Penalized estimation for non-identifiable models

Junichiro Yoshida · Nakahiro Yoshida

Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.

Supplementary material

7 Appendix: proof of Theorems 4 and 5, and Proposition 1

We first prepare some lemmas.

Lemma 2 *Let $\{N_t\}_{t\geq 0}$ be a counting process whose intensity is denoted by λ_t. Let $\{X_t\}_{t\geq 0}$ be an $\mathbb{R}^{\mathsf{d}}$-valued predictable process assumed to be locally bounded. Let $G\subset\mathbb{R}^{\mathsf{g}}$ be a bounded open domain admitting the Sobolev embedding. Let $f:\mathbb{R}^{\mathsf{d}}\times G\to\mathbb{R}^{\mathsf{f}}$ be a measurable map satisfying the following conditions.*
(i) For each $x\in\mathbb{R}^{\mathsf{d}}$, $f(x,\cdot)$ is of class $C^1(G)$.
(ii) $\sup_{\gamma\in G}\left|\partial_\gamma^i f(\cdot,\gamma)\right|$ $(i=0,1)$ are bounded on every bounded set of $\mathbb{R}^{\mathsf{d}}$.
Then for any $p\geq 1$,

$$\sup_{T>1} E\left[\sup_{\gamma\in G}\left|\frac{1}{\sqrt{T}}\int_0^T f(X_t,\gamma)d\tilde{N}_t\right|^p\right]<\infty,$$

where $\tilde{N}_t=N_t-\int_0^t\lambda_s ds$, provided that

$$\sup_{\gamma\in G,t\geq 0} E\left[\left|\partial_\gamma^i f(X_t,\gamma)\right|^p\lambda_t^q\right]<\infty \qquad (i=0,1)$$

for any $p,q\geq 1$ with $p\geq 2q$.

Proof Let $T>1$. Since the process $\sup_{\gamma\in G}\left|\partial_\gamma f(X,\gamma)\right|$ is locally bounded, we have

$$\partial_\gamma\int_0^T f(X_t,\gamma)d\tilde{N}_t \;=\; \int_0^T\partial_\gamma f(X_t,\gamma)d\tilde{N}_t \qquad (\gamma\in G).$$

Let p be any positive number with $p>\mathsf{g}$. Take some integer k with $2^k\geq p$. From Sobolev's inequality,

$$\begin{aligned}
E\left[\sup_{\gamma\in G}\left|\frac{1}{\sqrt{T}}\int_0^T f(X_t,\gamma)d\tilde{N}_t\right|^p\right] &\leq E\left[\sup_{\gamma\in G}\left|\frac{1}{\sqrt{T}}\int_0^T f(X_t,\gamma)d\tilde{N}_t\right|^{2^k}\right]\\
\lesssim E\left[\int_G\sum_{i=0,1}\left|\frac{1}{\sqrt{T}}\int_0^T\partial_\gamma^i f(X_t,\gamma)d\tilde{N}_t\right|^{2^k}d\gamma\right]&\\
\lesssim \sup_{\gamma\in G,i=0,1}E\left[\left|\frac{1}{\sqrt{T}}\int_0^T\partial_\gamma^i f(X_t,\gamma)d\tilde{N}_t\right|^{2^k}\right].& \qquad (56)
\end{aligned}$$

By the Burkholder-Davis-Gundy inequality, for each $i=0,1$,

$$
\begin{aligned}
&E\Bigg[\bigg|\frac{1}{\sqrt{T}}\int_0^T \partial_\gamma^i f(X_t,\gamma)d\tilde{N}_t\bigg|^{2^k}\Bigg]\\
\lesssim\ &E\Bigg[\bigg|\frac{1}{T}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^2 dN_t\bigg|^{2^{k-1}}\Bigg]\\
\lesssim\ &E\Bigg[\bigg|\frac{1}{T}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^2 \lambda_t dt\bigg|^{2^{k-1}}\Bigg]+E\Bigg[\bigg|\frac{1}{T}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^2 d\tilde{N}_t\bigg|^{2^{k-1}}\Bigg]\\
\leq\ &\sup_{t\geq 0}E\Bigg[\bigg|\big|\partial_\gamma^i f(X_t,\gamma)\big|^2\lambda_t\bigg|^{2^{k-1}}\Bigg]+E\Bigg[\bigg|\frac{1}{\sqrt{T}}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^2 d\tilde{N}_t\bigg|^{2^{k-1}}\Bigg].
\end{aligned}
$$

Repeating this evaluation,

$$
\begin{aligned}
&E\Bigg[\bigg|\frac{1}{\sqrt{T}}\int_0^T \partial_\gamma^i f(X_t,\gamma)d\tilde{N}_t\bigg|^{2^k}\Bigg]\\
\lesssim\ &\sum_{j=1}^{k-1}\sup_{t\geq 0}E\Bigg[\bigg|\big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^j}\lambda_t\bigg|^{2^{k-j}}\Bigg]+E\Bigg[\bigg|\frac{1}{\sqrt{T}}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^{k-1}} d\tilde{N}_t\bigg|^{2}\Bigg]\\
\lesssim\ &\sum_{j=1}^{k-1}\sup_{t\geq 0}E\Bigg[\bigg|\big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^j}\lambda_t\bigg|^{2^{k-j}}\Bigg]+E\Bigg[\frac{1}{T}\int_0^T \big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^k}\lambda_t dt\Bigg]\\
\leq\ &\sum_{j=1}^{k}\sup_{t\geq 0}E\Bigg[\bigg|\big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^j}\lambda_t\bigg|^{2^{k-j}}\Bigg].
\end{aligned}
$$

Thus, from (56),

$$
\begin{aligned}
&\sup_{T>1}E\Bigg[\sup_{\gamma\in G}\bigg|\frac{1}{\sqrt{T}}\int_0^T f(X_t,\gamma)d\tilde{N}_t\bigg|^{p}\Bigg]\\
\lesssim\ &\sup_{\gamma\in G,i=0,1}\sum_{j=1}^{k}\sup_{t\geq 0}E\Bigg[\bigg|\big|\partial_\gamma^i f(X_t,\gamma)\big|^{2^j}\lambda_t\bigg|^{2^{k-j}}\Bigg]<\infty.
\end{aligned}
$$

□

Lemma 3 *Take N, λ, X, G and f as in Lemma 2, and assume (i) and (ii) in Lemma 2. Also assume the ergodicity of X as (23). Then*

$$
\sup_{\gamma\in G}\bigg|\frac{1}{T}\int_0^T f(X_t,\gamma)dt-\int_{\mathbb{R}^d} f(x,\gamma)\nu(dx)\bigg|\to 0 \quad \textit{in } L^1(dP) \quad (T\to\infty) \tag{57}
$$

provided that

$$
\Big\{\sup_{\gamma\in G}|f(X_t,\gamma)|\Big\}_{t\geq 0} \quad \textit{is uniformly integrable.} \tag{58}
$$

In particular, (57) *holds if for any* $p \geq 1$,

$$\sup_{\gamma \in G, t \geq 0} E\left[\left|\partial_\gamma^i f(X_t, \gamma)\right|^p\right] < \infty \qquad (i = 0, 1). \tag{59}$$

Proof Let $M > 0$, and define a bounded function f_M as $f_M(x, \gamma) = \big(f(x, \gamma) \wedge M\big) \vee (-M)$. Then

$$\begin{aligned}
& E\left[\sup_{\gamma \in G}\left|\frac{1}{T}\int_0^T f(X_t, \gamma)dt - \int_{\mathbb{R}^d} f(x, \gamma)\nu(dx)\right|\right] \\
\leq\ & E\left[\sup_{\gamma \in G}\left|\frac{1}{T}\int_0^T f(X_t, \gamma)dt - \frac{1}{T}\int_0^T f_M(X_t, \gamma)dt\right|\right] \\
& + \sup_{\gamma \in G}\left|\int_{\mathbb{R}^d} f(x, \gamma)\nu(dx) - \int_{\mathbb{R}^d} f_M(x, \gamma)\nu(dx)\right| \\
& + E\left[\sup_{\gamma \in G}\left|\frac{1}{T}\int_0^T f_M(X_t, \gamma)dt - \int_{\mathbb{R}^d} f_M(x, \gamma)\nu(dx)\right|\right].
\end{aligned}$$

The first and second terms on the rightmost side are as small as we want by taking sufficiently large $M > 0$ since

$$\begin{aligned}
& \sup_{\gamma \in G}\left|\int_{\mathbb{R}^d} f(x, \gamma) 1_{\left\{|f(x,\gamma)| \geq M\right\}}\nu(dx)\right| \\
\leq\ & \int_{\mathbb{R}^d} \sup_{\gamma \in G}|f(x, \gamma)|\, 1_{\left\{\sup_{\gamma \in G}|f(x,\gamma)| \geq M\right\}}\nu(dx) \\
=\ & \lim_{L \to \infty}\lim_{T \to \infty} E\left[\frac{1}{T}\int_0^T L \wedge \sup_{\gamma \in G}|f(X_t, \gamma)|\, 1_{\left\{\sup_{\gamma \in G}|f(X_t,\gamma)| \geq M\right\}} dt\right] \\
\leq\ & \sup_{t \geq 0} E\left[\sup_{\gamma \in G}|f(X_t, \gamma)|\, 1_{\left\{\sup_{\gamma \in G}|f(X_t,\gamma)| \geq M\right\}}\right],
\end{aligned}$$

and since (58) holds. Let $\delta > 0$ and take a finite set $G_\delta \subset G$ such that $\sup_{\gamma_1 \in G}\min_{\gamma_2 \in G_\delta}|\gamma_1 - \gamma_2| < \delta$. Then

$$\begin{aligned}
& E\left[\sup_{\gamma \in G}\left|\frac{1}{T}\int_0^T f_M(X_t, \gamma)dt - \int_{\mathbb{R}^d} f_M(x, \gamma)\nu(dx)\right|\right] \\
\leq\ & E\left[\max_{\gamma \in G_\delta}\left|\frac{1}{T}\int_0^T f_M(X_t, \gamma)dt - \int_{\mathbb{R}^d} f_M(x, \gamma)\nu(dx)\right|\right] \\
& + E\left[\frac{1}{T}\int_0^T \sup_{\substack{\gamma_1, \gamma_2 \in G \\ |\gamma_1 - \gamma_2| < \delta}}\big|f_M(X_t, \gamma_1) - f_M(X_t, \gamma_2)\big| dt\right] \\
& + \int_{x \in \mathbb{R}^d} \sup_{\substack{\gamma_1, \gamma_2 \in G \\ |\gamma_1 - \gamma_2| < \delta}}\big|f_M(x, \gamma_1) - f_M(x, \gamma_2)\big|\nu(dx) \\
\overset{T \to \infty}{\to}\ & 2\int_{x \in \mathbb{R}^d} \sup_{\substack{\gamma_1, \gamma_2 \in G \\ |\gamma_1 - \gamma_2| < \delta}}\big|f_M(x, \gamma_1) - f_M(x, \gamma_2)\big|\nu(dx) \overset{\delta \to 0}{\to} 0.
\end{aligned}$$

Thus, (57) holds.

From Sobolev's inequality, (58) holds if (59) holds. □

Proof of Theorem 4. We define a new parameter space $\Xi = \Theta \times \mathcal{T}$ by

$$\begin{aligned}\Theta &= [0, M_g] \times [-L_\beta, M_\beta]^{|\mathcal{A}|} \times [0, M_\alpha]^{|\mathcal{A}|} \times [0, M_\alpha]^{\mathsf{a}-|\mathcal{A}|} \\ \mathcal{T} &= [-L_\beta, M_\beta]^{\mathsf{a}-|\mathcal{A}|},\end{aligned}$$

and also define new parameters $\theta = (\overline{\theta}, \underline{\theta}) \in \Theta$ and $\tau = (\tau_k)_{k\in\mathcal{A}^c} \in \mathcal{T}$ by

$$\overline{\theta} = \big(g,\, (\beta_i)_{i\in\mathcal{A}}\,,\, (\alpha_i)_{i\in\mathcal{A}}\big), \quad \underline{\theta} = (\alpha_k)_{k\in\mathcal{A}^c}, \quad \tau = (\tau_k)_{k\in\mathcal{A}^c} = (\beta_k)_{k\in\mathcal{A}^c}.$$

That is, we consider a parameter transformation as $(\theta, \tau) = \varphi(g, \alpha, \beta)$, where $\varphi : [0, M_g] \times [0, M_\alpha]^{\mathsf{a}} \times [-L_\beta, M_\beta]^{\mathsf{a}} \to \Xi$ is defined as

$$\varphi(g, \alpha, \beta) = \big(g,\, (\beta_i)_{i\in\mathcal{A}}\,,\, (\alpha_i)_{i\in\mathcal{A}}\,,\, (\alpha_k)_{k\in\mathcal{A}^c}\,,\, (\beta_k)_{k\in\mathcal{A}^c}\big). \tag{60}$$

Define estimators $\hat{\theta}_T$ and $\hat{\tau}_T$ taking values in Θ and $\mathcal{T}$, respectively, by $(\hat{\theta}_T, \hat{\tau}_T) = \varphi(\hat{g}_T, \hat{\alpha}_T, \hat{\beta}_T)$. We set $\mathsf{p} = 1+|\mathcal{A}|+\mathsf{a}$ and $\mathsf{p}_1 = 1+2|\mathcal{A}|$, respectively. We define $\mathcal{J}_1$, $\mathcal{J}_0$ and $\mathcal{J}$ as $\mathcal{J}_1 = \{1\} \cup \big\{2 + |\mathcal{A}|, ..., \mathsf{p}_1\big\}$, $\mathcal{J}_0 = \{\mathsf{p}_1 + 1, ..., \mathsf{p}\}$ and $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_0$, respectively. Define one of the true values $\theta^* = (\overline{\theta}^*, \underline{\theta}^*) \in \mathbb{R}^{\mathsf{p}_1} \times \mathbb{R}^{\mathsf{p}-\mathsf{p}_1}$ as

$$\overline{\theta}^* = \big(g^*,\, (\beta_i^*)_{i\in\mathcal{A}}\,,\, (\alpha_i^*)_{i\in\mathcal{A}}\big)\,, \qquad \underline{\theta}^* = (0)_{k\in\mathcal{A}^c}.$$

For $\mathsf{r} := \mathsf{p}$, let $\mathfrak{a}_T = \mathrm{diag}(\mathfrak{a}_{1,T}, ..., \mathfrak{a}_{\mathsf{r},T}) = T^{-\frac{1}{2}} I_{\mathsf{r}}$,[1] $\mathfrak{b}_T = T$ and $\rho_k = \frac{r}{q} > 1$ $(k \in \mathcal{J}_0)$. We take $a_T \in GL(\mathsf{p})$ as a deterministic diagonal matrix defined by

$$(a_T)_{jj} = \begin{cases} T^{-\frac{1}{2}} & (j \in \{1, ..., \mathsf{p}_1\}) \\ T^{-\frac{r}{2q}} & (j \in \{\mathsf{p}_1 + 1, ..., \mathsf{p}\}) \end{cases}.$$

Define U_T and U by (8) and (10), respectively. Then from Example 1 of Section 2.3 in Yoshida and Yoshida (2022), Condition [**A3**] holds, and $U = \mathbb{R} \times \mathbb{R}^{|\mathcal{A}|} \times \mathbb{R}^{|\mathcal{A}|} \times [0, \infty)^{\mathsf{a}-|\mathcal{A}|} \subset \mathbb{R}^{\mathsf{p}}$. Define c_i $(i \in \mathcal{J}_1)$ and d_k $(k \in \mathcal{J}_0)$ as

$$c_i = \begin{cases} \kappa_g 1_{\{r=1\}} & (i = 1) \\ \kappa_\alpha 1_{\{r=1\}} & \textit{otherwise} \end{cases}, \qquad d_k = \kappa_\alpha.$$

Define a random field $\mathbb{Z}$ as

$$\mathbb{Z}(u) = \overline{\Delta}[\overline{u}] - \frac{1}{2}\overline{\Gamma}[\overline{u}^{\otimes 2}] - q \sum_{i\in\mathcal{J}_1} c_i (\theta_i^*)^{q-1} u_i - \sum_{k\in\mathcal{J}_0} d_k |u_k|^q,$$

for any $u = (u_1, ..., u_{\mathsf{p}}) \in \mathbb{R}^{\mathsf{p}}$, where $\overline{u} = (u_1, ..., u_{\mathsf{p}_1})$ and $\overline{\Delta} \sim N_{\mathsf{p}_1}\big(0, \overline{\Gamma}\big)$. We define a U-valued random variable $\hat{u}$ by $\hat{u} = \big(\overline{\Gamma}^{-1}\overline{\Delta}^\dagger,\ 0\big)$, where $0 \in \mathbb{R}^{\mathsf{p}-\mathsf{p}_1}$.

[1] I_{m} denotes the m-dimensional identity matrix

Note that $\hat{u}$ becomes a unique maximizer of $\mathbb{Z}$ on U, and [**A4**] holds. Define a random field $\mathcal{H}_T : \Omega \times \Xi \to \mathbb{R} \cup \{-\infty\}$ as

$$\mathcal{H}_T(\theta,\tau) = \int_0^T \log \lambda_t\big(\phi(\theta,\tau)\big) dN_t - \int_0^T \lambda_t\big(\phi(\theta,\tau)\big) dt,$$

where ϕ denotes φ^{-1} for φ defined in (60). Then $\hat{\theta}_T$ is a maximizer of $\mathcal{H}_T(\theta, \hat{\tau}_T) - \sum_{j\in\mathcal{J}} \xi_{j,T} p_j(\theta_j)$ on Θ, where for any $j \in \mathcal{J}$,

$$\xi_{j,T} \ = \begin{cases} \kappa_g T^{\frac{r}{2}} & (j=1) \\ \kappa_\alpha T^{\frac{r}{2}} & otherwise \end{cases}, \qquad p_j(x) \ = \ |x|^q \qquad (x \in \mathbb{R}). \tag{61}$$

Note that q_j $(j \in \mathcal{J})$ defined in (iii) of Section 3.1 are determined as $q_j = q < 1$. In the following, we show [**H1**]-[**H5**] under [**P1**]-[**P3**], and use Theorem 2.

We first show [**H1**]. From Taylor's series, for any $(\theta, \tau) \in \Xi$,

$$\begin{aligned}
&\mathcal{H}_T(\theta,\tau) - \mathcal{H}_T(\theta^*,\tau) \\
&\leq \int_0^T \log \frac{\lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t^*} dN_t - \int_0^T \big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\} dt \\
&= \int_0^T \frac{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*}{\lambda_t^*} dN_t - \int_0^1 (1-s) \int_0^T \frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{\big\{s\lambda_t\big(\phi(\theta,\tau)\big) + (1-s)\lambda_t^*\big\}^2} \\
&\qquad\qquad \cdot dN_t ds - \int_0^T \big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\} dt \\
&= \int_0^T \frac{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*}{\lambda_t^*} d\tilde{N}_t - \int_0^1 (1-s) \int_0^T \frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{\big\{s\lambda_t\big(\phi(\theta,\tau)\big) + (1-s)\lambda_t^*\big\}^2} \\
&\qquad\qquad \cdot dN_t ds,
\end{aligned}$$

where $\tilde{N}$ is a martingale defined by $\tilde{N}_t = N_t - \int_0^t \lambda_s^* ds$. Take an arbitrary $R > 0$. The integrand of the second term in the rightmost side is evaluated as for any $(\theta,\tau) \in \Xi$, $0 \leq s \leq 1$ and $0 \leq t \leq T$,

$$\begin{aligned}
\frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{\big\{s\lambda_t\big(\phi(\theta,\tau)\big) + (1-s)\lambda_t^*\big\}^2} \ &\geq \ \frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{M_R g^*} \varphi_R(X_t) \\
&\geq \ \frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{M_R \lambda_t^*} \varphi_R(X_t) \qquad (\because \lambda_t^* \geq g^*),
\end{aligned}$$

where $\varphi_R : \mathbb{R}^{\mathsf{a}} \to [0,1]$ is a continuous function vanishing outside of $[-R-1, R+1]^{\mathsf{a}}$ and satisfying $\varphi_R \equiv 1$ on $[-R,R]^{\mathsf{a}}$, and $M_R > 0$ is a constant depending on R. Therefore,

$$\begin{aligned}
&\mathcal{H}_T(\theta,\tau) - \mathcal{H}_T(\theta^*,\tau) \\
&\leq \int_0^T \frac{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*}{\lambda_t^*} d\tilde{N}_t - \frac{1}{2} \int_0^T \frac{\big\{\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^*\big\}^2}{M_R \lambda_t^*} \varphi_R(X_t) dN_t.
\end{aligned}$$

In the following, we denote $\phi(\theta,\tau)$ by (g,α,β). As (27), we have

$$\lambda_t\big(\phi(\theta,\tau)\big) - \lambda_t^* \;=\; w(\beta,X_t)[h(\theta,\tau)] \qquad (t\geq 0, (\theta,\tau)\in\Xi),$$

where $h:\mathbb{R}^{\mathsf{p}}\to\mathbb{R}^{\mathsf{p}}$ is a continuous function defined as for any $(\theta,\tau)\in\mathbb{R}^{\mathsf{p}}$,

$$\begin{aligned} h(\theta,\tau) \;=\; &\bigg(g+\sum_{k\in\mathcal{A}^c}\alpha_k - g^*,\; \big(\alpha_i(\beta_i-\beta_i^*)\big)_{i\in\mathcal{A}},\; (\alpha_i-\alpha_i^*)_{i\in\mathcal{A}}, \\ &\quad (\alpha_k\beta_k)_{k\in\mathcal{A}^c}\bigg). \end{aligned} \tag{62}$$

Then

$$\begin{aligned} &\mathcal{H}_T(\theta,\tau)-\mathcal{H}_T(\theta^*,\tau) \\ &\leq\; \frac{1}{\sqrt{T}}\int_0^T \frac{w(\beta,X_t)}{\lambda_t^*}d\tilde{N}_t\big[\sqrt{T}h(\theta,\tau)\big] - \frac{1}{2M_RT}\int_0^T\frac{w(\beta,X_t)^{\otimes 2}}{\lambda_t^*}\varphi_R(X_t) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \cdot dN_t\big[\big(\sqrt{T}h(\theta,\tau)\big)^{\otimes 2}\big] \\ &=\; K_T(\theta,\tau)\big[\sqrt{T}h(\theta,\tau)\big]-\frac{1}{2}\big\{G(\theta,\tau)+r_T(\theta,\tau)\big\}\big[\big(\sqrt{T}h(\theta,\tau)\big)^{\otimes 2}\big] \end{aligned}$$

for any $t\geq 0$ and $(\theta,\tau)\in\Xi$, where for any $(\theta,\tau)\in\Xi$,

$$\begin{aligned} K_T(\theta,\tau) \;&=\; \frac{1}{\sqrt{T}}\int_0^T\frac{w(\beta,X_t)}{\lambda_t^*}d\tilde{N}_t, \\ G(\theta,\tau) \;&=\; \frac{1}{M_R}\int_{\mathbb{R}^{\mathsf{a}}} w(\beta,x)^{\otimes 2}\varphi_R(x)\nu(dx), \\ r_T(\theta,\tau) \;&=\; \frac{1}{M_R}\bigg\{\frac{1}{T}\int_0^T\frac{w(\beta,X_t)^{\otimes 2}}{\lambda_t^*}\varphi_R(X_t)d\tilde{N}_t \\ &\qquad +\frac{1}{T}\int_0^T w(\beta,X_t)^{\otimes 2}\varphi_R(X_t)dt - \int_{\mathbb{R}^{\mathsf{a}}} w(\beta,x)^{\otimes 2}\varphi_R(x)\nu(dx)\bigg\}. \end{aligned}$$

Condition [**P1**] implies that for any $p,q\geq 1$ with $p\geq 2q$ and for each $n=0,1$,

$$\begin{aligned} &\sup_{\beta\in[-L_\beta,M_\beta]^{\mathsf{a}},t\geq 0} E\bigg[\bigg|\frac{\partial_\beta^n w(\beta,X_t)}{\lambda_t^*}\bigg|^p(\lambda_t^*)^q\bigg] \\ &\leq (g^*)^{q-p}\sup_{\beta\in[-L_\beta,M_\beta]^{\mathsf{a}},t\geq 0} E\Big[\big|\partial_\beta^n w(\beta,X_t)\big|^p\Big] \qquad \big(\because \lambda_t^*\geq g^*\big) \\ &\lesssim (g^*)^{q-p}\sum_{j=1}^{\mathsf{a}}\sup_{\beta\in[-L_\beta,M_\beta]^{\mathsf{a}},t\geq 0} E\bigg[\Big\{\big(1+(X_t^j)^2\big)e^{\beta_j X_t^j}\Big\}^p\bigg] < \infty. \end{aligned} \tag{63}$$

Here the second inequality follows from the evaluation

$$
\begin{aligned}
&\left|\partial_\beta^n w(\beta,x)\right|^p \\
&= \left|\partial_\beta^n \int_0^1 \Big(1,\ \big(x_i e^{\{s\beta_i+(1-s)\beta_i^*\}x_i}\big)_{i\in\mathcal{A}},\ \big(e^{\beta_i^* x_i}\big)_{i\in\mathcal{A}},\ \big(x_k e^{s\beta_k x_k}\big)_{k\in\mathcal{A}^c}\Big) ds\right|^p \\
&\le \int_0^1 \left|\partial_\beta^n \Big(1,\ \big(x_i e^{\{s\beta_i+(1-s)\beta_i^*\}x_i}\big)_{i\in\mathcal{A}},\ \big(e^{\beta_i^* x_i}\big)_{i\in\mathcal{A}},\ \big(x_k e^{s\beta_k x_k}\big)_{k\in\mathcal{A}^c}\Big)\right|^p ds \\
&\lesssim \int_0^1 \left[\sum_{j=1}^{\mathsf{a}} \big\{(1+x_j^2)e^{\tilde{\beta}_{j,s}x_j}\big\}^p + \sum_{i\in\mathcal{A}} \big\{e^{\beta_i^* x_i}\big\}^p\right] ds
\end{aligned}
$$

for any $x=(x_1,...,x_{\mathsf{a}})\in\mathbb{R}^{\mathsf{a}}$ and any $\beta\in[-L_\beta,M_\beta]^{\mathsf{a}}$, where $\tilde{\beta}_{j,s}\in[-L_\beta,M_\beta]^{\mathsf{a}}$ $(j=1,...,\mathsf{a})$ are real numbers depending on s. From (63), we can use Lemma 2, and obtain

$$
\sup_{(\theta,\tau)\in\Xi} \left|K_T(\theta,\tau)\right| = O_P(1). \tag{64}
$$

Similarly, $\sup_{\beta\in[-L_\beta,M_\beta]^{\mathsf{a}}} \left|\frac{1}{\sqrt{T}}\int_0^T \frac{w(\beta,X_t)^{\otimes 2}}{\lambda_t^*}\varphi_R(X_t)d\tilde{N}_t\right| = O_P(1)$. Since for any $p>1$,

$$
\sup_{\beta\in[-L_\beta,M_\beta]^{\mathsf{a}},t\ge 0} E\left[\left|\partial_\beta^i\big\{w(\beta,X_t)^{\otimes 2}\big\}\varphi_R(X_t)\right|^p\right] < \infty \qquad (i=0,1),
$$

we can use Lemma 3, and obtain

$$
\frac{1}{T}\int_0^T w(\beta,X_t)^{\otimes 2}\varphi_R(X_t)dt - \int_{\mathbb{R}^{\mathsf{a}}} w(\beta,x)^{\otimes 2}\varphi_R(x)\nu(dx) \xrightarrow{P} 0,
$$

uniformly in $\beta\in[-L_\beta,M_\beta]^{\mathsf{a}}$ as $T\to\infty$. Therefore, $\sup_{(\theta,\tau)\in\Xi}\left|r_T(\theta,\tau)\right| = o_P(1)$. Moreover, [**P2**] implies the non-degeneracy of G for sufficiently large R. Thus, [**H1**] holds for $\mathfrak{a}_T = T^{-\frac{1}{2}}I_{\mathsf{r}}$.

Condition [**H2**] also holds. Indeed, continuing to denote $\phi(\theta,\tau)$ by (g,α,β), we have

$$
\begin{aligned}
\Theta^* &= \{\theta\in\Theta;\ \exists\tau\in\mathcal{T}, h(\theta,\tau)=0\} \\
&= \Big\{\theta\in\Theta;\ g+\sum_{k\in\mathcal{A}^c}\alpha_k = g^*,\ \big(\beta_i-\beta_i^*\big)_{i\in\mathcal{A}}=0,\ \big(\alpha_i-\alpha_i^*\big)_{i\in\mathcal{A}}=0\Big\},
\end{aligned}
$$

where h is given by (62). Therefore, $\Theta^*\cap\{\underline{\theta}=0\} = \Theta^*\cap\{\alpha_k=0\ (k\in\mathcal{J}_0)\} = \{\theta^*\}$. Condition **(a)** of [**H2**] obviously holds since h can be smoothly extended on $\mathbb{R}^{\mathsf{p}}$. Condition **(b)** of [**H2**] holds since $\alpha_i > 2^{-1}\alpha_i^* > 0$ $(i\in\mathcal{A})$ if θ is close to θ^*.

We show [**H3**]. From Remark 2, we only need to show [**H3**]$'$. Since $\kappa_g < \kappa_\alpha$ from [**P3**], for any $\theta \in \Theta^*$,

$$\begin{aligned}\kappa_g|g|^q + \kappa_\alpha \sum_{j=1}^{\mathsf{a}} |\alpha_j|^q &= \kappa_g\Big|g^* - \sum_{k\in\mathcal{A}^c}\alpha_k\Big|^q + \kappa_\alpha \sum_{k\in\mathcal{A}^c}|\alpha_k|^q + \kappa_\alpha\sum_{i\in\mathcal{A}}|\alpha_i^*|^q \\ &\geq \kappa_g|g^*|^q - \kappa_g\sum_{k\in\mathcal{A}^c}|\alpha_k|^q + \kappa_\alpha\sum_{k\in\mathcal{A}^c}|\alpha_k|^q + \kappa_\alpha\sum_{i\in\mathcal{A}}|\alpha_i^*|^q \\ &\geq \kappa_g|g^*|^q + \kappa_\alpha\sum_{i\in\mathcal{A}}|\alpha_i^*|^q,\end{aligned}$$

where the equations hold if and only if $\alpha_k = 0$ ($k \in \mathcal{A}^c$). Therefore, under $\theta \in \Theta^*$, the penalty term is minimized if and only if $\alpha_k = 0$ ($k \in \mathcal{A}^c$), that is, if and only if $\theta = \theta^*$. Thus, [**H3**]$'$ holds.

Condition [**H4**] holds for $\xi_{j,T}$ defined in (61) since $\kappa_\alpha > 0$ and $q_j = q < 1$ ($j \in \mathcal{J}$). Finally, we show [**H5**]. Let $\mathcal{G} = \{\phi, \Omega\}$, and take $\mathcal{N}$ as $\mathcal{N} = \mathrm{Int}(\Theta) \cap \big\{(\theta_1, ..., \theta_{\mathsf{p}}) \in \mathbb{R}^{\mathsf{p}}; \theta_1 > 2^{-1}g^*\big\}$. Note that $\mathcal{N}$ satisfies (16) and (17) and that for any $(\theta, \tau) \in \overline{\mathcal{N}} \times \mathcal{T}$ and any $t \geq 0$,

$$\lambda_t\big(\phi(\theta,\tau)\big) \;\geq\; 2^{-1}g^*. \tag{65}$$

For any $(\theta, \tau) \in \overline{\mathcal{N}} \times \mathcal{T}$,

$$\begin{aligned}&\partial_\theta \mathcal{H}_T(\theta^*, \tau) \\ &= \bigg(\int_0^T \frac{v(X_t)}{g^* + \sum_{i\in\mathcal{A}}\alpha_i^* e^{-\beta_i^* X_t^i}} d\tilde{N}_t, \int_0^T \frac{\big(e^{\tau_k X_t^k}\big)_{k\in\mathcal{A}^c}}{g^* + \sum_{i\in\mathcal{A}}\alpha_i^* e^{-\beta_i^* X_t^i}} d\tilde{N}_t,\bigg)'.\end{aligned}$$

Similarly as before, from [**P1**] and Lemma 2,

$$\sup_{\tau\in\mathcal{T}}\bigg|\frac{1}{\sqrt{T}}\int_0^T \frac{\big(e^{\tau_k X_t^k}\big)_{k\in\mathcal{A}^c}}{g^* + \sum_{i\in\mathcal{A}}\alpha_i^* e^{-\beta_i^* X_t^i}} d\tilde{N}_t\bigg| \;=\; O_P(1).$$

Moreover, from the martingale central limit theorem, we obtain

$$\frac{1}{\sqrt{T}}\int_0^T \frac{v(X_t)}{g^* + \sum_{i\in\mathcal{A}}\alpha_i^* e^{-\beta_i^* X_t^i}} d\tilde{N}_t \;\overset{d}{\to}\; \overline{\Delta},$$

checking Lindeberg's condition as for $S_t = S_t(T) := \frac{1}{\sqrt{T}}\int_0^t \frac{v(X_s)}{g^* + \sum_{i\in\mathcal{A}}\alpha_i^* e^{-\beta_i^* X_s^i}} d\tilde{N}_s$ and for any $a > 0$,

$$\begin{aligned}E\bigg[\sum_{t\leq T}\big(\Delta S_t\big)^2 1_{\{|\Delta S_t|>a\}}\bigg] &\leq a^{-1}E\bigg[\sum_{t\leq T}\big|\Delta S_t\big|^3\bigg] \\ &\leq a^{-1}E\bigg[\int_0^T\bigg|\frac{1}{\sqrt{T}}\frac{v(X_t)}{\lambda_t^*}\bigg|^3 dN_t\bigg] \\ &\leq a^{-1}T^{-\frac{1}{2}}\sup_{t\geq 0}E\bigg[\bigg|\frac{v(X_t)}{\lambda_t^*}\bigg|^3\lambda_t^*\bigg] \;\to\; 0\end{aligned}$$

as $T \to \infty$. Thus, the first half of the argument of [**H5**] holds. Furthermore, for any $(\theta, \tau) \in \overline{\mathcal{N}} \times \mathcal{T}$,

$$\begin{aligned}
&\partial_\theta^2 \mathcal{H}_T(\theta,\tau)\\
&= -\int_0^T \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} dN_t + \int_0^T \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)} dN_t - \int_0^T \partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big) dt\\
&= -\int_0^T \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} \lambda_t\big(\phi(\theta^*,\tau)\big) dt - \int_0^T \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} d\tilde{N}_t\\
&\quad + \int_0^T \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)} d\tilde{N}_t + \int_0^T \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)} \lambda_t\big(\phi(\theta^*,\tau)\big) dt\\
&\quad - \int_0^T \partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big) dt.
\end{aligned}$$

Similarly as before, from [**P1**] and (65), we can use Lemma 2, and obtain

$$\begin{aligned}
&\sup_{(\theta,\tau)\in\overline{\mathcal{N}}\times\mathcal{T}} \left| -\frac{1}{\sqrt{T}} \int_0^T \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} d\tilde{N}_t + \frac{1}{\sqrt{T}} \int_0^T \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)} d\tilde{N}_t \right|\\
&= O_P(1).
\end{aligned}$$

Therefore, for any $(\theta, \tau) \in \overline{\mathcal{N}} \times \mathcal{T}$,

$$\begin{aligned}
\frac{1}{T}\partial_\theta^2 \mathcal{H}_T(\theta,\tau) = -\frac{1}{T}\int_0^T \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} \lambda_t\big(\phi(\theta^*,\tau)\big) dt + \frac{1}{T}\int_0^T \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)}\\
\cdot \lambda_t\big(\phi(\theta^*,\tau)\big) dt - \frac{1}{T}\int_0^T \partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big) dt + o_P(1).
\end{aligned}$$

From [**P1**] and (65), we can apply Lemma 3 to

$$\begin{aligned}
f(\gamma, X_t) \;=\; & \frac{\partial_\theta \lambda_t\big(\phi(\theta,\tau)\big)^{\otimes 2}}{\lambda_t\big(\phi(\theta,\tau)\big)^2} \lambda_t\big(\phi(\theta^*,\tau)\big) + \frac{\partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big)}{\lambda_t\big(\phi(\theta,\tau)\big)} \lambda_t\big(\phi(\theta^*,\tau)\big)\\
& - \partial_\theta^2 \lambda_t\big(\phi(\theta,\tau)\big),
\end{aligned}$$

where $\gamma = (\theta, \tau) \in \overline{\mathcal{N}} \times \mathcal{T}$. Therefore, for any $R > 0$,

$$\sup_{\substack{(\theta,\tau)\in\overline{\mathcal{N}}\times\mathcal{T}\\ |a_T^{-1}(\theta-\theta^*)|\le R}} \left| a_T' \partial_\theta^2 \mathcal{H}_T(\theta,\tau) a_T + \begin{pmatrix} \Gamma(\theta^*) & O \\ O & O \end{pmatrix} \right| \xrightarrow{P} 0.$$

Thus, the second half of the argument of [**H5**] holds. Then using Theorem 2, we have $a_T^{-1}(\hat{\theta}_T - \theta^*) \xrightarrow{d} \hat{u}$. This implies (25) and (26).

We see easily that [**S**] holds from Example 1. Therefore, using Theorem 3, we have $\lim_{T\to\infty} P\big[\hat{\alpha}_{k,T} = 0 \ (k \in \mathcal{A}^c)\big] = 1$. □

Proof of Theorem 5. From (30), we can take some large $L > 0$ such that the following matrix is non-degenerate:

$$\int_{[0,\infty)^{\mathsf{a}}} \big\{(x_j)_{j\in D}\big\}^{\otimes 2}\varphi_L(x)\big\{\alpha^*\cdot x1_{\{\alpha^*\cdot x>0\}} + 1_{\{\alpha^*\cdot x=0\}}\big\}\nu(dx), \qquad (66)$$

where $\varphi_L : \mathbb{R}^{\mathsf{a}} \to [0,1]$ is a continuous function vanishing outside of $[-L-1, L+1]^{\mathsf{a}}$ and satisfying $\varphi_L \equiv 1$ on $[-L,L]^{\mathsf{a}}$. Choose some constant $M_L > 1$ satisfying that

$$\lambda_t(\alpha)\varphi_L(X_t) \ \le\ M_L \qquad \big(t\ge 0,\ \alpha\in[0,M_\alpha]^{\mathsf{a}}\big). \qquad (67)$$

We define an estimating function $\widetilde{\Psi}_T$ as

$$\widetilde{\Psi}_T(\alpha) \ =\ \Psi_T(\alpha) + \int_0^T \left\{\lambda_t(\alpha) - \frac{\{\lambda_t(\alpha)\}^2}{2M_L}\varphi_L(X_t)\right\}1_{\{\lambda_t^*=0\}}dt \quad \big(\alpha\in[0,M_\alpha]^{\mathsf{a}}\big).$$

Note that $\widetilde{\Psi}_T \ge \Psi_T$. For each T, let $\widetilde{\alpha}_T = (\widetilde{\alpha}_{1,T},...,\widetilde{\alpha}_{\mathsf{a},T})$ be an arbitrary $[0,M_\alpha]^{\mathsf{a}}$-valued random variable that asymptotically maximizes $\widetilde{\Psi}_T(\alpha)$.

In order to show Theorem 5, it is sufficient to show that under [**L1**] and [**L2**],

$$\big(T^{\frac12}(\widetilde{\alpha}_{i,T}-\alpha_i^{**})_{i\in\mathcal{J}_1}, T^{\frac{r}{2q}}(\widetilde{\alpha}_{k,T})_{k\in\mathcal{J}_0}\big) \xrightarrow{d} \big(\overline{\Gamma}^{-1}\overline{\Delta}^{\dagger}, 0\big), \qquad (68)$$

$$\lim_{T\to\infty} P\big[\widetilde{\alpha}_{k,T}=0 \ \ (k\in\mathcal{J}_0),\ \ \widetilde{\alpha}_{i,T}\ne 0 \ \ (i\in\mathcal{J}_1)\big] = 1. \qquad (69)$$

In fact, assume (68) and (69) for any asymptotic maximizer $\widetilde{\alpha}_T$ of $\widetilde{\Psi}_T$. Then since $\lambda^*_\cdot = \sum_{i\in\mathcal{J}_1}\alpha_i^{**}X^i_\cdot$ almost surely and therefore $1_{\{\lambda^*_\cdot=0\}} = 1_{\{X^i_\cdot=0\ (i\in\mathcal{J}_1)\}}$ almost surely, we have

$$\begin{aligned} & P\big[\widetilde{\Psi}_T(\widetilde{\alpha}_T) = \Psi_T(\widetilde{\alpha}_T)\big] \\ =\ & P\left[\int_0^T \left\{\lambda_t(\widetilde{\alpha}_T) - \frac{\{\lambda_t(\widetilde{\alpha}_T)\}^2}{2M_L}\varphi_L(X_t)\right\}1_{\left\{X_t^i=0\ (i\in\mathcal{J}_1)\right\}}dt \ = 0\right] \\ \ge\ & P\big[\widetilde{\alpha}_{k,T}=0 \ \ (k\in\mathcal{J}_0)\big] \to 1 \qquad (T\to\infty) \qquad \big(\because (69)\big). \end{aligned}$$

Therefore, the following evaluation asymptotically holds: $\widetilde{\Psi}_T(\hat{\alpha}_T) \ \le\ \widetilde{\Psi}_T(\widetilde{\alpha}_T) = \Psi_T(\widetilde{\alpha}_T) \ \le\ \Psi_T(\hat{\alpha}_T)$. Thus, together with the inequality $\widetilde{\Psi}_T \ge \Psi_T$, $\widetilde{\Psi}_T(\hat{\alpha}_T)$ is asymptotically equal to $\widetilde{\Psi}_T(\widetilde{\alpha}_T)$. This means that $\hat{\alpha}_T$ also asymptotically maximizes $\widetilde{\Psi}_T$. Then (68) and (69) holds when substituting $\hat{\alpha}_T$ for $\widetilde{\alpha}_T$. Therefore, Theorem 5 holds.

In the following, we show (68) and (69) under [**L1**] and [**L2**]. Define two parameter spaces Θ and $\mathcal{T}$ by $\Theta = [0,M_\alpha]^{\mathsf{a}}$ and $\mathcal{T} = \{1\}$, respectively, and we consider new parameters $\theta\in\Theta$ and $\tau\in\mathcal{T}$ by $\theta=\alpha$ and $\tau=1$, respectively. Define estimators $\hat{\theta}_T$ and $\hat{\tau}_T$ taking values in Θ and $\mathcal{T}$, respectively, as $\hat{\theta}_T = \widetilde{\alpha}_T$, $\hat{\tau}_T = 1$. In the following, we omit τ. Define $\theta^*\in\Theta$ as $\theta^*=\alpha^{**}$. We take p and p_1 as $\mathsf{p}=\mathsf{a}$ and $\mathsf{p}_1 = |\mathcal{J}_1|$, respectively. For notational simplicity, assume that $\mathcal{J}_1 = \{1,...,\mathsf{p}_1\}$ and $\mathcal{J}_0 = \{\mathsf{p}_1+1,...,\mathsf{p}\}$. Define $\mathfrak{a}_T = \mathrm{diag}(\mathfrak{a}_{1,T},...,\mathfrak{a}_{\mathsf{r},T})$

as $\mathfrak{a}_T = T^{-\frac{1}{2}} I_{\mathsf{r}}$. ($\mathsf{r}$ is already defined as $\mathsf{r} = |D|$ in Section 5.) Let $\mathfrak{b}_T := T$ and $\rho_k := \frac{r}{q} > 1$ $(k \in \mathcal{J}_0)$, and take $a_T \in GL(\mathsf{p})$ as a deterministic diagonal matrix defined by

$$(a_T)_{jj} = \begin{cases} T^{-\frac{1}{2}} & \big(j \in \mathcal{J}_1 = \{1, ..., \mathsf{p}\} \setminus \mathcal{J}_0\big) \\ T^{-\frac{r}{2q}} & (j \in \mathcal{J}_0) \end{cases}. \tag{70}$$

Define U_T and U by (8) and (10), respectively. Then from Example 1 of Section 2.3 in Yoshida and Yoshida (2022), Condition [**A3**] holds, and $U = \big\{u = (u_1, ..., u_{\mathsf{p}}) \in \mathbb{R}^{\mathsf{p}}; u_k \geq 0\ (k \in \mathcal{J}_0)\big\}$. Define $c_i \in \mathbb{R}$ $(i \in \mathcal{J}_1)$ and $d_k \in \mathbb{R}$ $(k \in \mathcal{J}_0)$ as $c_i = \kappa_i 1_{\{r=1\}}$ and $d_k = \kappa_k$, respectively. Define a random field $\mathbb{Z}$ as

$$\mathbb{Z}(u) = \overline{\Delta}[(u_i)_{i\in\mathcal{J}_1}] - \frac{1}{2}\overline{\Gamma}\big[\big((u_i)_{i\in\mathcal{J}_1}\big)^{\otimes 2}\big] - q\sum_{i\in\mathcal{J}_1} c_i(\alpha_i^{**})^{q-1}u_i - \sum_{k\in\mathcal{J}_0} d_k|u_k|^q,$$

for any $u = (u_1, ..., u_{\mathsf{p}}) \in \mathbb{R}^{\mathsf{p}}$. We define a U-valued random variable $\hat{u}$ given by $\hat{u} = \big(\overline{\Gamma}^{-1}\overline{\Delta}^{\dagger},\ 0\big)$. Note that with probability 1, $\hat{u}$ becomes a unique maximizer of $\mathbb{Z}$ on U, and [**A4**] holds. Define a random field $\mathcal{H}_T : \Omega \times \Theta \to \mathbb{R} \cup \{-\infty\}$ as

$$\begin{aligned} \mathcal{H}_T(\alpha) \;=\; & \int_0^T \log\lambda_t(\alpha)dN_t - \int_0^T \lambda_t(\alpha)dt \\ & + \int_0^T \left\{\lambda_t(\alpha) - \frac{\{\lambda_t(\alpha)\}^2}{2M_L}\varphi_L(X_t)\right\}1_{\{\lambda_t^*=0\}}dt. \end{aligned}$$

Then the estimation function $\widetilde{\Psi}_T$ can be expressed as

$$\widetilde{\Psi}_T(\alpha) \;=\; \mathcal{H}_T(\alpha) - T^{\frac{r}{2}}\sum_{j=1}^{\mathsf{a}}\kappa_j\alpha_j^q \;=\; \mathcal{H}_T(\alpha) - \sum_{j=1}^{\mathsf{p}}\xi_{j,T}p_j(\alpha_j) \qquad (\alpha \in \Theta),$$

where for any $j \in \mathcal{J} := \mathcal{J}_1 \cup \mathcal{J}_0 = \{1, ..., \mathsf{p}\}$,

$$\xi_{j,T} \;=\; \kappa_j T^{\frac{r}{2}}, \qquad p_j(x) \;=\; |x|^q \qquad (x \in \mathbb{R}). \tag{71}$$

Note that q_j $(j \in \mathcal{J})$ defined in (iii) of Section 3.1 are determined as $q_j = q < 1$.

From Theorem 2, if [**H1**]-[**H5**] hold for $\widetilde{\Psi}_T$ under [**L1**] and [**L2**], then $a_T^{-1}(\widetilde{\alpha}_T - \alpha^{**}) \xrightarrow{d} \hat{u}$, i.e., (68) holds. Therefore, in the following, we show [**H1**]-[**H5**] under [**L1**] and [**L2**].

We first show [**H1**]. We have $\int_0^T \log\lambda_t(\alpha)dN_t = \int_0^T \log\lambda_t(\alpha)1_{\{\lambda_t^*>0\}}dN_t$ almost surely since $\int_0^T 1_{\{\lambda_t^*=0\}}dN_t = 0$ with probability one. Then from Taylor's series,

$$
\begin{aligned}
&\mathcal{H}_T(\alpha) - \mathcal{H}_T(\alpha^{**}) \\
&= \int_0^T \log \lambda_t(\alpha) 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T \lambda_t(\alpha) dt + \int_0^T \left\{ \lambda_t(\alpha) - \frac{\{\lambda_t(\alpha)\}^2}{2M_L} \varphi_L(X_t) \right\} \\
&\quad \cdot 1_{\{\lambda_t^*=0\}} dt - \int_0^T \log \lambda_t^* 1_{\{\lambda_t^*>0\}} dN_t + \int_0^T \lambda_t^* dt \\
&= \int_0^T \log \frac{\lambda_t(\alpha)}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T \left\{ \lambda_t(\alpha) - \lambda_t^* \right\} 1_{\{\lambda_t^*>0\}} dt \\
&\quad - \int_0^T \frac{\{\lambda_t(\alpha)\}^2}{2M_L} \varphi_L(X_t) 1_{\{\lambda_t^*=0\}} dt \\
&= \int_0^T \frac{\lambda_t(\alpha) - \lambda_t^*}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^1 (1-s) \int_0^T \frac{\left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2}{\left\{ s\lambda_t(\alpha) + (1-s)\lambda_t^* \right\}^2} 1_{\{\lambda_t^*>0\}} \\
&\quad \cdot dN_t ds - \int_0^T \left\{ \lambda_t(\alpha) - \lambda_t^* \right\} 1_{\{\lambda_t^*>0\}} dt - \int_0^T \frac{\{\lambda_t(\alpha)\}^2}{2M_L} \varphi_L(X_t) 1_{\{\lambda_t^*=0\}} dt \\
&= \int_0^T \frac{\lambda_t(\alpha) - \lambda_t^*}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t - \int_0^1 (1-s) \int_0^T \frac{\left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2}{\left\{ s\lambda_t(\alpha) + (1-s)\lambda_t^* \right\}^2} 1_{\{\lambda_t^*>0\}} \\
&\qquad \cdot dN_t ds - \int_0^T \frac{\{\lambda_t(\alpha)\}^2}{2M_L} \varphi_L(X_t) 1_{\{\lambda_t^*=0\}} dt,
\end{aligned}
$$

where $\tilde{N}$ is a martingale defined by $\tilde{N}_t = N_t - \int_0^t \lambda_s^* ds$. The integrand of the second term in the rightmost side is evaluated as for any $\alpha \in \Theta$, $0 \le s \le 1$ and $0 \le t \le T$,

$$
\frac{\left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2}{\left\{ s\lambda_t(\alpha) + (1-s)\lambda_t^* \right\}^2} 1_{\{\lambda_t^*>0\}} \ \ge \ \frac{\left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2}{M_L^2} \varphi_L(X_t) 1_{\{\lambda_t^*>0\}} \qquad \big(\because (67) \big).
$$

Therefore, noting that $M_L^2 > M_L$, we have

$$
\begin{aligned}
&\mathcal{H}_T(\alpha) - \mathcal{H}_T(\alpha^*) \\
&\le \int_0^T \frac{\lambda_t(\alpha) - \lambda_t^*}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t - \frac{1}{2M_L^2} \int_0^T \left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2 \varphi_L(X_t) 1_{\{\lambda_t^*>0\}} dN_t \\
&\quad - \int_0^T \frac{\{\lambda_t(\alpha)\}^2}{2M_L^2} \varphi_L(X_t) 1_{\{\lambda_t^*=0\}} dt \\
&= \int_0^T \frac{\lambda_t(\alpha) - \lambda_t^*}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t - \frac{1}{2M_L^2} \int_0^T \left\{ \lambda_t(\alpha) - \lambda_t^* \right\}^2 \varphi_L(X_t) \big\{ 1_{\{\lambda_t^*>0\}} dN_t \\
&\qquad + 1_{\{\lambda_t^*=0\}} dt \big\}.
\end{aligned}
$$

From (33), we have

$$
\lambda_t(\alpha) - \lambda_t^* = (X_t^j)_{j \in D}[h(\alpha)] \qquad (t \ge 0, \alpha \in [0, M_\alpha]^{\mathsf{a}}),
$$

where $h = (h_1, ..., h_{\mathsf{r}}) : \Theta \to \mathbb{R}^{\mathsf{r}}$ is a continuous function defined by

$$h(\alpha) = (\alpha - \alpha^*)A = (\alpha - \alpha^{**})A \qquad (\alpha \in [0, M_\alpha]^{\mathsf{a}}). \tag{72}$$

Then

$$\begin{aligned}
&\mathcal{H}_T(\alpha) - \mathcal{H}_T(\alpha^*)\\
\le\ & \frac{1}{\sqrt{T}}\int_0^T \frac{(X_t^j)_{j\in D}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t\big[\sqrt{T}h(\alpha)\big] - \frac{1}{2M_L^2 T}\\
&\quad \cdot \int_0^T \big\{(X_t^j)_{j\in D}\big\}^{\otimes 2}\varphi_L(X_t)\big\{1_{\{\lambda_t^*>0\}}dN_t + 1_{\{\lambda_t^*=0\}}dt\big\}\big[\big(\sqrt{T}h(\alpha)\big)^{\otimes 2}\big]\\
=\ & K_T\big[\sqrt{T}h(\alpha)\big] - \frac{1}{2}\big\{G + r_T\big\}\big[\big(\sqrt{T}h(\alpha)\big)^{\otimes 2}\big] \qquad\qquad (t \ge 0, \alpha \in \Theta),
\end{aligned}$$

where

$$\begin{aligned}
K_T\ &=\ \frac{1}{\sqrt{T}}\int_0^T \frac{(X_t^j)_{j\in D}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t,\\
G\ &=\ \frac{1}{M_L^2}\int_{[0,\infty)^{\mathsf{a}}} \big\{(x_j)_{j\in D}\big\}^{\otimes 2}\varphi_L(x)\left\{\alpha^*\cdot x\, 1_{\{\alpha^*\cdot x>0\}} + 1_{\{\alpha^*\cdot x=0\}}\right\}\nu(dx),\\
r_T\ &=\ \frac{1}{M_L^2 T}\int_0^T \big\{(X_t^j)_{j\in D}\big\}^{\otimes 2}\varphi_L(X_t) 1_{\{\lambda_t^*>0\}} d\tilde{N}_t\\
&\qquad + \frac{1}{M_L^2 T}\int_0^T \big\{(X_t^j)_{j\in D}\big\}^{\otimes 2}\varphi_L(X_t)\left\{\lambda_t^* 1_{\{\lambda_t^*>0\}} + 1_{\{\lambda_t^*=0\}}\right\}dt - G.
\end{aligned}$$

Since from [**L2**],

$$\begin{aligned}
\sup_{T>0} E\big[|K_T|^2\big]\ &=\ \sup_{T>0} E\left[\frac{1}{T}\int_0^T \left|\frac{(X_t^j)_{j\in D}}{\lambda_t^*}\right|^2 \lambda_t^* 1_{\{\lambda_t^*>0\}}dt\right]\\
&\le\ \sup_{t\ge 0} E\left[\left|\frac{(X_t^j)_{j\in D}}{\lambda_t^*}\right|^2 \lambda_t^* 1_{\{\lambda_t^*>0\}}dt\right]\\
&=\ \int_{[0,\infty)^{\mathsf{a}}} \frac{\big|(x_j)_{j\in D}\big|^2}{\alpha^*\cdot x} 1_{\{\alpha^*\cdot x>0\}}\nu(dx)\ <\ \infty,
\end{aligned}$$

we obtain $K_T = O_P(1)$. Similarly,

$$\frac{1}{M_L^2 T}\int_0^T \big\{(X_t^j)_{j\in D}\big\}^{\otimes 2}\varphi_L(X_t) 1_{\{\lambda_t^*>0\}} d\tilde{N}_t\ =\ o_P(1).$$

From the ergodicity, we have

$$\frac{1}{M_L^2 T}\int_0^T \big\{(X_t^j)_{j\in D}\big\}^{\otimes 2}\varphi_L(X_t)\left\{\lambda_t^* 1_{\{\lambda_t^*>0\}} + 1_{\{\lambda_t^*=0\}}\right\} - G \xrightarrow{P} 0.$$

Therefore, $r_T = o_P(1)$. Moreover, (66) implies the non-degeneracy of G. Thus, [**H1**] holds.

Second, we show [**H2**]. From Lemma 5 described below, $\mathrm{Ker}A \cap \langle \{e_j\}_{j\in\mathcal{J}_1}\rangle = \{0\}$. Therefore, for any $\alpha \in \Theta$ with $\alpha_k = 0$ $(k \in \mathcal{J}_0)$,

$$|h(\alpha)| \;=\; |(\alpha - \alpha^{**})A| \;=\; \left|\left(\sum_{j\in\mathcal{J}_1} \alpha_j e_j - \alpha^{**}\right)A\right| \;\geq\; \epsilon_0 |(\alpha_j)_{j\in\mathcal{J}_1} - (\alpha_j^{**})_{j\in\mathcal{J}_1}|,$$

where ϵ_0 is some positive constant. Therefore, **(a)** and **(b)** of [**H2**] hold, and $\Theta^* \cap \{\underline{\theta} = 0\} = \{\alpha \in \Theta; h(\alpha) = 0\} \cap \{\alpha_k = 0 \ (k \in \mathcal{J}_0)\} = \{\theta^*\}$. Thus, [**H2**] holds.

From [**L1**], Condition [**H3**]$'$ holds for $\Theta^* = \{\alpha \in \Theta; h(\alpha) = 0\} = \{\alpha^* + \mathrm{Ker}(A)\} \cap [0, M_\alpha]^{\mathsf{a}}$, which implies [**H3**]. Condition [**H4**] obviously holds for $\xi_{j,T}$ defined in (71) since $\kappa_j > 0$ and $q_j = q < 1$ $(j = 1, ..., \mathsf{a})$.

Finally, we show [**H5**]. Take $\mathcal{G} = \{\phi, \Omega\}$, and take $\mathcal{N}$ as

$$\mathcal{N} \;=\; \mathrm{Int}(\Theta) \cap \left\{\alpha \in \mathbb{R}^{\mathsf{a}}; |\alpha_i - \alpha_i^{**}| < 2^{-1}\alpha_i^{**} \quad (i \in \mathcal{J}_1)\right\}.$$

Note that $\mathcal{N}$ satisfies (16) and (17) and that for any $\alpha \in \overline{\mathcal{N}}$, $\lambda_t^* > 0$ implies $\lambda_t(\alpha) > 0$ since $\lambda_t(\alpha) \geq 2^{-1}\lambda_t^*$. We have

$$\begin{aligned}
\partial_\theta \mathcal{H}_T(\alpha^{**}) \;&=\; \int_0^T \frac{X_t}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T X_t dt + \int_0^T X_t 1_{\{\lambda_t^*=0\}} dt \\
&=\; \int_0^T \frac{X_t}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t.
\end{aligned}$$

Similarly as before, from [**L2**],

$$\frac{1}{\sqrt{T}}\left|\int_0^T \frac{(X_t^j)_{j\in\mathcal{J}_0}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t\right| \;=\; O_P(1).$$

Moreover, from [**L2**] and the martingale central limit theorem, we have

$$S_T := \frac{1}{\sqrt{T}} \int_0^T \frac{(X_t^j)_{j\in\mathcal{J}_1}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t \;\overset{d}{\to}\; \overline{\Delta},$$

checking Lindeberg's condition as for any $a > 0$,

$$\begin{aligned}
E\sum_{t\leq T} \big(\Delta S_t\big)^2 1_{\{|\Delta S_t|>a\}} \;&\leq\; a^{-1} E \sum_{t\leq T} \big|\Delta S_t\big|^3 \\
&\leq\; a^{-1} E\left[\int_0^T \left|\frac{1}{\sqrt{T}} \frac{(X_t^j)_{j\in\mathcal{J}_1}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}}\right|^3 dN_t\right] \\
&=\; a^{-1} T^{-\frac{1}{2}} E\left[\left|\frac{(X_0^j)_{j\in\mathcal{J}_1}}{\lambda_0^*}\right|^3 \lambda_0^* 1_{\{\lambda_0^*>0\}}\right] \to 0
\end{aligned}$$

as $T \to \infty$. Thus, the first half of the argument of [**H5**] holds.

Let us show the second half of the argument of [**H5**]. That is, take any $R > 0$ and we show

$$\sup_{\substack{\alpha\in\overline{\mathcal{N}} \\ |a_T^{-1}(\alpha-\alpha^{**})|\leq R}} \left| a_T' \partial_\theta^2 \mathcal{H}_T(\alpha) a_T + \begin{pmatrix} \Gamma & O \\ O & O \end{pmatrix} \right| \;\overset{P}{\to}\; 0, \qquad (73)$$

for $\partial_\theta^2 \mathcal{H}_T(\alpha)$ that satisfies the following equation:

$$\begin{aligned}
&\partial_\theta^2 \mathcal{H}_T(\alpha) \\
&= -\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T \frac{X_t^{\otimes 2}}{M_L} \varphi_L(X_t) 1_{\{\lambda_t^*=0\}} dt \\
&= -\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T \frac{X_t^{\otimes 2}}{M_L} \varphi_L(X_t) 1_{\left\{X_t^i=0 \ (i\in\mathcal{J}_1)\right\}} dt \\
&= -\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t - \int_0^T \frac{\left\{(X_t^j 1_{\{j\in\mathcal{J}_0\}})_{j=1,\dots,\mathsf{p}}\right\}^{\otimes 2}}{M_L} \varphi_L(X_t) \\
&\qquad \cdot 1_{\left\{X_t^i=0 \ (i\in\mathcal{J}_1)\right\}} dt \qquad (\alpha \in \overline{\mathcal{N}}).
\end{aligned}$$

Since a_T is defined as (70) and $\frac{r}{q} > 1$, we have

$$\begin{aligned}
&a_T' \partial_\theta^2 \mathcal{H}_T(\alpha) a_T \\
&= a_T' \bigg(-\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t \bigg) a_T - \frac{1}{T^{\frac{r}{q}}} \int_0^T \frac{\left\{(X_t^j 1_{\{j\in\mathcal{J}_0\}})_{j=1,\dots,\mathsf{p}}\right\}^{\otimes 2}}{M_L} \\
&\qquad \cdot \varphi_L(X_t) 1_{\left\{X_t^i=0 \ (i\in\mathcal{J}_1)\right\}} dt \\
&= (\sqrt{T} a_T)' \bigg(-\frac{1}{T} \int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t \bigg) (\sqrt{T} a_T) + o(1).
\end{aligned}$$

Thus, for (73), it suffices to show

$$\sup_{\substack{\alpha\in\overline{\mathcal{N}} \\ |a_T^{-1}(\alpha-\alpha^{**})|\le R}} \left| -\frac{1}{T} \int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t + \Gamma \right| \overset{P}{\to} 0, \tag{74}$$

where Γ is a $\mathsf{a} \times \mathsf{a}$ matrix defined as $\Gamma = \int_{[0,\infty)^{\mathsf{a}}} \frac{x^{\otimes 2}}{\alpha^* \cdot x} 1_{\{\alpha^* \cdot x > 0\}} \nu(dx)$. Take any $\epsilon > 0$. Then there exists some $T_0 = T_0(R, \epsilon)$ such that for any $T \ge T_0$ and any $\alpha \in \overline{\mathcal{N}}$ with $|a_T^{-1}(\alpha - \alpha^{**})| \le R$,

$$(1-\epsilon)\lambda_\cdot^* \le \lambda_\cdot(\alpha) \ \le \lambda_\cdot^* + \epsilon \sum_{j\in D} X_\cdot^j$$

almost surely. Therefore, for any $T \ge T_0$ and any $\alpha \in \overline{\mathcal{N}}$ with $|a_T^{-1}(\alpha - \alpha^{**})| \le R$,

$$A_T[u^{\otimes 2}] \ \le \ \frac{1}{T} \int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t[u^{\otimes 2}] \le B_T[u^{\otimes 2}] \qquad (u \in \mathbb{R}^{\mathsf{a}}, |u| \le 1),$$

where

$$\begin{aligned}
A_T &= \frac{1}{T} \int_0^T \frac{X_t^{\otimes 2}}{\left(\lambda_t^* + \epsilon \sum_{j\in D} X_t^j\right)^2} 1_{\{\lambda_t^*>0\}} dN_t, \\
B_T &= \frac{1}{T} (1-\epsilon)^{-2} \int_0^T \frac{X_t^{\otimes 2}}{(\lambda_t^*)^2} 1_{\{\lambda_t^*>0\}} dN_t,
\end{aligned}$$

for any $T > 1$, any $u \in \mathbb{R}^{\mathsf{a}}$ with $|u| \le 1$ and any $\alpha \in \overline{\mathcal{N}}$. Since from [**L2**],

$$\begin{aligned} & E\bigg[\bigg|\frac{1}{\sqrt{T}}\int_0^T \frac{X_t^{\otimes 2}}{\big(\lambda_t^* + \epsilon\sum_{j\in D} X_t^j\big)^2} 1_{\{\lambda^*>0\}} d\tilde{N}_t\bigg|^2\bigg] \\ \lesssim\ & E\bigg[\frac{1}{T}\int_0^T \frac{|X_t|^4}{\big(\lambda_t^* + \epsilon\sum_{j\in D} X_t^j\big)^4}\lambda_t^* 1_{\{\lambda_t^*>0\}} dt\bigg] \\ \le\ & E\bigg[\frac{1}{T}\int_0^T \frac{|X_t|^4}{(\lambda_t^*)^4}\lambda_t^* 1_{\{\lambda_t^*>0\}} dt\bigg] \\ =\ & \int_{[0,\infty)^{\mathsf{a}}} \frac{|x|^4}{\big(\alpha^*\cdot x\big)^3} 1_{\{\alpha^*\cdot x>0\}} \nu(dx)\ <\ \infty \end{aligned}$$

and since

$$\begin{aligned} & \int_{[0,\infty)^{\mathsf{a}}} \frac{|x^{\otimes 2}|}{\big(\alpha^*\cdot x + \epsilon\sum_{j\in D} x_j\big)^2}\, \alpha^*\cdot x\ 1_{\{\alpha^*\cdot x>0\}}\ \nu(dx) \\ \le\ & \int_{[0,\infty)^{\mathsf{a}}} \frac{|x^{\otimes 2}|}{\alpha^*\cdot x}\, 1_{\{\alpha^*\cdot x>0\}}\ \nu(dx) < \infty, \end{aligned}$$

we have

$$\begin{aligned} A_T\ =\ & \frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\big(\lambda_t^* + \epsilon\sum_{j\in D} X_t^j\big)^2} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t \\ & + \frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\big(\lambda_t^* + \epsilon\sum_{j\in D} X_t^j\big)^2}\lambda_t^* 1_{\{\lambda_t^*>0\}} dt \\ \stackrel{P}{\to}\ & \int_{[0,\infty)^{\mathsf{a}}} \frac{x^{\otimes 2}}{\big(\alpha^*\cdot x + \epsilon\sum_{j\in D} x_j\big)^2}\, \alpha^*\cdot x\, 1_{\{\alpha^*\cdot x>0\}}\ \nu(dx)\ \ =:\ A(\epsilon). \end{aligned}$$

Similarly,

$$\begin{aligned} B_T\ =\ & (1-\epsilon)^{-2}\bigg\{\frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{(\lambda_t^*)^2} 1_{\{\lambda_t^*>0\}} d\tilde{N}_t + \frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t^*} 1_{\{\lambda_t^*>0\}} dt\bigg\} \\ \stackrel{P}{\to}\ & (1-\epsilon)^{-2}\int_{[0,\infty)^{\mathsf{a}}} \frac{x^{\otimes 2}}{\alpha^*\cdot x} 1_{\{\alpha^*\cdot x>0\}} \nu(dx)\ \ =:\ B(\epsilon). \end{aligned}$$

Then for any $T \geq T_0$,

$$\begin{aligned}
&\sup_{\substack{\alpha\in\overline{\mathcal{N}}\\ |a_T^{-1}(\alpha-\alpha^{**})|\leq R}} \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left| -\frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right| \\
\leq\ & \sup_{\substack{\alpha\in\overline{\mathcal{N}}\\ |a_T^{-1}(\alpha-\alpha^{**})|\leq R}} \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left(-\frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right)^+ \\
&+ \sup_{\substack{\alpha\in\overline{\mathcal{N}}\\ |a_T^{-1}(\alpha-\alpha^{**})|\leq R}} \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left(-\frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right)^- \\
\leq\ & \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left(-A_T[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right)^+ + \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left(-B_T[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right)^- \\
\lesssim\ & \left| -A_T + \Gamma \right| + \left| -B_T + \Gamma \right| \overset{P}{\to} \left| -A(\epsilon) + \Gamma \right| + \left| -B(\epsilon) + \Gamma \right|
\end{aligned}$$

as $T \to \infty$, where $f^+ := f \vee 0$ and $f^- := f \wedge 0$ for any $\mathbb{R}$-valued function f. Since $\lim_{\epsilon\to+0} A(\epsilon) = \lim_{\epsilon\to+0} B(\epsilon) = \Gamma$, we have

$$\sup_{\substack{\alpha\in\overline{\mathcal{N}}\\ |a_T^{-1}(\alpha-\alpha^{**})|\leq R}} \sup_{u\in\mathbb{R}^{\mathsf{a}},|u|\leq 1} \left| -\frac{1}{T}\int_0^T \frac{X_t^{\otimes 2}}{\lambda_t(\alpha)^2} 1_{\{\lambda_t^*>0\}} dN_t[u^{\otimes 2}] + \Gamma[u^{\otimes 2}] \right| \overset{P}{\to} 0.$$

Therefore, (74) holds. Thus, the second half of the argument of [**H5**] holds. Then using Theorem 2, we obtain (68).

Furthermore, [**S**] holds from Example 1. Therefore, using Theorem 3, we have $\lim_{T\to\infty} P\big[(\widetilde{\alpha}_{k,T})_{k\in\mathcal{J}_0} = 0\big] = 1$. Since (68) holds and $\alpha_i^{**} \neq 0$ $(i \in \mathcal{J}_1)$, we obtain (69). □

Proof of Proposition 1. From the following Lemma 4,

$$\begin{aligned}
&\left\{ \alpha \in \{\alpha^* + \mathrm{Ker}(A)\} \cap [0,\infty)^{\mathsf{a}}; Pe(\alpha) = \inf_{\tilde{\alpha}\in\{\alpha^*+\mathrm{Ker}(A)\}\cap[0,\infty)^{\mathsf{a}}} Pe(\tilde{\alpha}) \right\} \\
\subset\ & \{pr_E(\alpha^*); E \in \mathcal{E}\} \cap [0,\infty)^{\mathsf{a}}.
\end{aligned}$$

Under $[\mathbf{L1}]^{\#}$, the set on the right-hand side has the unique minimizer $pr_{E_0}(\alpha^*)$ of Pe on the set itself. Therefore, $pr_{E_0}(\alpha^*)$ uniquely minimizes Pe on $\{\alpha^* + \mathrm{Ker}(A)\} \cap [0,\infty)^{\mathsf{a}}$. Since $pr_{E_0}(\alpha^*) \in [0, M_\alpha]^{\mathsf{a}}$ under $[\mathbf{L1}]^{\#}$, $pr_{E_0}(\alpha^*)$ also uniquely minimizes Pe on $\{\alpha^* + \mathrm{Ker}(A)\} \cap [0, M_\alpha]^{\mathsf{a}}$. Therefore, [**L1**] holds and $\alpha^{**} = pr_{E_0}(\alpha^*)$. □

Lemma 4 *For any $M \in (0,\infty) \cup \{\infty\}$,*

$$\begin{aligned}
&\left\{ \alpha \in \{\alpha^* + \mathrm{Ker}(A)\} \cap [0,M)^{\mathsf{a}}; Pe(\alpha) = \inf_{\tilde{\alpha}\in\{\alpha^*+\mathrm{Ker}(A)\}\cap[0,M)^{\mathsf{a}}} Pe(\tilde{\alpha}) \right\} \\
\subset\ & \{pr_E(\alpha^*); E \in \mathcal{E}\}. \qquad (75)
\end{aligned}$$

Proof Take any $\alpha \in \{\alpha^* + \mathrm{Ker}(A)\} \cap [0, M)^{\mathsf{a}}$ satisfying

$$Pe(\alpha) = \inf_{\tilde{\alpha} \in \{\alpha^* + \mathrm{Ker}(A)\} \cap [0,M)^{\mathsf{a}}} Pe(\tilde{\alpha}).$$

Define F as the set of all $j \in \{1, ..., \mathsf{a}\}$ with $\alpha_j \neq 0$. We prove

$$\langle \{e_j\}_{j \in F} \rangle \cap \mathrm{Ker}A = \{0\} \tag{76}$$

by contradiction. Suppose that $\langle \{e_j\}_{j \in F} \rangle \cap \mathrm{Ker}A \neq \{0\}$. Then there exists $(c_j)_{j \in F} \in \mathbb{R}^{|F|} \setminus \{0\}^{|F|}$ such that $\sum_{j \in F} c_j e_j \in \mathrm{Ker}A$. Take $\underline{\lambda} < 0$ and $\overline{\lambda} > 0$ such that for any $\lambda \in (\underline{\lambda}, \overline{\lambda})$,

$$\alpha - \lambda \sum_{j \in F} c_j e_j \in [0, M)^{\mathsf{a}}. \tag{77}$$

Note that for any $\lambda \in (\underline{\lambda}, \overline{\lambda})$,

$$\alpha - \lambda \sum_{j \in F} c_j e_j \in \{\alpha^* + \mathrm{Ker}(A)\}. \tag{78}$$

Define two functions $f : (\underline{\lambda}, \overline{\lambda}) \to \mathbb{R}^{\mathsf{a}}$ and $g : (\underline{\lambda}, \overline{\lambda}) \to \mathbb{R}$ as $f(\lambda) = \alpha - \lambda \sum_{j \in F} c_j e_j$ and $g(\lambda) = Pe\big(f(\lambda)\big)$, respectively. Then $g(0)$ cannot be the local minimum of g since $g''(\lambda) < 0$ for any $\lambda \in (\underline{\lambda}, \overline{\lambda})$. Therefore, there exists some $\lambda_0 \in (\underline{\lambda}, \overline{\lambda})$ such that

$$Pe(\alpha) \;=\; g(0) \;>\; g(\lambda_0) \;=\; Pe\big(f(\lambda_0)\big).$$

Furthermore, $f(\lambda_0) \in \{\alpha^* + \mathrm{Ker}(A)\} \cap [0, M)^{\mathsf{a}}$ from (77) and (78). This contradicts the minimality of α. Thus, (76) holds.

From (76), we can take some $E_1 \in \mathcal{E}$ with $E_1 \supset F$. Then $0 = (\alpha - \alpha^*)A = \big(\alpha - pr_{E_1}(\alpha^*)\big)A$. Since $\alpha \in \langle \{e_j\}_{j \in F} \rangle \subset \langle \{e_j\}_{j \in E_1} \rangle$ and $\langle \{e_j\}_{j \in E_1} \rangle \cap \mathrm{Ker}A = 0$, we have $\alpha = pr_{E_1}(\alpha^*)$, and hence $\alpha \in \{pr_E(\alpha^*); E \in \mathcal{E}\}$. Thus, (75) holds. □

Lemma 5 *Assume* [**L1**]. *Then*

$$\mathrm{Ker}A \cap \langle \{e_j\}_{j \in \mathcal{J}_1} \rangle = \{0\}. \tag{79}$$

Moreover, if [**L2**] *holds, then* $\overline{\Gamma}$ *is non-degenerate.*

Proof From Lemma 4, under [**L1**], $\alpha^{**} \in \{pr_E(\alpha^*); E \in \mathcal{E}\}$. Therefore, there exists some $E \in \mathcal{E}$ such that $\mathcal{J}_1 \subset E$. Thus, $\mathrm{Ker}A \cap \langle \{e_j\}_{j \in \mathcal{J}_1} \rangle \subset \mathrm{Ker}A \cap \langle \{e_j\}_{j \in E} \rangle = \{0\}$, and (79) holds.

We show that $\overline{\Gamma}$ is non-degenerate under [**L2**]. Assume $\overline{\Gamma}[v^{\otimes 2}] = 0$ for some $v \in \mathbb{R}^{|\mathcal{J}_1|}$. Then

$$\big|(x_i)_{i \in \mathcal{J}_1} \cdot v\big| \, 1_{\left\{ \sum_{j=1}^{\mathsf{a}} \alpha^*_j x_j > 0 \right\}} \;=\; 0 \quad \nu\text{-}a.e.\, x, \tag{80}$$

where ν-$a.e.\,x$ denotes almost everywhere $x \in [0,\infty)^{\mathsf{a}}$ with respect to the measure ν. Since the probability that $\lambda_t^* = \sum_{j=1}^{\mathsf{a}} \alpha_j^* X_t^j = \sum_{j=1}^{\mathsf{a}} \alpha_j^{**} X_t^j$ $(t \geq 0)$ is equal to one, we have

$$1_{\left\{\sum_{j=1}^{\mathsf{a}} \alpha_j^* x_j = 0\right\}} = 1_{\left\{\sum_{j=1}^{\mathsf{a}} \alpha_j^{**} x_j = 0\right\}} = 1_{\{x_i=0\ (i\in\mathcal{J}_1)\}} \quad \nu\text{-}a.e.\,x.$$

Therefore, from (80), we have $(x_i)_{i\in\mathcal{J}_1} \cdot v = 0$ ν-$a.e.\,x$. Thus,

$$\begin{aligned}
0 &= \int_{[0,\infty)^{\mathsf{a}}} \big((x_i)_{i\in\mathcal{J}_1}\big)^{\otimes 2} \nu(dx)\ [v^{\otimes 2}] = \int_{[0,\infty)^{\mathsf{a}}} \left\{\big(e_i \cdot x\big)_{i\in\mathcal{J}_1}\right\}^{\otimes 2} \nu(dx)\ [v^{\otimes 2}] \\
&= \int_{[0,\infty)^{\mathsf{a}}} \left\{\big(e_i A \cdot (x_j)_{j\in D}\big)_{i\in\mathcal{J}_1}\right\}^{\otimes 2} \nu(dx)\ [v^{\otimes 2}] \qquad \big(\because (32)\big) \\
&= (e_i A)_{i\in\mathcal{J}_1} \int_{[0,\infty)^{\mathsf{a}}} \big((x_j)_{j\in D}\big)^{\otimes 2} \nu(dx)\ \big((e_i A)_{i\in\mathcal{J}_1}\big)'\ [v^{\otimes 2}].
\end{aligned}$$

Now $\{e_i A\}_{i\in\mathcal{J}_1}$ is linearly independent from (79). Therefore, from (30), we obtain $v = 0$, which implies the non-degeneracy of $\overline{\Gamma}$. □