}_T = \mathbb{P}_T(\widehat{e}_t) \xrightarrow{P} 0$. Because $\mathbb{P}_T(|e_t|) = \mathbb{P}_0(|e|) + o_p(1)$ by the weak law of large numbers, the conclusion (iii) follows.

(iv) Since $\int x^2 d\widehat{F}_e(x) = \mathbb{P}_T\left(\widehat{e}_t^2\right) - (\bar{e}_T)^2 = \mathbb{P}_T\left(\widehat{e}_t^2\right) + o_p(1)$ and $\mathbb{P}_T(e_t^2) = \mathbb{P}_0(e^2) + o_p(1) = 1 + o_p(1)$, to show (iv) it is sufficient to show that $\mathbb{P}_T\left(\widehat{e}_t^2\right) - \mathbb{P}_T(e_t^2) = o_p(1)$. From (D.31) we have

$$\widehat{e}_t^2 - e_t^2 = (E_{1,t} + E_{2,t} + E_{3,t})^2 + 2(E_{1,t} + E_{2,t} + E_{3,t})e_t, \tag{D.34}$$

which implies that

$$\begin{aligned}\left|\mathbb{P}_T\left(\widehat{e}_t^2\right) - \mathbb{P}_T(e_t^2)\right| \le& \mathbb{P}_T\left(\left|\widehat{e}_t^2 - e_t^2\right|\right) \\ \le& 3\sum_{i=1}^3 \mathbb{P}_T(E_{i,t}^2) + 2\sqrt{\mathbb{P}_T\{(\sum_{i=1}^3 E_{i,t})^2\}}\sqrt{\mathbb{P}_T(e_t^2)} \\ \le& 3\sum_{i=1}^3 \mathbb{P}_T(E_{i,t}^2) + 2\sqrt{3\sum_{i=1}^3 \mathbb{P}_T(E_{i,t}^2)}\sqrt{1 + o_p(1)},\end{aligned}$$

by the C_r and Cauchy-Schwartz inequalities. Therefore, $\mathbb{P}_T\left(\widehat{e}_t^2\right) - \mathbb{P}_T(e_t^2) = o_p(1)$ if $\mathbb{P}_T(E_{i,t}^2) = o_p(1)$ for $i = 1, 2, 3$. Since this can be shown in the almost same way as for showing $\mathbb{P}_T(|E_{i,t}|) = o_p(1)$ in the proof of (i), we omit the detailed proof here for simplicity.

(v) Note that

$$\int |x|^r d\widehat{F}_e(x) \le \sum_{i=0}^r \binom{r}{i} |\bar{e}_T|^i \mathbb{P}_T(\widehat{e}_t^{r-i}), \tag{D.35}$$

and $|\bar{e}_T|^i = |\mathbb{P}_T(\widehat{e}_t)|^i = o_p(1)$ for each $1 \le i \le r$. Using the expansion (D.31) and the fact that $\mathbb{P}_T(|e_t|^i) = \mathbb{P}_0(|e_t|^i) + o_p(1)$, it is straightforward to show that $\mathbb{P}_T(\widehat{e}_t^i) = O_p(1)$ for each $1 \le i \le r$. Therefore, the desired result (v) is verified. □

APPENDIX E: PROOFS FOR SECTION 6

E.1. Proof of Theorem 6.1. In this subsection, we present the proof for Theorem 6.1 of the main paper on the convergence of the four-regime based LS estimator under the segmented models with less than four regimes.

PROOF. Suppose that the true model is

$$Y = \sum_{k=1}^{K_0} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_{k0} \mathbb{1}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0)\} + \varepsilon, \tag{E.1}$$

where the number of regimes $K_0 \le 4$ and the number of splitting hyperplanes $L_0 \le 2$. In particular, $R_k(\boldsymbol{\gamma}_0) = \mathcal{Z}_1 \times \mathcal{Z}_2$ for the global linear model ($K_0 = 1$), the splitting coefficient $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_{10}$ or $\boldsymbol{\gamma}_{20}$ for $L_0 = 1$, and $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{10}^{\mathrm{T}}, \boldsymbol{\gamma}_{20}^{\mathrm{T}})^{\mathrm{T}}$ for $L_0 = 2$.

For a candidate $\boldsymbol{\theta} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$, we let $\{R_j^{(4)}(\boldsymbol{\gamma})\}_{j=1}^4$ be the four regimes under $\boldsymbol{\gamma}$, and denote $\mathcal{G} = \{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2\}$ and $\mathcal{B} = \{\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_4\}$. Then, the population of the LS criterion function based on the four-regime model is

$$\mathbb{M}(\boldsymbol{\theta}) = \mathbb{E}[\{Y - \sum_{j=1}^{4} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_j \mathbb{1}\{\boldsymbol{Z} \in R_j^{(4)}(\boldsymbol{\gamma})\}\}^2].$$

Suppose that when the data is generated from Model (E.1) with $K_0 \le 4$, $\mathbb{M}(\boldsymbol{\theta})$ is minimized at $\boldsymbol{\theta}_* = (\boldsymbol{\gamma}_*^{\mathrm{T}}, \boldsymbol{\beta}_*^{\mathrm{T}})^{\mathrm{T}}$. Let $\mathcal{G}_* = \{\boldsymbol{\gamma}_{1*}, \boldsymbol{\gamma}_{2*}\}$ and $\mathcal{B}_* = \{\boldsymbol{\beta}_{1*}, \cdots, \boldsymbol{\beta}_{4*}\}$, representing the true parameters under the four-segment model. In the case of $K_0 = 4$, we have shown that $\boldsymbol{\theta}_* = \boldsymbol{\theta}_0$ in Proposition 1. Now we show that when $K_0 < 4$, the true parameters $\boldsymbol{\gamma}_0$ and $\boldsymbol{\beta}_0$ are elements of $\mathcal{G}_*$ and $\mathcal{B}_*$, respectively. That is, we are to show that $d(\boldsymbol{\gamma}_0, \mathcal{G}_*) = 0$ and $d(\boldsymbol{\beta}_{k0}, \mathcal{B}_*) = 0$ for $k = 1, 2$. Without loss of generality, we take $L_0 = 1$ and $K_0 = 2$ in this proof, which makes Model (E.1) to be the two-regime model (6.3) of the main paper. The proof for the other degenerated models can be shown similarly.

Note that

$$\begin{aligned}
\mathbb{M}(\boldsymbol{\theta}) &= \mathbb{E}[\{Y - \sum_{j=1}^{4} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_j \mathbb{1}\{\boldsymbol{Z} \in R_j^{(4)}(\boldsymbol{\gamma})\}\}^2] \\
&= \mathbb{E}[\varepsilon^2 + \{\sum_{k=1}^{2} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_{k0} \mathbb{1}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0)\} - \sum_{j=1}^{4} \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_j \mathbb{1}\{\boldsymbol{Z} \in R_j^{(4)}(\boldsymbol{\gamma})\}\}^2] \\
&= \mathbb{E}(\varepsilon^2) + \sum_{k=1}^{2} \sum_{j=1}^{4} \mathbb{E}[\{\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{\beta}_{k0} - \boldsymbol{\beta}_j)\}^2 \mathbb{1}\{\boldsymbol{Z} \in R_k(\boldsymbol{\gamma}_0) \cap R_j^{(4)}(\boldsymbol{\gamma})\}] \\
&= \mathbb{E}(\varepsilon^2) + \sum_{k=1}^{2} \sum_{j=1}^{4} A_{k,j}(\boldsymbol{\theta}), \text{ say,}
\end{aligned} \tag{E.2}$$

where the second equality is due to $\mathbb{E}(\varepsilon | \boldsymbol{X}, \boldsymbol{Z}) = 0$. At $\boldsymbol{\theta} = \boldsymbol{\theta}_*$, it can be shown that $A_{k,j}(\boldsymbol{\theta}_*) = 0$ for any k, j. Hence $\mathbb{M}(\boldsymbol{\theta}_*) = \mathbb{E}(\varepsilon^2)$.

Suppose that $d(\boldsymbol{\gamma}_0, \mathcal{G}) \ne 0$, namely $\boldsymbol{\gamma}_1 \ne \boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_2 \ne \boldsymbol{\gamma}_0$. Then the true splitting hyperplane $H_0 : \boldsymbol{z}^{\mathrm{T}} \boldsymbol{\gamma}_0 = 0$ will partition through at least one region $R_j^{(4)}(\boldsymbol{\gamma})$ for $j \in \{1, \cdots, 4\}$. By Assumption S2 (i) we have $\mathbb{P}\left\{\boldsymbol{Z} \in R_1(\boldsymbol{\gamma}_0) \cap R_j^{(4)}(\boldsymbol{\gamma})\right\} > 0$ and $\mathbb{P}\left\{\boldsymbol{Z} \in R_2(\boldsymbol{\gamma}_0) \cap R_j^{(4)}(\boldsymbol{\gamma})\right\} > 0$. Therefore,

$$A_{1,j}(\boldsymbol{\theta}) \ge \lambda_0 \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{10}\|^2, \quad A_{2,j}(\boldsymbol{\theta}) \ge \lambda_0 \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{20}\|^2$$

according to Assumption S2 (ii). Since $\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20}$, either $A_{1,j}(\boldsymbol{\theta}) > 0$ or $A_{2,j}(\boldsymbol{\theta}) > 0$. Consequently, $\mathbb{M}(\boldsymbol{\theta}) \geq \mathbb{M}(\boldsymbol{\theta}_*) + A_{k,h}(\boldsymbol{\theta}) + A_{j,h}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_*)$.

Suppose that $d(\boldsymbol{\gamma}_0, \mathcal{G}) = 0$, namely $\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_0$ or $\boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_0$, while $d(\boldsymbol{\beta}_{k0}, \mathcal{B}) \neq 0$ for $k \in \{1, 2\}$. In such case, there exits $j \in \{1, \cdots, 4\}$ such that $R_j^{(4)}(\boldsymbol{\gamma}) \subset R_k(\boldsymbol{\gamma}_0)$. Hence

$$A_{k,j}(\boldsymbol{\theta}) = \mathbb{E}\left[\{\boldsymbol{X}_t^{\mathrm{T}}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{k0})\}^2 \mathbb{1}\{\boldsymbol{Z}_t \in R_k(\boldsymbol{\gamma}_0)\}\right] \geq \lambda_0 \|\boldsymbol{\beta}_j - \boldsymbol{\beta}_{k0}\|^2 > 0,$$

by Assumption S2 (ii). Therefore, $\mathbb{M}(\boldsymbol{\theta}) \geq \mathbb{M}(\boldsymbol{\theta}_*) + A_{k,j}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_*)$.

Combining the two cases yields that

$$\mathbb{M}(\boldsymbol{\theta}) > \mathbb{M}(\boldsymbol{\theta}_*) \quad \text{for any } \boldsymbol{\theta} \in \Theta \tag{E.3}$$

if either $d(\boldsymbol{\gamma}_0, \mathcal{G}) \neq 0$ or $d(\boldsymbol{\beta}_{k0}, \mathcal{B}) \neq 0$ for some $k \in \{1, 2\}$. Therefore, $\boldsymbol{\theta}_*$ as the minimizer of $\mathbb{M}(\boldsymbol{\theta})$ must satisfy $d(\boldsymbol{\gamma}_0, \mathcal{G}_*) = 0$ and $d(\boldsymbol{\beta}_{k0}, \mathcal{B}_*) = 0$ for $k \in \{1, 2\}$.

Having established the minimizer of the least square criterion function under the population level, the rest of the proof for the convergence rate of the LS estimator under Assumptions S3 and S4, follows the similar arguments as in Appendix B for the four-regime case. □

E.2. Proof of Theorem 6.2.

PROOF. Suppose that the true model is given by (E.1) with the K_0 true regimes being $\{R_{k0}\}_{k=1}^{K_0}$ and the true regression coefficients are $\{\boldsymbol{\beta}_{k0}\}_{k=1}^{K_0}$ respectively. Let the estimated regimes and the estimated regression coefficients under the four-regime model be $\{\widehat{R}_j^{(4)}\}_{j=1}^4$ and $\{\widehat{\boldsymbol{\beta}}_j^{(4)}\}_{j=1}^4$, respectively.

For any $1 \leq K \leq 4$, let

$$\mathcal{C}_T(K) = \log\left(\frac{S_T(K)}{T}\right) + \frac{\lambda_T}{T}K,$$

where $\lambda_T \to \infty$ and $\lambda_T/T \to 0$ as $T \to \infty$, and

$$S_T(4) = \sum_{t=1}^T \left[Y_t - \sum_{k=1}^4 \boldsymbol{X}_t^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_k^{(4)} \mathbb{1}\left\{\boldsymbol{Z}_t \in \widehat{R}_k^{(4)}\right\}\right]^2.$$

For $1 \leq K \leq 3$, define recursively

$$S_T(K) = S_T(K+1) + D_T^{(K+1)}(\widehat{i}, \widehat{h}),$$

where $(\widehat{i}_{K+1}, \widehat{h}_{K+1}) = \arg\min_{\mathcal{A}_{k+1}} D_T^{(K+1)}(i, h)$ and

$$\begin{aligned}
D_T^{(K)}(i,h) =& \min_{\boldsymbol{\beta} \in \mathbb{B}} \sum_{t=1}^T [Y_t - \boldsymbol{X}_t^{\mathrm{T}} \boldsymbol{\beta} \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_i^{(K)} \cup \widehat{R}_h^{(K)}\}]^2 \\
&- \sum_{t=1}^T [Y_t - \sum_{k=i,h} \boldsymbol{X}_t^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_k^{(K)} \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_k^{(K)}\}]^2 \\
=:& S_{i,h}^{(K)} - T_{i,h}^{(K)}, \quad \text{say.}
\end{aligned}$$

First, we claim that for $K \geq K_0$, for each $1 \leq h \leq K_0$, there exists an index set $\mathcal{Q}_h^{(K)} \subset \{1, \cdots, K\}$ such that

$$\mathbb{P}\left\{\boldsymbol{Z}_t \in R_{h0} \triangle \cup_{i \in \mathcal{Q}_h^{(K)}} \widehat{R}_i^{(K)}\right\} = O(T^{-1}) \text{ and } \max_{i \in \mathcal{Q}_h^{(K)}} \|\boldsymbol{\beta}_{h0} - \widehat{\boldsymbol{\beta}}_i^{(K)}\| = O_p(T^{-1/2}). \tag{E.4}$$

We will prove the claim recursively. Specifically, we are to show that if (E.4) holds for $K = \tilde{K}$, then it also holds for $K = \tilde{K} - 1$, by showing that the index pair for merged regimes (i, h) from the $\tilde{K}$-segments model to the $(\tilde{K} - 1)$-segments model satisfies $(i, h) \in \mathcal{Q}_k^{(\tilde{K})}$ for some $1 \le k \le \tilde{K}$.

We start with $K = 4$, where (E.4) is ensured by Theorem 6.1. For the case of $K = 3$, we now show that $D_T^{(4)}(i, h) < D_T^{(4)}(i', h')$ if $\{i, h\} \subset \mathcal{Q}_k^{(4)}$ for some $1 \le k \le K_0$ and $\{i', h'\} \not\subset \mathcal{Q}_k^{(4)}$ for any $1 \le k \le K_0$, which implies that the selected merged regimes leading to the submodel with $K = 3$ are $\widehat{R}_i^{(4)}$ and $\widehat{R}_h^{(4)}$ which are asymptotically contained in the same regime R_{k0}.

Case (1). If two indices $\{i_k, h_k\} \subset \mathcal{Q}_k^{(4)}$, with some standard algebra, we can obtain

$$D_T^{(4)}(i_k, h_k) = S_{i_k,h_k}^{(4)} - T_{i_k,h_k}^{(4)} = \mathbf{H}_T(\widehat{R}_{i_k}^{(4)})^{\mathrm{T}} \boldsymbol{\Xi}_T^{-1} \mathbf{H}_T(\widehat{R}_{i_k}^{(4)}),$$

where

$$\mathbf{H}_T(\widehat{R}_{i_k}^{(4)}) = \{\mathbf{I}_p - \mathbf{G}_T(\widehat{R}_i^{(4)})\mathbf{G}_{k,T}^{-1}\}\sqrt{T}\mathbb{E}_T\left\{\varepsilon_t \boldsymbol{X}_t \mathbb{1}(\boldsymbol{Z}_t \in \widehat{R}_{i_k}^{(4)})\right\} \quad \text{and}$$

$$\boldsymbol{\Xi}_T = \mathbf{G}_T(\widehat{R}_{i_k}^{(4)}) - \mathbf{G}_T(\widehat{R}_{i_k}^{(4)})\mathbf{G}_{k,T}^{-1}\mathbf{G}_T(\widehat{R}_{i_k}^{(4)}),$$

with $\mathbf{G}_T(\widehat{R}_{i_k}^{(4)}) = \mathbb{E}_T[\boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_{i_k}^{(4)}\}]$ and $\mathbf{G}_{k,T} = \mathbb{E}_T[\boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \mathbb{1}\{\boldsymbol{Z}_t \in R_{k0}\}]$ for each $1 \le k \le K_0$ and $i_k \in \mathcal{Q}_k^{(4)}$. Using the martingale central limit theorem and the uniform law of large numbers, it can be easily seen that

$$D_T^{(4)}(i_k, h_k) = O_p(1), \quad \text{if } \{i_k, h_k\} \subset \mathcal{Q}_k^{(4)} \ \text{ for each } 1 \le k \le K_0. \tag{E.5}$$

Case (2). If the two indices $\{i, h\} \not\subset \mathcal{Q}_k^{(4)}$ for any $1 \le k \le K_0$. Suppose that $i \in \mathcal{Q}_{\tilde{i},}^{(4)}$ and $h \in \mathcal{Q}_{\tilde{h}}^{(4)}$, for some $1 \le \tilde{i}, \tilde{h} \le K_0$. Then Theorem 6.1 implies that that $\mathbb{P}\{\boldsymbol{Z}_t \in \widehat{R}_i^{(4)} \setminus R_{\tilde{i}0}\} = O_p(1/T)$, $\|\boldsymbol{\beta}_{\tilde{i}0} - \widehat{\boldsymbol{\beta}}_i\| = O_p(1/\sqrt{T})$, and the same consistency also holds for $\widehat{R}_h^{(4)}$ and $\widehat{\boldsymbol{\beta}}_h$. Then standard algebra leads to

$$T_{i,h}/T = \mathbb{E}_T[\varepsilon_t^2 \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_i^{(4)} \cup \widehat{R}_h^{(4)}\}] + o_p(1), \text{ and} \tag{E.6}$$

$$S_{i,h}/T = \mathbb{E}_T[\varepsilon_t^2 \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_i^{(4)} \cup \widehat{R}_h^{(4)}\}] + \boldsymbol{\delta}_{\tilde{i}\tilde{h},0}^{\mathrm{T}} \mathbf{G}_T(\widehat{R}_i^{(4)})\mathbf{G}_T(\widehat{R}_{i\cup h}^{(4)})^{-1}\mathbf{G}_T(\widehat{R}_h^{(4)})\boldsymbol{\delta}_{\tilde{i}\tilde{h},0} + o_p(1),$$

where $\mathbf{G}_T(\widehat{R}_{i\cup h}^{(4)}) = \mathbb{E}_T[\boldsymbol{X}_t \boldsymbol{X}_t^{\mathrm{T}} \mathbb{1}\{\boldsymbol{Z}_t \in \widehat{R}_i^{(4)} \cup \widehat{R}_h^{(4)}\}]$. By Assumption S2 and the ULLN, the smallest eigenvalue of $\mathbf{G}_T(\widehat{R}_i^{(4)})\mathbf{G}_T(\widehat{R}_{i\cup h}^{(4)})^{-1}\mathbf{G}_T(\widehat{R}_h^{(4)})$ is asymptotically bounded away from some constant $\lambda_1 > 0$. Since $\boldsymbol{\delta}_{\tilde{i}\tilde{h},0} = \boldsymbol{\beta}_{\tilde{i}0} - \boldsymbol{\beta}_{\tilde{h}0} \neq \mathbf{0}$ as required in Assumption S2, from (E.6) we obtain

$$D_T^{(4)}(i, h) = S_{i,h} - T_{i,h} = O_p(T), \quad \text{if } \{i, h\} \not\subset \mathcal{Q}_k^{(4)} \ \text{ for any } 1 \le k \le K_0. \tag{E.7}$$

This together with (E.5) and (E.7) implies that the optimal regime merger from $K = 4$ to $K = 3$ is the pair of regimes that are contained in the same $\mathcal{Q}_k^{(4)}$ for some $1 \le k \le K_0$. Hence, (E.4) with $K = 3$ is verified. Using the same argument the claim (E.4) with $K = 2$ and 1 can also be established, respectively, provided that $K \ge K_0$.

(E.4) implies that with some relabelling,

$$\mathbb{P}\{\widehat{R}_k^{(K_0)} \triangle R_{k0}\} = O(T^{-1}) \text{ and } \|\boldsymbol{\beta}_{k0} - \widehat{\boldsymbol{\beta}}_k\| = O_p(T^{-1/2}), \tag{E.8}$$

for each $1 \le k \le K_0$. which reveals that the back-elimination procedure consistently estimates the true model, if it can be shown that $\mathbb{P}(\widehat{K} = K_0) \to 1$ as $T \to \infty$.

We now show that $\mathbb{P}\{\mathcal{C}_T(K) < \mathcal{C}_T(K_0)\} \to 0$ when $K \neq K_0$, which ensures the model selection consistency.

(i) First, if $K < K_0$, by the definition of $\mathcal{C}(K)$, we have

$$\mathbb{P}\{\mathcal{C}_T(K) < \mathcal{C}_T(K_0)\} = \mathbb{P}\left\{\log\left(\frac{S_T(K)}{S_T(K_0)}\right) < \frac{\lambda_T(K_0 - K)}{T}\right\}. \tag{E.9}$$

As $\lambda_T/T \to 0$, to show the above probability converges to 0, it suffices to show that $\mathbb{P}\{S_T(K) > S_T(K_0)\} \to 1$. Note that (E.4) means that $|\mathcal{Q}_h^{(K_0)}| = 1$ for each $1 \le h \le K_0$. Similar to (E.7), it is straightforward to show that $D_T^{(K)}(\hat{i}_K, \hat{h}_K) > 0$ for each $1 \le K \le K_0$, meaning that any under-segmented models have increased sum of squared residuals. As $S_T(K) - S_T(K_0) = \sum_{k=K+1}^{K_0} D_T^{(k)}(\hat{i}_k, \hat{h}_k)$, we have $\mathbb{P}\{S_T(K) > S_T(K_0)\} \to 1$, which implies (E.9) converges to 0 as $\lambda_T/T \to 0$.

(ii) If $K > K_0$, meaning that the K-regime model is over-segmented, we have

$$\begin{aligned}\mathbb{P}\{\mathcal{C}_T(K) < \mathcal{C}_T(K_0)\} &= \mathbb{P}\left\{\log\left(\frac{S_T(K_0)}{S_T(K)}\right) > \frac{\lambda_T(K - K_0)}{T}\right\}\\ &= \mathbb{P}\left\{\frac{S_T(K_0) - S_T(K)}{S_T(K)/T} > T\left(e^{\frac{\lambda_T(K-K_0)}{T}} - 1\right)\right\},\end{aligned} \tag{E.10}$$

and $S_T(K_0) = S_T(K) + \sum_{k=K_0+1}^{K} D_T^{(k)}(\hat{i}_k, \hat{h}_k)$. Because of (E.5) we have $S_T(K_0) - S_T(K) = O_p(1)$. In addition, $S_T(K_0)/T = \mathbb{E}_T(\varepsilon_t^2) = O_p(1)$. By the Taylor expansion, $T\left(e^{\frac{\lambda_T(K-K_0)}{T}} - 1\right) = O(\lambda_T) \to \infty$. Hence, the probability in (E.10) converges to 0, .

Combining Cases (i) can (ii), we have $\mathbb{P}\{\mathcal{C}_T(K) < \mathcal{C}_T(K_0)\} \to 0$ if $K \neq K_0$. Since $\widehat{K} = \arg\min_{1 \le K \le 4} \mathcal{C}(K)$ and $1 \le K_0 \le 4$, it implies that $\mathbb{P}(\widehat{K} = K_0) \to 1$ as $T \to \infty$. This together with (E.8) completes the proof. □

APPENDIX F: AUXILIARY ASSUMPTIONS

F.1. Sufficient conditions for some assumptions. In this part, we provide some sufficient conditions for Assumptions 2.(i), 3.(ii), and 4.(i).

ASSUMPTION S1. (i) For each $l \in \{1,2\}$, let $q_l = \boldsymbol{Z}^\top \boldsymbol{\gamma}_{l0}$. There exists some $j \in \{1, \cdots, d_l\}$, such that for almost surely $\boldsymbol{Z}_{-1,l}$, the conditional density $f_{q_l|\boldsymbol{Z}_{-1,l}}(q)$ is continuous at $q = 0$ and $f_{q_{l,t}|\boldsymbol{Z}_{-1,l}}(0) \ge c_0$ for almost surely $\boldsymbol{Z}_{-1,l}$, where c_0 is a positive constant.

(ii) For each $l \in \{1,2\}$, there exists $c_1 > 0$ and $j \in [d_l]$ such that the conditional density $f_{q_l|\boldsymbol{Z}_{-1,l}}(q|\boldsymbol{z}) < c_1$ for almost surely $q \in \mathbb{R}$ and $\boldsymbol{z} \in \mathcal{Z}_{-1,l}$, where $\mathcal{Z}_{-1,l}$ is the support for the distribution of $\boldsymbol{Z}_{-1,l}$ and is a compact set in $\mathbb{R}^{d_l - 1}$.

(iii) For each $l \in \{1,2\}$, there exist some $j_l \in [d_l]$ and $c_2 > 0$ such that the conditional density $f_{(q_1,q_2)|(\boldsymbol{Z}_{-j_1,1},\boldsymbol{Z}_{-j_2,2})}(q_1, q_2|\boldsymbol{z}_1, \boldsymbol{z}_2) < c_2$ for almost surely $(q_1, q_2) \in \mathbb{R}^2$ and $(\boldsymbol{z}_1, \boldsymbol{z}_2) \in \mathcal{Z}_{-j_1,1} \times \mathcal{Z}_{-j_2,2}$, where $\mathcal{Z}_{-j_l,l}$ is the support for the distribution of $\boldsymbol{Z}_{-j_l.l}$ and is a compact set in $\mathbb{R}^{d_l-1}$ for each $l \in \{1,2\}$.

The following lemma shows that Assumption S1 implies Assumptions 2.(i), 3.(ii), 4.(i) and 4.(iii).

LEMMA F.1. *(i) Under Assumption S1 (i), there exists some constant $\delta_1 > 0$, if $\epsilon < \delta$, then $\mathbb{P}(|q_l| < \epsilon|\boldsymbol{Z}_{-1,l}) \ge c_0\epsilon/2$ almost surely, implying Assumptions 2.(i) and 4.(i).*

(ii) Under Assumption S1 (ii), there exist some positive constants δ_2 and c_1 such that if $\gamma_1, \gamma_2 \in \mathcal{N}(\gamma_{l0}; \delta_2)$, then $|\mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 < 0) - \mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\gamma_2 < 0)| \le c_3\|\gamma_1 - \gamma_2\|$, which ensures Assumptions 3.(ii).

(iii) Under Assumption S1 (iii),there exist some positive constants δ_3 and c_2 such that if $\gamma_1, \gamma_2 \in \mathcal{N}(\gamma_{10}; \delta_0)$ and $\gamma_3, \gamma_4 \in \mathcal{N}(\gamma_{20}, \delta_2)$, then $\mathbb{P}(\boldsymbol{Z}_1^{\mathrm{T}}\gamma_1 < 0 < \boldsymbol{Z}_1^{\mathrm{T}}\gamma_2, \boldsymbol{Z}_2^{\mathrm{T}}\gamma_3 < 0 < \boldsymbol{Z}_2^{\mathrm{T}}\gamma_4) \le c_2\|\gamma_1 - \gamma_2\|\|\gamma_3 - \gamma_4\|$, which ensures Assumptions 4.(iii).

PROOF. (i) The continuity of $f_{q_{l,t}|\boldsymbol{Z}_{-j,l}}(q)$ at $q = 0$ in Assumption S1 (i) implies that there exists $\delta_1 > 0$ such that $f_{q_l|\boldsymbol{Z}_{-1,l}}(|q|) \ge f_{q_l|\boldsymbol{Z}_{-1,l}}(0) - c_1/2 \ge c_1/2$. The assertion then follows from $\mathbb{P}(|q_l| < \epsilon|\boldsymbol{Z}_{-i,l}) = \int_{-\epsilon}^{\epsilon} f_{q_l|\boldsymbol{Z}_{-1,l}}(q)dq$.

(ii) Let $\Delta_l(\gamma) = \boldsymbol{Z}_l^{\mathrm{T}}(\gamma_{l0} - \gamma)$. Then for any $\gamma_1, \gamma_2 \in \mathcal{N}(\gamma_{l0}; \delta_1)$, where $\delta_1 < \delta_0/B$,

$$\mathbb{P}\left\{\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 > 0 > \boldsymbol{Z}_l^{\mathrm{T}}\gamma_2\right\} = \mathbb{P}\left\{\Delta_l(\gamma_1) < q_l < \Delta_l(\gamma_2)\right\} = E_{\boldsymbol{Z}_{-1,l}}\left\{\int_{\Delta_l(\gamma_1)}^{\Delta_l(\gamma_2)} f_{q_l|\boldsymbol{Z}_{-1,l}}(q)dq\right\}.$$

Let $M > 0$ be the constant such that $\|\boldsymbol{z}\|_\infty < M$ for all $\boldsymbol{z} \in \mathcal{Z}_{-j,l}$ and let $\delta_1 = \delta_0/M$, which ensures $\|\Delta_l(\gamma)\|_\infty \le \delta_0$ whenever $\gamma \in \mathcal{N}(\gamma_{l0}; \delta_1)$. It is then straightforward to see that $\mathbb{P}\left\{\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 > 0 > \boldsymbol{Z}_l^{\mathrm{T}}\gamma_2\right\} \le c_1 M\|\gamma_1 - \gamma_2\|$. Similary, $\mathbb{P}\left\{\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 < 0 < \boldsymbol{Z}_l^{\mathrm{T}}\gamma_2\right\}$ can be bounded in the same way. Since $|\mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 < 0) - \mathbb{P}(\boldsymbol{Z}_l^{\mathrm{T}}\gamma_2 < 0)| = \mathbb{P}\left\{\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 > 0 > \boldsymbol{Z}_l^{\mathrm{T}}\gamma_2\right\} + \mathbb{P}\left\{\boldsymbol{Z}_l^{\mathrm{T}}\gamma_1 < 0 < \boldsymbol{Z}_l^{\mathrm{T}}\gamma_2\right\}$, the desired result follows.

(iii) It follows from the similar argument as in (ii) and thus is omitted. □

F.2. Assumptions for degenerated models. The following assumption adapts Assumptions 2-4 of the main article for the segmented regression with the number of regimes $K_0 = 4$ and the number of splitting hyperplanes $L_0 = 2$ to the degenerated models with $1 \le K_0 \le 3$ and $0 \le L_0 \le 2$, which include Model (6.1)–(6.5) in the main article. Let $(Y, \boldsymbol{X}, \boldsymbol{Z}) \sim \mathbb{P}_0$. Suppose the data generated from a model

$$Y = \sum_{k=1}^{K_0} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}_{k0}\mathbb{1}\{\boldsymbol{Z} \in R_k(\gamma_0)\} + \varepsilon, \tag{F.1}$$

where the number of regimes $1 \le K_0 \le 3$ and the number of splitting hyperplanes $0 \le L_0 \le 2$. In particular, $R_k(\gamma_0) = \mathcal{Z}_1 \times \mathcal{Z}_2$ for the global linear model ($K_0 = 1$), the splitting coefficient $\gamma_0 = \gamma_{10}$ or γ_{20} when $L_0 = 1$, and $\gamma_0 = (\gamma_{10}^{\mathrm{T}}, \gamma_{20}^{\mathrm{T}})^{\mathrm{T}}$ when $L_0 = 2$. We use $\mathcal{L}_0 \subset \{1, 2\}$ to indicate the indices of the splitting hyperplanes. For instance, if the true model has two hyperplanes then $\mathcal{L}_0 = \{1, 2\}$; and if it has only one hyperplane $H_{20} = \{\boldsymbol{z}_2^{\mathrm{T}}\gamma_{20} = 0\}$ then $\mathcal{L}_0 = \{2\}$. The following assumptions are needed for Theorem 6.1.

ASSUMPTION S2. For each $i \in \mathcal{L}_0$ and $k, h \in \{1, \cdots, K_0\}$, the following conditions hold. (i) If $L_0 = 2$, then $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are not identically distributed. (ii) There exists a $j \in [d_i]$ such that $\mathbb{P}(|q_i| \le \epsilon|\boldsymbol{Z}_{-j,i}) > 0$ for almost surely $\boldsymbol{Z}_{-j,i}$ for any $\epsilon > 0$, where $\boldsymbol{Z}_{-j,i}$ is the vector after excluding $\boldsymbol{Z}_i$'s jth element. Without loss of generality, assume $j = 1$. (iii) For any $\gamma \in \Gamma_1 \times \Gamma_2$, if $\mathbb{P}\{\boldsymbol{Z} \in R_k(\gamma_0) \cap R_h(\gamma)\} > 0$, then the smallest eigenvalue of $\mathbb{E}\{\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}|\boldsymbol{Z} \in R_k(\gamma_0) \cap R_h(\gamma)\} \ge \lambda_0$ for some constant $\lambda_0 > 0$. (iv) If $(k, h) \in \mathcal{S}(i)$, then $\|\boldsymbol{\beta}_{k0} - \boldsymbol{\beta}_{h0}\| > c_0$ for some constant $c_0 > 0$, where $\mathcal{S}(i)$ is defined in (3).

ASSUMPTION S3. (i) $\mathbb{E}(Y^4) < \infty$, $\mathbb{E}(\|\boldsymbol{X}\|^4) < \infty$ and $\max_{i \in \mathcal{L}_0}\mathbb{E}(\|\boldsymbol{Z}_i\|) < \infty$. (ii) For each $i \in \mathcal{L}_0$, $\mathbb{P}(\boldsymbol{Z}_i^{\mathrm{T}}\gamma_1 < 0 < \boldsymbol{Z}_i^{\mathrm{T}}\gamma_2) \le c_1\|\gamma_1 - \gamma_2\|$ if $\gamma_1, \gamma_{2,} \in \mathcal{N}(\gamma_{i0}; \delta_0)$, for some constants $\delta_0, c_1 > 0$.

ASSUMPTION S4. (i) For $i \in \mathcal{L}_0$, there exist constants $\delta_1, c_2 > 0$ such that if $\epsilon \in (0, \delta_1)$ then $\mathbb{P}(|q_i| < \epsilon | \boldsymbol{Z}_{-1,i}) \geq c_2 \epsilon$ almost surely. (ii) For $i \in \mathcal{L}_0$, there exists a neighborhood $\mathcal{N}_i = \mathcal{N}(\boldsymbol{\gamma}_{i0}; \delta_2)$ for some $\delta_2 > 0$, such that $\inf_{\boldsymbol{\gamma} \in \mathcal{N}_i} \mathbb{E}(\|\boldsymbol{X}\| | \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\gamma} = 0) > 0$ almost surely. (iii) If $L_0 = 2$, then $\mathbb{P}(\boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{\gamma}_1 < 0 < \boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{\gamma}_2, \boldsymbol{Z}_2^{\mathrm{T}} \boldsymbol{\gamma}_3 < 0 < \boldsymbol{Z}_2^{\mathrm{T}} \boldsymbol{\gamma}_4) \leq c_3 \|\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2\| \|\boldsymbol{\gamma}_3 - \boldsymbol{\gamma}_4\|$ for some constant $c_3 > 0$ if $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \in \mathcal{N}_1$ and $\boldsymbol{\gamma}_3, \boldsymbol{\gamma}_4 \in \mathcal{N}_2$. (iv) For some constant $r > 8$, $\sup_{\boldsymbol{\gamma} \in \mathcal{N}_i} \mathbb{E}(\|\boldsymbol{X}\|^r | \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\gamma} = 0) < \infty$ and $\sup_{\boldsymbol{\gamma} \in \mathcal{N}_i} \mathbb{E}(\varepsilon^r | \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\gamma} = 0) < \infty$ almost surely.

APPENDIX G: EXTENSION TO GENERAL SEGMENTED REGRESSIONS

In this section, we discuss the extension of the proposed four-regime segmented regression to general segmented regressions with more than $L = 2$ splits. The range of numbers of regimes split by L hyperplanes is presented by the following result, whose proof can be seen in Buck (1943).

THEOREM G.1. *Suppose that there are $L \geq 1$ hyperplanes $H_l = \{\boldsymbol{z} \in \mathcal{Z} : \boldsymbol{z}^{\mathrm{T}} \boldsymbol{\gamma}_i = 0\}_{i=1}^{L}$. Then the number of regimes K split by these L hyperplanes satisfies*

$$L - 1 \leq K \leq \sum_{i=0}^{\min(L,d)} \binom{L}{i}. \tag{G.1}$$

REMARK G.1. The above bound in sharp and can be attained in general hyperplane arrangement (see e.g., Orlik and Terao, 2013). It reveals the challenges in the general segmented linear regressions. First, in the large or high dimensional setting where $d > L$, the right-hand of (G.1) becomes 2^L. It implies that each possible combination of the signs of the $\{\boldsymbol{z}^{\mathrm{T}} \boldsymbol{\gamma}_i, 1 \leq i \leq L\}$ determines a specific region. Under such a circumstance, the computation burdens will be quite high in both optimization and model selection to select among the models with $1 \leq K \leq 2^L$. Moreover, the increase of K can bring more risk of overfitting.

On the other hand, under the regime where $d < L$, the maximum number of region $K_{\max}$ is $\sum_{i=0}^{d} \binom{L}{i} = O(L^d)$. The main difficulty is in specifying the model form of segmented models , since it can be challenging to know which hyperplanes constitute the boundaries of each regime due to the complications of hyperplane arrangements. One possible solution is to via some data-driven method to determine the boundaries of each regime, while it brings more computational complexity and requires further studies.

APPENDIX H: ADDITIONAL SIMULATION RESULTS

H.1. Simulations under models with less than four regimes. This section presents results for the estimation based on the four-regime model when the underlying models were degenerated with less than four regimes. The true parameters for the degenerated were specified in Section 7.2 of the main paper. The data generating processes for $\{\boldsymbol{X}_t, \boldsymbol{Z}_{1,t}, \boldsymbol{Z}_{2,t}, \varepsilon_t\}_{t=1}^{T}$ included three the independence, the AR(0.2) and the MA(0.2) settings as that in Section 7.1 of the main paper. Table S2 summarizes the empirical averages of the L_2-distance between the sets of the true parameters and their estimates under the four-regime model: $D(\mathcal{G}_0, \widehat{\mathcal{G}})$ and $D(\mathcal{B}_0, \widehat{\mathcal{B}})$. In addition, to evaluate the cost of not knowing the number of the underlying regimes, we also estimated $\boldsymbol{\gamma}_0$ and $\boldsymbol{\beta}_0$ in the so-called oracle setting, in which the three degenerated models were known to have three or two regimes and the parameters were estimated by the LS estimators of the corresponding models, denoted by $\widehat{\boldsymbol{\gamma}}^{\mathrm{3REG}}, \widehat{\boldsymbol{\beta}}^{\mathrm{3REG}}$ and $\widehat{\boldsymbol{\gamma}}^{\mathrm{2REG}}, \widehat{\boldsymbol{\beta}}^{\mathrm{2REG}}$, respectively. The three-regime LS estimators were obtained via a new MIQP algorithm presented in Appendix C of the SM, while $\widehat{\boldsymbol{\beta}}^{\mathrm{2REG}}$ of the two-regime estimators were calculated by the algorithm of Lee et al. (2021).

Table S2 shows that the estimation errors as reflected by the distance measures $D(\mathcal{G}_0, \widehat{\mathcal{G}})$ and $D(\mathcal{B}_0, \widehat{\mathcal{B}})$ reduced as the sample sizes T was increased, confirming that the parameters of the degenerated models could be consistently estimated by the four-regime model. By comparing $D(\mathcal{G}_0, \widehat{\mathcal{G}})$ with $\|\boldsymbol{\gamma}_0 - \widehat{\boldsymbol{\gamma}}^{\text{3REG}}\|$ and $\|\boldsymbol{\gamma}_0 - \widehat{\boldsymbol{\gamma}}^{\text{2REG}}\|$ in Table S2, we found that the estimation errors for $\boldsymbol{\gamma}_0$ based on the four-regime model were about the same as those of $\widehat{\boldsymbol{\gamma}}^{\text{3REG}}$ or $\widehat{\boldsymbol{\gamma}}^{\text{2REG}}$, respectively, meaning that the four-regime estimators achieved similar level of accuracy as the estimators from the models with correctly specified number of regimes, for the boundary coefficient estimation. The reason is that the four-regime estimator can efficiently use the data points located near the underlying boundaries as $\widehat{\boldsymbol{\gamma}}^{\text{3REG}}$ or $\widehat{\boldsymbol{\gamma}}^{\text{2REG}}$ did. On the other hand, Table S2 shows that the estimation accuracy for the regression coefficients based on the four-regime model were inferior to the estimators based on the models with the true number of regimes when the sample size was small. This was expected since the four-regime based estimation made redundant regime partitions, and hence did not effectively used the subsample belonged to the same underlying regime.

TABLE S2

Empirical average $D(\mathcal{G}_0, \widehat{\mathcal{G}}), D(\mathcal{B}_0, \widehat{\mathcal{B}})$, which represent the L_2 distance between the set of true parameters and their estimates by the four-regime model, and $\|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}^{3REG}\|_2, \|\boldsymbol{\gamma}_0 - \widehat{\boldsymbol{\gamma}}^{3REG}\|_2$, or $\|\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}^{2REG}\|_2$ and $\|\boldsymbol{\gamma}_0 - \widehat{\boldsymbol{\gamma}}^{2REG}\|_2$ (multiplied by 10) under the independent (IND), auto-regressive (AR) and moving average (MA) settings for $\{\boldsymbol{X}_t^0, \boldsymbol{Z}_{1,t}^0, \boldsymbol{Z}_{2,t}^0\}_{t=1}^T$ of the three-regime model (a.2) and the two-regime model (b). The numbers inside the parentheses are the standard errors of the simulated averages.

	Three-regime model (a.1)											
T	IND				AR				MA			
	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$
200	0.55	4.53	0.53	4.33	0.67	4.14	0.61	3.96	0.68	4.29	0.66	3.99
	0.26	1.24	0.22	0.95	0.35	0.82	0.33	0.84	0.30	1.15	0.27	0.84
400	0.30	3.09	0.32	2.95	0.30	2.85	0.32	2.73	0.30	2.84	0.31	2.74
	0.18	0.72	0.17	0.61	0.15	0.67	0.17	0.57	0.17	0.64	0.16	0.60
800	0.14	2.24	0.16	2.15	0.15	1.92	0.15	2.01	0.15	1.96	0.15	1.95
	0.07	0.51	0.06	0.48	0.08	0.47	0.08	0.45	0.06	0.37	0.05	0.37
1600	0.08	1.49	0.08	1.48	0.08	1.32	0.07	1.31	0.07	1.38	0.07	1.38
	0.04	0.35	0.04	0.34	0.04	0.28	0.04	0.27	0.04	0.28	0.04	0.27
	Three-regime model (a.2)											
T	IND				AR				MA			
	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{3\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{3\text{REG}}$
200	2.08	4.91	2.09	4.64	2.29	4.58	2.27	4.23	2.21	4.53	2.25	4.28
	(1.52)	(1.09)	(1.58)	(1.01)	(1.64)	(1.08)	(1.70)	(0.95)	(1.61)	(1.05)	(1.59)	(0.97)
400	1.00	3.35	1.03	3.20	1.12	3.04	1.13	2.91	1.12	3.06	1.10	2.91
	(0.85)	(0.73)	(0.87)	(0.71)	(0.81)	(0.63)	(0.80)	(0.63)	(0.82)	(0.67)	(0.78)	(0.63)
800	0.49	2.31	0.48	2.26	0.53	2.08	0.51	1.98	0.49	2.14	0.49	2.06
	(0.35)	(0.47)	(0.35)	(0.46)	(0.39)	(0.45)	(0.38)	(0.44)	(0.35)	(0.44)	(0.36)	(0.43)
1600	0.26	1.62	0.26	1.58	0.24	1.48	0.24	1.44	0.24	1.51	0.23	1.47
	(0.18)	(0.33)	(0.18)	(0.33)	(0.16)	(0.31)	(0.17)	(0.29)	(0.17)	(0.31)	(0.17)	(0.30)
	Two-regime model											
T	IND				AR				MA			
	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{2\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{2\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{2\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{2\text{REG}}$	$D(\mathcal{G}_0, \widehat{\mathcal{G}})$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\gamma}}^{2\text{REG}}$	$\widehat{\boldsymbol{\beta}}^{2\text{REG}}$
200	0.55	3.25	0.54	2.85	0.59	2.95	0.60	2.54	0.64	3.26	0.64	2.59
	(0.44)	(0.83)	(0.43)	(0.74)	(0.45)	(0.78)	(0.48)	(0.68)	(0.49)	(0.80)	(0.48)	(0.68)
400	0.30	2.28	0.30	1.97	0.29	2.10	0.31	1.78	0.28	2.31	0.31	1.83
	(0.24)	(0.59)	(0.23)	(0.49)	(0.23)	(0.49)	(0.22)	(0.46)	(0.20)	(0.54)	(0.21)	(0.49)
800	0.14	1.69	0.14	1.41	0.14	1.25	0.15	1.23	0.15	1.49	0.14	1.29
	(0.10)	(0.43)	(0.11)	(0.35)	(0.12)	(0.32)	(0.13)	(0.32)	(0.11)	(0.36)	(0.11)	(0.32)
1600	0.07	1.02	0.07	0.97	0.06	0.93	0.07	0.88	0.07	0.94	0.07	0.90
	(0.05)	(0.27)	(0.06)	(0.25)	(0.05)	(0.23)	(0.05)	(0.22)	(0.05)	(0.24)	(0.05)	(0.23)

	Global linear model					
T	IND		AR		MA	
	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\beta}}^{\text{GLR}}$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\beta}}^{\text{GLR}}$	$D(\mathcal{B}_0, \widehat{\mathcal{B}})$	$\widehat{\boldsymbol{\beta}}^{\text{GLR}}$
200	1.81	1.33	1.48	1.22	1.87	1.19
	(0.67)	(0.49)	(0.55)	(0.47)	(0.62)	(0.45)
400	1.23	0.92	1.02	0.87	1.24	0.86
	(0.44)	(0.33)	(0.37)	(0.33)	(0.46)	(0.33)
800	0.83	0.69	0.78	0.59	0.85	0.64
	(0.23)	(0.23)	(0.27)	(0.22)	(0.30)	(0.22)
1600	0.62	0.46	0.51	0.43	0.54	0.43
	(0.24)	(0.18)	(0.18)	(0.15)	(0.18)	(0.16)

To gain further insights on the performances of the four-regime estimates under the degenerated models, we investigated the simulation results in more details by comparing adjacent estimated regression coefficients. Figure S1 displays the box plots of the squared distances between the estimated adjacent regression coefficients $\|\widehat{\boldsymbol{\beta}}_j - \widehat{\boldsymbol{\beta}}_k\|^2$ where the underlying samples were generated from the three-regime model (a.2). Figure S1 shows that as the sample size T was increased, $\|\widehat{\boldsymbol{\beta}}_1 - \widehat{\boldsymbol{\beta}}_2\|^2, \|\widehat{\boldsymbol{\beta}}_2 - \widehat{\boldsymbol{\beta}}_3\|^2$ and $\|\widehat{\boldsymbol{\beta}}_4 - \widehat{\boldsymbol{\beta}}_1\|^2$ converged to $\|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20}\|^2, \|\boldsymbol{\beta}_{20} - \boldsymbol{\beta}_{30}\|^2$ and $\|\boldsymbol{\beta}_{30} - \boldsymbol{\beta}_{10}\|^2$, respectively, while $\|\widehat{\boldsymbol{\beta}}_3 - \widehat{\boldsymbol{\beta}}_4\|^2$ decreased to 0, indicating that $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ were consistent estimates of $\boldsymbol{\beta}_{10}$ and $\boldsymbol{\beta}_{20}$, respectively, and both $\widehat{\boldsymbol{\beta}}_3$ and $\widehat{\boldsymbol{\beta}}_4$ converged to $\boldsymbol{\beta}_{30}$. Similar results for the two-regime model are also shown in Figure S2, which reveals that the estimated regression coefficients under the four-regime model could still provide consistent estimates to the underlying coefficients of the degenerated models.

Fig S1: Box plots for the squared distances of the estimated adjacent regression coefficient for the three-regime model (a.2). The red dashed lines indicate the squared distances of the true regression coefficients for the adjacent estimated regimes.

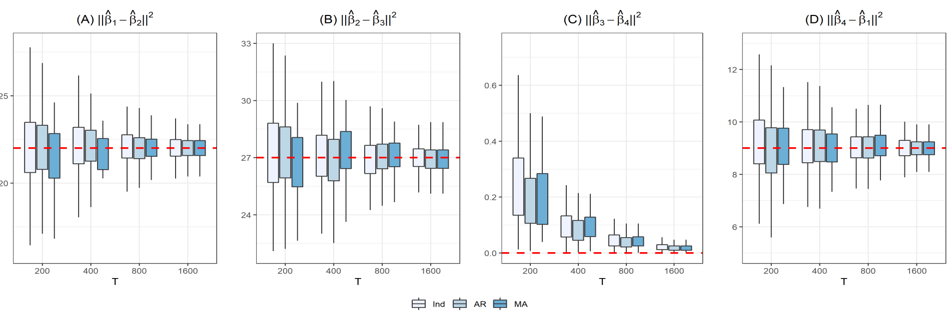


Fig S2: Box plots for the squared distances of the estimated adjacent regression coefficient for the two-regime model. The red dashed lines indicate the squared distances of the true regression coefficients of the three-regression model, with $\|\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20}\|^2 = 22$.

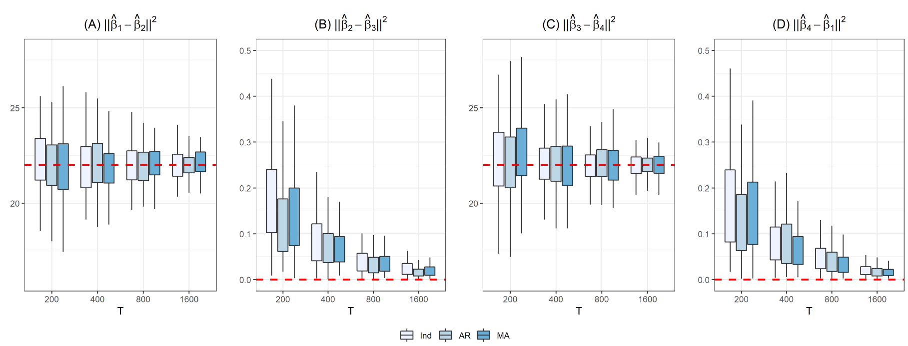


TABLE S3

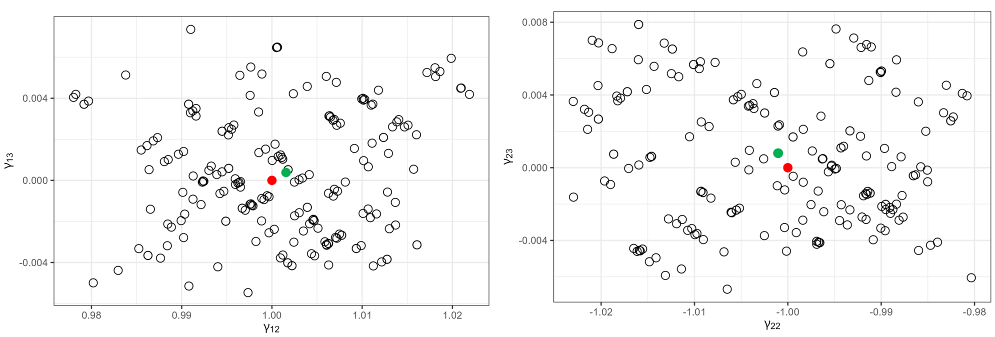
Empirical Model specification results under segmented models with regimes from $K_0 = 4$ to $K_0 = 1$ under 500 times replications. The performances were evaluated by the empirical average of the estimated number of regimes $\widehat{K}$, the discrepancy between the true regimes and the estimated regimes $D(\mathcal{R}, \widehat{\mathcal{R}})$ and the L_2 estimation error of regression coefficients $D(\mathcal{B}, \widehat{\mathcal{B}})$. The penalty parameter λ_T in the model selection criterion was set as $\lambda_T \in \{5, 5\log(T), 5\log^2(T)\}$. The numbers inside the parentheses are the standard errors of the simulated averages.

Model	T	$\lambda_T = 5$			$\lambda_T = 5\log(T)$			$\lambda_T = 5\log^2(T)$		
		$\widehat{K}$	$D(\mathcal{R}, \widehat{\mathcal{R}})$	$D(\mathcal{B}, \widehat{\mathcal{B}})$	$\widehat{K}$	$D(\mathcal{R}, \widehat{\mathcal{R}})$	$D(\mathcal{B}, \widehat{\mathcal{B}})$	$\widehat{K}$	$D(\mathcal{R}, \widehat{\mathcal{R}})$	$D(\mathcal{B}, \widehat{\mathcal{B}})$
Model (2.1) ($K_0 = 4$)	200	4.00 (0.00)	0.03 (0.02)	0.61 (0.12)	3.99 (0.08)	0.03 (0.04)	0.62 (0.16)	2.78 (0.87)	0.87 (0.91)	2.24 (1.05)
	400	4.00 (0.00)	0.01 (0.01)	0.41 (0.08)	4.00 (0.00)	0.01 (0.01)	0.41 (0.08)	3.92 (0.27)	0.05 (0.13)	0.53 (0.43)
	800	4.00 (0.00)	0.01 (0.00)	0.29 (0.05)	4.00 (0.00)	0.01 (0.00)	0.29 (0.05)	4.00 (0.00)	0.01 (0.00)	0.29 (0.05)
	1600	4.00 (0.00)	0.00 (0.00)	0.20 (0.04)	4.00 (0.00)	0.00 (0.00)	0.20 (0.04)	4.00 (0.00)	0.00 (0.00)	0.20 (0.04)
Model (6.1) ($K_0 = 3$)	200	3.44 (0.50)	0.12 (0.11)	0.50 (0.11)	3.00 (0.00)	0.02 (0.02)	0.48 (0.11)	2.85 (0.38)	0.13 (0.30)	0.75 (0.69)
	400	3.39 (0.49)	0.10 (0.11)	0.34 (0.07)	3.00 (0.00)	0.01 (0.01)	0.33 (0.07)	3.00 (0.00)	0.01 (0.01)	0.33 (0.07)
	800	3.33 (0.47)	0.08 (0.11)	0.23 (0.05)	3.00 (0.00)	0.01 (0.00)	0.22 (0.05)	3.00 (0.00)	0.01 (0.00)	0.22 (0.05)
	1600	3.33 (0.47)	0.08 (0.11)	0.16 (0.03)	3.00 (0.00)	0.00 (0.00)	0.16 (0.03)	3.00 (0.00)	0.00 (0.00)	0.16 (0.03)
Model (6.2) ($K_0 = 3$)	200	3.00 (0.00)	0.02 (0.01)	0.47 (0.12)	3.00 (0.00)	0.02 (0.01)	0.47 (0.12)	2.71 (0.46)	0.19 (0.27)	0.97 (0.80)
	400	3.00 (0.00)	0.01 (0.01)	0.31 (0.07)	3.00 (0.00)	0.01 (0.01)	0.31 (0.07)	3.00 (0.00)	0.01 (0.01)	0.31 (0.07)
	800	3.00 (0.00)	0.01 (0.00)	0.22 (0.05)	3.00 (0.00)	0.01 (0.00)	0.22 (0.05)	3.00 (0.00)	0.01 (0.00)	0.22 (0.05)
	1600	3.00 (0.00)	0.00 (0.00)	0.15 (0.03)	3.00 (0.00)	0.00 (0.00)	0.15 (0.03)	3.00 (0.00)	0.00 (0.00)	0.15 (0.03)
Model (6.3) ($K_0 = 2$)	200	3.38 (0.59)	0.14 (0.11)	0.35 (0.10)	2.03 (0.17)	0.01 (0.01)	0.30 (0.08)	2.00 (0.00)	0.01 (0.01)	0.30 (0.08)
	400	3.54 (0.51)	0.13 (0.11)	0.24 (0.07)	2.01 (0.08)	0.01 (0.01)	0.20 (0.05)	2.00 (0.00)	0.01 (0.00)	0.20 (0.05)
	800	3.53 (0.53)	0.12 (0.11)	0.16 (0.04)	2.00 (0.06)	0.00 (0.00)	0.14 (0.04)	2.00 (0.00)	0.00 (0.00)	0.14 (0.04)
	1600	3.50 (0.55)	0.13 (0.12)	0.12 (0.03)	2.00 (0.00)	0.00 (0.00)	0.10 (0.03)	2.00 (0.00)	0.00 (0.00)	0.10 (0.03)
Model (6.4) ($K_0 = 2$)	200	2.93 (0.80)	0.23 (0.18)	0.37 (0.10)	2.00 (0.04)	0.02 (0.02)	0.34 (0.11)	2.00 (0.00)	0.02 (0.01)	0.34 (0.11)
	400	2.80 (0.76)	0.20 (0.18)	0.25 (0.07)	2.00 (0.00)	0.01 (0.01)	0.24 (0.07)	2.00 (0.00)	0.01 (0.01)	0.24 (0.07)
	800	2.68 (0.70)	0.17 (0.17)	0.17 (0.05)	2.00 (0.00)	0.00 (0.00)	0.16 (0.05)	2.00 (0.00)	0.00 (0.00)	0.16 (0.05)
	1600	2.70 (0.69)	0.18 (0.17)	0.15 (0.14)	2.00 (0.06)	0.01 (0.03)	0.14 (0.14)	2.00 (0.00)	0.01 (0.03)	0.14 (0.14)
Model (6.5) ($K_0 = 1$)	200	1.98 (0.70)	0.28 (0.19)	0.17 (0.07)	1.04 (0.01)	0.02 (0.00)	0.13 (0.05)	1.00 (0.00)	0.00 (0.00)	0.13 (0.05)
	400	1.93 (0.68)	0.27 (0.19)	0.12 (0.04)	1.02 (0.00)	0.01 (0.00)	0.09 (0.03)	1.00 (0.00)	0.00 (0.00)	0.09 (0.03)
	800	1.84 (0.59)	0.25 (0.18)	0.08 (0.03)	1.00 (0.00)	0.00 (0.00)	0.07 (0.02)	1.00 (0.00)	0.00 (0.00)	0.07 (0.02)
	1600	1.85 (0.64)	0.25 (0.19)	0.06 (0.02)	1.00 (0.00)	0.00 (0.00)	0.05 (0.02)	1.00 (0.00)	0.00 (0.00)	0.05 (0.02)

H.2. Multiple solutions selected by the MIQP. In our estimation procedure, it is required to produce multiple solutions for $\boldsymbol{\gamma}$ and then take their averages to approximate the centroid of the least squares set $\widehat{\mathcal{G}}$. In this part, we demonstrate the performance of such an approximation by the following simulation.

The data generation process for $\{(Y_i, \boldsymbol{X}_i, \boldsymbol{Z}_i)\}_{i=1}^T$ was the same as the independence setting as that in Section 7.1. The sample size used in this simulation was $T = 800$. The true splitting coefficients were $\boldsymbol{\gamma}_{10} = (1, 1, 0)^\mathrm{T}$ and $\gamma_{20} = (1, -1, 0)^\mathrm{T}$, respectively. By setting the parameters *SolutionNumber* = 200 and *PoolGap* = 0 in the MIQP solver in GUROBI, we obtained 200 solutions whose objective values all attained the minimum, which ensured that these solutions were selected from the $\widehat{\mathcal{G}}$. Figure S3 displays that the selected values were nearly uniformly distributed, and their averages approximated to the true values colored in red and the center of $\widehat{\mathcal{G}}$.

Fig S3: Distributions of the selected 200 optimal solutions for the splitting coefficients. The first elements of γ_1 and γ_2 were omitted since they were normalized as 1. The true values were indicated in red, and the averages of the multiple solutions were indicated in green.



APPENDIX I: ADDITIONAL CASE STUDY RESULTS

The following Table S4 reports basic summary statistics of the involved variables in the training and testing sets.

TABLE S4

Sample means of the training and testing sets of the meteorological variables. The numbers inside the parentheses are the sample standard deviations and the numbers inside the square brackets are the sample correlations with the covariates and $PM_{2.5}$.

Season	$PM_{2.5}$	TEMP	DEWP	PRES	WD	IWS	log(BLH)	RAIN
Training sets								
	48.31	15.87	-1.47	1008.09	3.51	5.89	5.39	0.09
Spring	(49.00)	(8.05)	(9.40)	(6.78)	(1.28)	(8.95)	(1.79)	(0.84)
		[-0.21]	[0.22]	[-0.10]	[0.06]	[-0.20]	[-0.25]	[-0.06]
	38.70	27.34	16.46	998.76	3.40	4.27	5.31	0.46
Summer	(27.08)	(4.41)	(5.50)	(4.34)	(1.33)	(7.53)	(1.62)	(2.70)
		[0.01]	[0.52]	[-0.12]	[0.02]	[-0.19]	[-0.10]	[-0.01]
	49.93	15.42	5.82	1013.51	3.14	3.99	4.87	0.15
Fall	(36.19)	(9.34)	(9.81)	(8.00)	(1.43)	(7.43)	(1.53)	[1.99]
		[0.04]	[0.26]	[-0.25]	[-0.07]	[-0.19]	[-0.12]	[-0.07]
	58.77	0.07	-14.62	1021.05	3.26	4.89	4.58	0.00
Winter	(56.65)	(5.03)	(7.07)	(6.68)	(1.29)	(8.46)	(1.56)	(0.01)
		[-0.03]	[0.57]	[-0.40]	[0.11]	[-0.27]	[-0.33]	[-0.01]
Testing sets								
	54.99	16.87	0.82	1005.28	3.64	11.59	5.46	0.09
Spring	(42.81)	(6.33)	(10.29)	(6.09)	(1.24)	(20.94)	(1.74)	(0.84)
		[-0.28]	[0.54]	[-0.57]	[-0.01]	[-0.33]	[-0.04]	[-0.02]
	41.61	26.96	17.16	997.93	3.31	4.95	5.41	0.32
Summer	(29.42)	(3.95)	(4.16)	(3.33)	(1.28)	(8.08)	(1.50)	(1.21)
		[0.06]	[0.57]	[0.30]	[0.07]	[-0.29]	[-0.02]	[-0.01]
	37.77	13.75	4.25	1014.92	3.25	4.50	4.89	0.17
Fall	(32.64)	(7.74)	(10.71)	(6.06)	(1.40)	(11.34)	(1.47)	(1.79)
		[-0.09]	[0.16]	[-0.12]	[-0.16]	[-0.15]	[-0.11]	[-0.02]
	56.48	-0.31	-14.61	1021.99	3.18	5.09	4.67	0.01
Winter	(83.69)	(4.14)	(6.51)	(4.86)	(1.25)	(7.75)	(1.56)	(0.08)
		[-0.15]	[0.35]	[-0.36]	[0.05]	[-0.22]	[-0.32]	[-0.02]

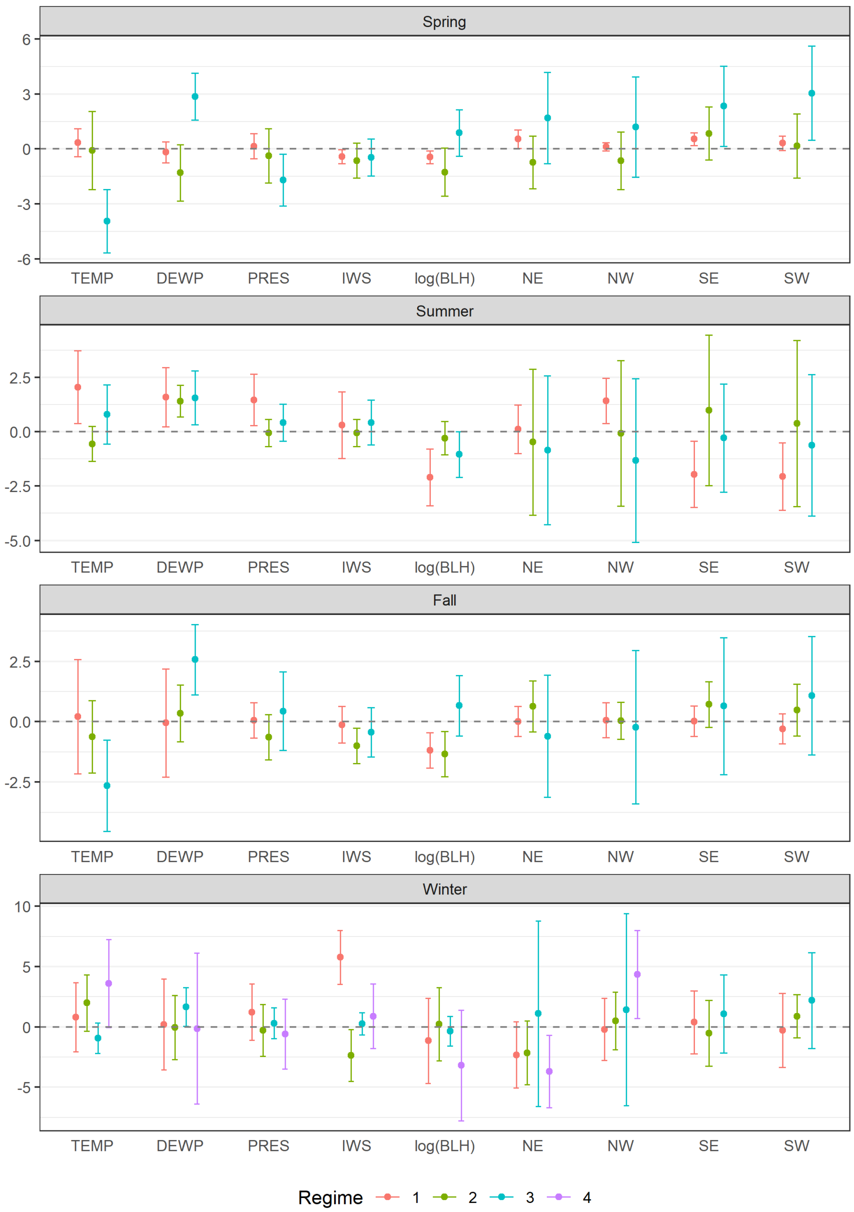
Table S5 reports some important statistics of each estimated regimes for the four seasons, including thes ample sizes, the fitting RMSEs and the sample means of $PM_{2.5}$ and the regression covariates of the estimated regimes.

TABLE S5

The sample sizes, the fitting RMSEs and the sample means of $PM_{2.5}$ and the regression covariates in the four seasons. The numbers inside the parentheses are the sample standard deviations of the sample means above them. RAIN was not included in seasons except for summer since their precipitation was rather sparse.

		T	RMSE	PM2.5	TEMP	DEWP	PRES	log(BLH)	RAIN
Spring	1	119	10.4	67.7	15.5	10.9	1006.7	5.6	
				(40.0)	(5.5)	(4.1)	(5.9)	(1.3)	
	2	793	12.7	61.5	15.9	2.6	1006.6	5.0	
				(55.7)	(8.8)	(7.0)	(6.6)	(1.7)	
	3	528	10.6	23.7	15.9	-10.5	1010.6	5.9	
				(23.0)	(7.4)	(4.7)	(6.5)	(1.8)	
Summer	1	180	9.4	61.9	28.6	23.6	995.9	4.2	0.12
				(37.3)	(3.4)	(1.1)	(3.2)	(5.6)	(0.6)
	2	910	8.4	42.6	27.1	17.9	998.6	2.7	0.2
				(22.8)	(4.4)	(2.9)	(4.2)	(3.2)	(1.2)
	3	343	4.7	16.1	27.1	8.8	1000.6	8.3	8.6
				(10.6)	(4.6)	(3.4)	(4.3)	(13.0)	(10.1)
Fall	1	252	9.2	61.1	15.7	13.3	1011.3	3.8	
				(36.9)	(5.1)	(4.2)	(5.1)	(0.9)	
	2	738	8.9	53.7	18.5	9.4	1011.4	5.0	
				(36.2)	(9.1)	(6.2)	(6.4)	(1.5)	
	3	448	9.1	37.4	10.2	-4.4	1018.3	5.3	
				(32.1)	(9.4)	(8.8)	(9.5)	(1.6)	
Winter	1	288	16.3	94.4	3.0	-9.7	1018.0	5.0	
				(62.1)	(5.2)	(6.8)	(5.8)	(1.7)	
	2	194	11.2	54.9	1.5	-16.2	1021.1	5.2	
				(43.3)	(6.0)	(5.7)	(7.4)	(1.7)	
	3	760	11.7	34.5	-2.0	-21.6	1025.9	4.8	
				(45.0)	(4.9)	(6.6)	(7.0)	(1.6)	
	4	157	15.8	71.6	-3.6	-16.2	1022.2	3.5	
				(55.3)	(4.4)	(5.3)	(7.3)	(1.0)	

Fig S4: Estimated regression coefficients (indicated by dots) and their 95% confidence intervals (indicated by bars) of each regime. The estimated coefficients of the Lag term were all significantly above 0 and thus not reported in this figure.

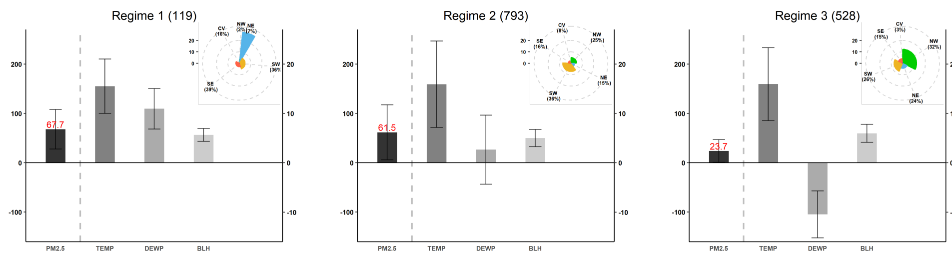


The following Figure S5 displays the estimated meteorological regimes on $PM_{2.5}$. It shows that in spring, for instance, Regime 1 had the highest DEWP and the highest proportion of CV among the three regimes, which is a known condition to encourage secondary generation of $PM_{2.5}$ and to constitute a unfavourable atmospheric diffusion condition, and thus resulted in high $PM_{2.5}$. Regime 2 had reduced percentages of CV and lower DEWP level compared to Regime 1, which alleviated the polluting level and led to better diffusion of $PM_{2.5}$ and can be regarded as a transitional state from either the high pollution to low pollution or vice versa. In Regime 3, the northerly winds occupied the leading position and DEWP was significantly lower. It is noted that the northerly wind brings cleaner and cooler air from the north, and under such circumstances the $PM_{2.5}$ concentration could be effectively reduced via the removal process at a lack of secondary generation. Therefore, Regime 3 represented a cleaning regime.

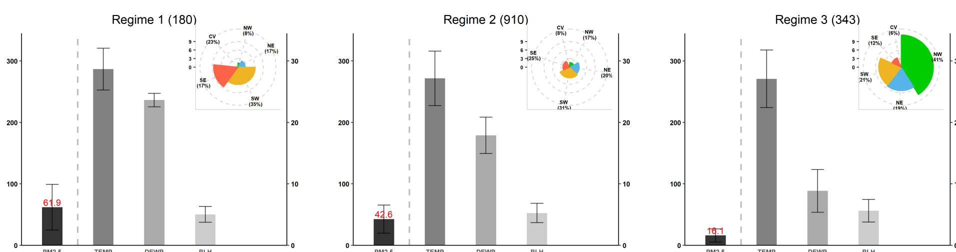
It is found that the regime-splitting for summer and fall shared the same pattern with the spring, namely Regime 1 with high $PM_{2.5}$ level accompanied by a large proportion of CV and high DEWP, indicating an air stagnation; Regime 2 is a transitional regime which had reduced DEWP and increased winds with about 50% southerly winds; and Regime 3 (cleaning) tended to had significantly large amount of strong northerly, in particular northwesterly wind and low DEWP, which are known favorable conditions to lower the $PM_{2.5}$. For winter, Regime 1 was still the most polluting regime and Regime 3 represented the cleaning regimes as the other seasons. However, the transitional regime was divided to two regimes: Regimes 2 and 4 with dominated wind directions being southeasterly and southwesterly wind, respectively, representing two different transitional modes. Regime 4 had more southwesterly wind which would bring the accumulated $PM_{2.5}$ along the foot of Taihang Mountain to Beijing, bringing in more transported air pollutants. As validated in Figure S5, Regime 4 of winter indeed had heavier $PM_{2.5}$ than Regime 2.

Fig S5: For each regime, the bars indicate sample means of $PM_{2.5}$ (scales on the left side), TEMP, DEWP and $\log$(BLH) (scales on the left side) and the lengths of error bar are twice of the sample deviations. The wind rose plots displays distribution of wind directions (via width of angles) and average speed (via length of radius). Sample sizes of each regime are reported in the parentheses of its subtitle.

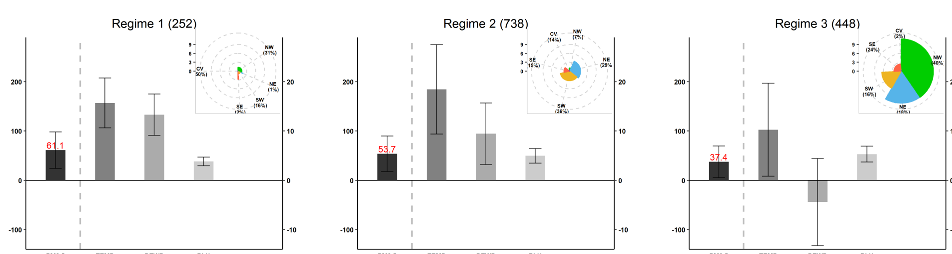
(a) Spring



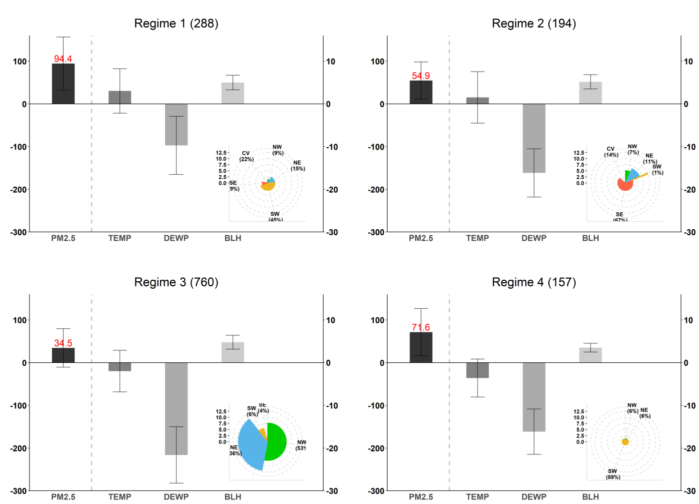
(b) Summer



(c) Fall



(d) Winter



REFERENCES

Bickel, P. J. and M. J. Wichura (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics 42*(5), 1656–1670.

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Buck, R. C. (1943). Partition of space. *The American Mathematical Monthly 50*(9), 541–544.

Chernozhukov, V. and H. Hong (2004). Likelihood estimation and inference in a class of nonregular econometric models. MIT Department of Economics Working Paper.

Doob, J. L. (1953). *Stochastic Processes*. London;New York;: Wiley.

Doukhan, P. (1995). *Mixing: properties and examples*, Volume 85. Springer Science & Business Media.

Györfi, L., W. Härdle, P. Sarda, and P. Vieu (1989). *Nonparametric curve estimation from time series*, Volume 60 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic press.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*(3), 726–748.

Hsing, T. (1995). On the asymptotic independence of the sum and rare values of weakly dependent stationary random variables. *Stochastic Process. Appl. 60*(1), 49–63.

Knight, K. (1999). Epi-convergence and stochastic equisemicontinuity. Preprint.

Lan, Y., M. Banerjee, and G. Michailidis (2009). Change-point estimation under adaptive sampling. *The Annals of Statistics 37*(4), 1752 – 1791.

Lee, S., Y. Liao, M. H. Seo, and Y. Shin (2021). Factor-driven two-regime regression. *Ann. Statist. 49*(3), 1656–1678.

Meyer, R. M. (1973). A Poisson-type limit theorem for mixing sequences of dependent "rare" events. *Ann. Probability 1*, 480–483.

Orlik, P. and H. Terao (2013). *Arrangements of Hyperplanes*, Volume 300. Springer Science & Business Media.

Peligrad, M. (1982). Invariance principles for mixing sequences of random variables. *The Annals of Probability*, 968–981.

Resnick, S. I. (2008). *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Rockafellar, R. T. and R. J. Wets (1998). *Variational analysis*. Springer-Verlag, Berlin.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer Science & Business Media.

Wald, A. (1944). On cumulative sums of random variables. *The Annals of Mathematical Statistics 15*(3), 283–296.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *Journal of Econometrics 167*(1), 274–294.

Yu, P. and X. Fan (2021). Threshold regression with a threshold boundary. *Journal of Business & Economic Statistics 39*(4), 953–971.