Econometrica Supplementary Material

SUPPLEMENT TO "MAXIMUM LIKELIHOOD ESTIMATION IN MARKOV REGIME-SWITCHING MODELS WITH COVARIATE-DEPENDENT TRANSITION PROBABILITIES"

(*Econometrica*, Vol. 90, No. 4, July 2022, 1681–1710)

DEMIAN POUZO
Department of Economics, University of California, Berkeley

ZACHARIAS PSARADAKIS
Department of Economics, Mathematics and Statistics, Birkbeck, University of London

MARTIN SOLA
Department of Economics, Universidad Torcuato Di Tella

SM.1. ERGODICITY AND STATIONARITY

LET $(\zeta_t)_{t=-\infty}^{\infty}$ be a Markov chain with transition kernel $\zeta \mapsto \mathbb{P}(\zeta, \cdot) \in \mathcal{P}(\mathbb{Z})$ and $\zeta_t \in \mathbb{Z} \subseteq \mathbb{R}^d$ for some $d > 0$. Also, for any probability measure P over $\mathbb{Z}$ and any $f : \mathbb{Z} \to \mathbb{R}$, let $P[f](z) \equiv \int f(u)P(z, du)$ (if it exists).

ASSUMPTION 9: *There exist constants* $\gamma \in (0, 1)$, $\lambda \in (0, 1)$, $b > 0$ *and* $R > 2b/(1 - \gamma)$, *a function* $\mathbf{V} : \mathbb{Z} \to [1, \infty)$, *and a probability measure* ϱ *such that*: (*i*) $\mathbb{P}[\mathbf{V}](\zeta) \leq \gamma \mathbf{V}(\zeta) + b1\{\zeta \in \mathcal{C}\}$ *for all* $\zeta \in \mathbb{Z}$ *with* $\mathcal{C} \equiv \{\zeta \in \mathbb{Z} : \mathbf{V}(\zeta) \leq R\}$; (*ii*) $\inf_{\zeta \in \mathcal{C}} \mathbb{P}(\zeta, \cdot) \geq \lambda \varrho(\cdot)$, *with* $\varrho(\mathcal{C}) > 0$.

The next result is used for the proof of Lemma 1; it contains well-known results that are stated and proved here for convenience. In particular, the first part of Lemma 10 is a restatement of Theorem 1.2 in Hairer and Mattingly (2011). The second part of Lemma 10 and Assumption 9(ii) imply that $\mathbb{P}$ is Harris recurrent (see Athreya and Lahiri (2006, Ch. 14)) and aperiodic (see Thierney (1996, p. 65)). The proof follows from standard arguments.

Let $v \mapsto \|v\|_{\mathbf{V}} \equiv \sup_{\zeta} \frac{|v(\zeta)|}{1+\mathbf{V}(\zeta)}$. Also, for any $A \subseteq \mathbb{Z}$, let $T_A = \inf\{t \geq 0 \colon \zeta_t \in A\}$. Finally, for any two sequences $(X_n)_n$ and $(Y_n)_n$, write $X_n \precsim Y_n$ if $X_n \leq CY_n$ for some universal positive finite constant C.

LEMMA 10: *If Assumption 9 holds, then*:

(i) $\mathbb{P}$ *admits a unique invariant measure* ν^*, *and there exist constants* $\gamma \in (0, 1)$ *and* $C > 0$ *such that*

$$\left\|\mathbb{P}^n[v] - \nu^*[v]\right\|_{\mathbf{V}} \leq C\gamma^n \left\|v - \nu^*[v]\right\|_{\mathbf{V}}$$

for every measurable function v *such that* $\|v\|_{\mathbf{V}} < \infty$, *where* $\nu^*[v] \equiv \int v(\zeta)\nu^*(d\zeta)$.

(ii) $\mathbb{P}(\zeta, \{T_{\mathcal{C}} < \infty\}) = 1$ *for all* $\zeta \in \mathbb{Z}$, *and* $\mathbb{P}(\zeta_0, \mathcal{C}) > 0$ *for all* $\zeta_0 \in \mathcal{C}$.

PROOF OF LEMMA 10: The proof follows standard arguments and is thus omitted; it can be found in Pouzo, Psaradakis, and Sola (2021). *Q.E.D.*

Demian Pouzo: dpouzo@berkeley.edu
Zacharias Psaradakis: z.psaradakis@bbk.ac.uk
Martin Sola: msola@utdt.edu

 https://doi.org/10.3982/ECTA17249

PROOF OF LEMMA 1: Let $(\zeta_t)_{t=-\infty}^{\infty}$ be the stochastic process given by $\zeta_t \equiv (X_t, S_t)$. This process is a Markov chain with transition kernel $\mathbb{X} \times \mathbb{S} \ni \zeta \mapsto \mathbb{P}(\zeta, \cdot) \in \mathcal{P}(\mathbb{X} \times \mathbb{S})$ given by

$$\mathbb{P}\big((x,s), \{\zeta_{t+1} \in A_1 \times A_2\}\big) = \sum_{s' \in A_2} Q_*\big(x,s,s'\big) P_*\big(x,s',A_1\big),$$

for any Borel sets $A_1 \subseteq \mathbb{X}$ and $A_2 \subseteq \mathbb{S}$.

By Lemma 10, there exists a unique invariant measure ν, provided that the conditions of Assumption 9 are met. In order to verify the first part of Assumption 9, consider $\mathbf{V}(\zeta) = \mathcal{U}(x)$, and $\mathcal{C} \equiv \mathcal{C}_1 \times \mathbb{S}$ with $\mathcal{C}_1 \equiv \{x \in \mathbb{X} \colon \mathcal{U}(x) \leq R\}$. By Assumption 2(i),

$$\mathbb{P}[\mathbf{V}](\zeta) = \int_{\mathbb{X}} \mathcal{U}(x') \left\{ \sum_{s' \in \mathbb{S}} Q_*\big(x,s,s'\big) P_*\big(x,s',dx'\big) \right\} \leq \gamma \mathcal{U}(x) + 2b' 1\{x \in \mathcal{C}_1\}.$$

Thus, $b \equiv 2b'$. Regarding Assumption 9(ii), observe that, by Assumption 1(i), for C and any $s \in \mathbb{S}$,

$$\mathbb{P}\big((x,s), C \times \{s'\}\big) \geq \underline{q}(x) P_*\big(x,s',C\big),$$

and, by Assumption 2(iii), $P_*(x,s',C) \geq \lambda' \varpi(C)$ and $\lambda' \in (0,1)$. Also note that, by Assumption 1, $\underline{q}$ is continuous and $\underline{q}(x) > 0$ for all $x \in \mathbb{X}$. Furthermore, by Assumption 2(ii), $\mathcal{U}$ is lower semicompact, because $\{x \in \mathbb{X} : \mathcal{U}(x) \leq R\}$ is closed ($x \mapsto \mathcal{U}(x)$ is lower semicontinuous), and is also bounded. Therefore, $\inf_{x:\mathcal{U}(x)\leq R} \underline{q}(x) = \min_{x:\mathcal{U}(x)\leq R} \underline{q}(x) \geq c > 0$ (because it is a minimization of a continuous function on compact set). Therefore,

$$\mathbb{P}\big(\zeta, C \times \{s'\}\big) \geq c\lambda' \varpi(C) \frac{1}{|\mathbb{S}|},$$

and, by putting $\varrho = \varpi(\cdot)\frac{1}{|\mathbb{S}|}$ and $\lambda \equiv c\lambda'$, Assumption 9(ii) follows since $\varpi(\mathcal{C}_1) > 0$. Since ν is unique, it is trivially ergodic. Therefore, the process with initial probability measure ν is stationary. Ergodicity of $(\zeta_t)_t$ follows from Theorem 14.2.11 in Athreya and Lahiri (2006) (recall that $\mathbb{P}$ is Harris recurrent and aperiodic). Since X_t is a deterministic function of ζ_t, X_0^∞ is also stationary and ergodic. Finally, observe that

$$\int \sup_{0\leq f\leq 1} \big|\mathbb{P}^n[f](\zeta) - \nu[f]\big| \nu(d\zeta) \precsim \gamma^n \int \big|1 + \mathcal{U}(x)\big| \nu(d\zeta).$$

Since $\mathcal{U}$ satisfies Assumption 9(i), it follows that $\int \mathbb{P}[\mathcal{U}](\zeta)\nu(d\zeta) \leq \gamma\nu[\mathcal{U}] + K$. Since ν is the invariant measure of $\mathbb{P}$ and $\gamma \in (0,1)$, this implies that $\nu(d\zeta) \leq K/(1-\gamma)$. Therefore,

$$\int \sup_{0\leq f\leq 1} \big|\mathbb{P}^n[f](\zeta) - \nu[f]\big| \nu(d\zeta) \precsim \gamma^n,$$

thereby implying that $(\zeta_t)_t$ is β-mixing with rate $\beta_n = O(\gamma^n)$ (see Davydov (1973)). Since X_t is a deterministic function of ζ_t, the same holds for X_0^∞. *Q.E.D.*

SM.2. PROOFS OF SUPPLEMENTARY LEMMAS IN APPENDIX A.1

To prove Lemmas 2 and 3, we use the following result.

LEMMA 11: *Suppose Assumptions* 1 *and* 4(ii) *hold. Then, for all* $t \in \mathbb{N}$ *and* $-n \leq -m \leq t-1$, *a.s.*-$\bar{P}_*^\nu$,

$$\sup_{\theta \in \Theta} \left| \log p_t^\nu\left(X_t \mid X_{-m}^{t-1}, \theta\right) - \log p_t^\nu\left(X_t \mid X_{-n}^{t-1}, \theta\right) \right| \leq C(X_{t-1}, X_t) \prod_{i=-m}^{t-1} \left(1 - \underline{q}(X_i)\right).$$

PROOF OF LEMMA 11: Observe that, for any $n \in \mathbb{N}$,

$$\log p_t^\nu\left(X_t \mid X_{-n}^{t-1}, \theta\right) = \log \sum_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t) \bar{P}_\theta^\nu\left(s \mid X_{-n}^{t-1}\right),$$

and since $\log x - \log y \leq x/y - 1$, it suffices to study $\frac{\sum_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t)(\bar{P}_\theta^\nu(S_t = s \mid X_{-m}^{t-1}) - \bar{P}_\theta^\nu(S_t = s \mid X_{-n}^{t-1}))}{\sum_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t) \bar{P}_\theta^\nu(s \mid X_{-n}^{t-1})}$. This expression can be bounded above by

$$\frac{\max_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t)}{\min_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t)} \left\| \bar{P}_\theta^\nu\left(S_t = \cdot \mid X_{-m}^{t-1}\right) - \bar{P}_\theta^\nu\left(S_t = \cdot \mid X_{-n}^{t-1}\right) \right\|_1.$$

By Assumption 4(ii), $\sup_{\theta \in \Theta} \frac{\max_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t)}{\min_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t)} \leq C(X_{t-1}, X_t)$ a.s.-$\bar{P}_*^\nu$. So it suffices to bound $\|\bar{P}_\theta^\nu(S_t = \cdot \mid X_{-m}^{t-1}) - \bar{P}_\theta^\nu(S_t = \cdot \mid X_{-n}^{t-1})\|_1$. By Lemma B.2.2 in Stachurski (2009) and the fact that $-n \leq -m$, this is bounded by

$$\frac{1}{2} \sup_{b, c \in \mathbb{S}^2} \left\| \bar{P}_\theta^\nu\left(S_t = \cdot \mid S_{-m} = b, X_{-m}^{t-1}\right) - \bar{P}_\theta^\nu\left(S_t = \cdot \mid S_{-m} = c, X_{-m}^{t-1}\right) \right\|_1.$$

Hence,

$$\sup_{\theta \in \Theta} \left| \log p_t^\nu\left(X_t \mid X_{-m}^{t-1}, \theta\right) - \log p_t^\nu\left(X_t \mid X_{-n}^{t-1}, \theta\right) \right| \leq C' \alpha_{\theta, t, -m}\left(X_{-m}^{t-1}\right),$$

where $\alpha_{\theta, t, -m}(X_{-m}^{t-1})$ is defined in expression (12). By applying Lemmas 6 and 5 and the fact that $\alpha_{\theta, t, -m}(X_{-m}^{t-1}) \leq \prod_{j=-m}^{t-1} \alpha_{\theta, j, j+1}(X_{-m}^{t-1})$, it follows that

$$\sup_{\theta \in \Theta} \left| \log p_t^\nu\left(X_t \mid X_{-m}^{t-1}, \theta\right) - \log p_t^\nu\left(X_t \mid X_{-n}^{t-1}, \theta\right) \right| \leq C(X_{t-1}, X_t) \prod_{i=-m}^{t-1} \left(1 - \underline{q}(X_i)\right),$$

a.s.-$\bar{P}_*^\nu$. *Q.E.D.*

We now prove Lemmas 2 and 3.

PROOF OF LEMMA 2: The result follows from Lemma 11, with $m = 0$ and $n = M$, and Assumption 5. *Q.E.D.*

PROOF OF LEMMA 3: Recall that, by Lemma 1, the process $X_{-\infty}^{\infty}$ is ergodic and stationary under $\bar{P}_*^\nu$.

Part (i). Consider a $\delta > 0$ and an open cover $\{B(\theta, \delta) \colon \theta \in \Theta\}$ where $B(\theta, \delta)$ is an open ball centered around θ with radius $\delta > 0$. Since Θ is compact (Assumption 3), there exists a finite subcover $B_j \equiv B(\theta_j, \delta)$ with $j = 1, \ldots, J$. Also note that, pointwise in $\theta\in\Theta$, $\ell_T^\nu(X_{-\infty}^T,\theta)-E_{\bar{P}_*^\nu}[\ell_T^\nu(X_{-\infty}^T,\theta)]\to 0$ a.s.-$\bar{P}_*^\nu$ by the ergodic theorem and the fact that $X_{-\infty}^\infty\mapsto\ell_T^\nu(X_{-\infty}^T,\theta)\in L^1(\bar{P}_{\theta*}^\nu)$. Thus, it suffices to show that there exists a $T(j,\epsilon)$ such that, for all $t\geq T(j,\epsilon)$,

$$\bar{P}_*^\nu\left(\sup_{\theta\in B_j}T^{-1}\sum_{t=1}^T\big(l_t\big(X_{-\infty}^t,\theta\big)-E_{\bar{P}_*^\nu}\big[l_t\big(X_{-\infty}^t,\theta\big)\big]\big)>\epsilon\right)\leq\epsilon,$$

where $l_t(X_{-\infty}^t,\theta)\equiv\log\frac{p^\nu(X_t|X_{-\infty}^{t-1},\theta)}{p^\nu(X_t|X_{-\infty}^{t-1},\theta_j)}$. Observe that, for any j,

$$\begin{aligned}&\sup_{\theta\in B_j}\sum_{t=1}^T\big(l_t\big(X_{-\infty}^t,\theta\big)-E_{\bar{P}_*^\nu}\big[l_t\big(X_{-\infty}^t,\theta\big)\big]\big)\\&\quad\leq\sum_{t=1}^T\sup_{\theta\in B_j}\big(l_t\big(X_{-\infty}^t,\theta\big)-E_{\bar{P}_*^\nu}\big[l_t\big(X_{-\infty}^t,\theta\big)\big]\big)\equiv\sum_{t=1}^T\bar{l}_t\big(X_{-\infty}^t\big).\end{aligned}$$

Moreover, observe that

$$\sup_{\theta\in B_j}\log\frac{p^\nu\big(X_t\mid X_{-\infty}^{t-1},\theta\big)}{p^\nu\big(X_t\mid X_{-\infty}^{t-1},\theta_j\big)}\leq\sup_{\theta\in B_j}\frac{p^\nu\big(X_t\mid X_{-\infty}^{t-1},\theta\big)}{p^\nu\big(X_t\mid X_{-\infty}^{t-1},\theta_j\big)}-1.$$

By Assumption 4(i), for any $\epsilon>0$ there exists a $\delta>0$ such that $E_{\bar{P}_*^\nu}[\sup_{\theta\in B_j}\frac{p^\nu(X_0|X_{-\infty}^{-1},\theta)}{p^\nu(X_0|X_{-\infty}^{-1},\theta_j)}]\leq 1+\epsilon$ for any $j\in\{1,\dots,J\}$ and any t. Therefore, we can choose a $\delta>0$ such that

$$E_{\bar{P}_*^\nu}\left[\sup_{\theta\in B_j}\log\frac{p^\nu\big(X_0\mid X_{-\infty}^{-1},\theta\big)}{p^\nu\big(X_0\mid X_{-\infty}^{-1},\theta_j\big)}\right]\leq\epsilon/4.$$

This in turn implies that $E_{\bar{P}_*^\nu}[\bar{l}_t(X_{-\infty}^t)]\leq\epsilon/2$. This result and the ergodic theorem establish that $\lim_{T\to\infty}T^{-1}\sum_{t=1}^T\bar{l}_t(X_{-\infty}^t)\leq\epsilon/2$ a.s.-$\bar{P}_*^\nu$. This implies the result in (11).

Part (ii). Follows directly from the ergodic theorem and the fact that $X_{-\infty}^\infty\mapsto\log p^\nu(X_t\mid X_{-\infty}^{t-1},\theta_*)$ is in $L^1(\bar{P}_*^\nu)$. *Q.E.D.*

SM.3. PROPERTIES OF $p_\theta(X_1|X_{-\infty}^0)$

For any $t\in\mathbb{N}_0\equiv\mathbb{N}\cup\{0\}$, any $X_{-\infty}^t$ and any $\theta\in\Theta$, $p^\nu(X_t|X_{-\infty}^{t-1},\theta)$ is defined as $\liminf_{M\to\infty}p_\theta^\nu(X_t|X_{-M}^{t-1})$; $p_*^\nu(X_t|X_{-\infty}^{t-1})$ is defined analogously.

LEMMA 12: *Suppose Assumptions* 4(ii) *and* 5 *hold*. *Then*:

(1) *For any* $t\in\mathbb{N}_0$, $x\mapsto p^\nu(\cdot|X_{-\infty}^{t-1},\theta)$ *and* $x\mapsto p_*^\nu(\cdot|X_{-\infty}^{t-1})$ *are densities, a.s.-*$\bar{P}_*^\nu$.

(2) *Suppose* Θ *is compact and that for each* $n\in\mathbb{N}_0$, $\theta\mapsto p_\theta^\nu(X_1|X_{-n}^0)$ *is uniformly continuous a.s.-*$\bar{P}_*^\nu$. *Then, for any* $\theta_0\in\Theta$ *and* $\epsilon>0$, *there exists a* $\delta>0$ *such that*

$$\bar{P}_*^\nu\Big(\sup_{\theta_0\in\Theta}\sup_{\theta\in B(\theta_0,\delta)}\big|p^\nu\big(X_1|X_{-\infty}^0,\theta\big)-p^\nu\big(X_1|X_{-\infty}^0,\theta_0\big)\big|>\epsilon\Big)<\epsilon.$$

(3) *Suppose* Θ *is compact and that for each* $n\in\mathbb{N}_0$, $\theta\mapsto p_\theta^\nu(X_1|X_{-n}^0)$ *is uniformly continuous a.s.-*$\bar{P}_*^\nu$. *Suppose also that there exists functions* $(x_1,x_0)\mapsto(\overline{p}(x_0,x_1),\underline{p}(x_0,x_1))$

such that for any $p \in \{p_\theta : \theta \in \Theta\} \cup p_*$, $\underline{p}(x_0, x_1) \leq p(x_0, s, x_1) \leq \overline{p}(x_0, x_1)$ *for all* $s \in \mathbb{S}$, *and* $E_{\bar{P}^\nu_*}[\overline{p}(X_0, X_1)/\underline{p}(X_0, X_1)] < \infty$. *Then Assumptions* 3 *and* 4 *hold*.

PROOF: (1) We need to show that the functions integrate to 1. By analogous steps to those in the proof of Lemma 11, it follows that

$$\begin{aligned} &\left| \int \{ p^\nu(x|X^{t-1}_{-\infty}, \theta) - p^\nu_\theta(x|X^{t-1}_{-n}) \} \, dx \right| \\ &\quad \leq \int \sum_s p_\theta(X_{t-1}, s, x) \, dx \limsup_{M \to \infty} \prod_{i=-n}^{t-1} (1 - \underline{q}(X_i)) \, dx \\ &\quad = \prod_{i=-n}^{t-1} (1 - \underline{q}(X_i)) |\mathbb{S}|. \end{aligned}$$

Since this holds for any n such that $-n \leq t-1$, we can take averages and obtain, for any $\epsilon > 0$, that

$$\begin{aligned} &\bar{P}^\nu_* \left(\frac{1}{M+1} \sum_{n=0}^{M} \left| \int p^\nu(x|X^{t-1}_{-\infty}, \theta) \, dx - 1 \right| > \epsilon \right) \\ &\quad \leq \bar{P}^\nu_* \left(\frac{1}{M+1} \sum_{n=0}^{M} \prod_{i=-n}^{0} (1 - \underline{q}(X_i)) |\mathbb{S}| > \epsilon \right). \end{aligned}$$

Since this holds for any M, by taking $M \to \infty$, stationarity and Assumption 5 imply that the RHS vanishes. Thus, it follows that $\bar{P}^\nu_*(|\int p^\nu(x|X^{t-1}_{-\infty}, \theta) \, dx - 1| \geq \epsilon) = 0$. As the $\epsilon > 0$ is arbitrary, this implies that $\int p^\nu(x|X^{t-1}_{-\infty}, \theta) \, dx = 1$, a.s.-$\bar{P}^\nu_*$. Following the same logic, an analogous result can be established for $p^\nu_*(\cdot|X^{t-1}_{-\infty})$.

(2) Similarly, for any n such that $-n \leq t-1$,

$$\sup_{\theta \in \Theta} \left| p^\nu(X_t|X^{t-1}_{-\infty}, \theta) - p^\nu_\theta(X_t|X^{t-1}_{-n}) \right| \leq \sum_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t) \prod_{i=-n}^{t-1} (1 - \underline{q}(X_i)).$$

Hence, for any θ_0 in Θ,

$$\begin{aligned} &\sup_{\theta \in B(\theta_0, \delta)} \left| p^\nu(X_t|X^{t-1}_{-\infty}, \theta) - p^\nu(X_t|X^{t-1}_{-\infty}, \theta_0) \right| \\ &\quad \leq \sup_{\theta \in B(\theta_0, \delta)} \sum_{s \in \mathbb{S}} p_\theta(X_{t-1}, s, X_t) \frac{1}{1+M} \sum_{n=0}^{M} \prod_{i=-n}^{t-1} (1 - \underline{q}(X_i)) \\ &\qquad + \frac{1}{1+M} \sum_{n=0}^{M} \sup_{\theta \in B(\theta_0, \delta)} \left| p^\nu_\theta(X_t|X^{t-1}_{-n}, \theta) - p^\nu_{\theta_0}(X_t|X^{t-1}_{-n}) \right|. \end{aligned}$$

For any $\gamma > 0$, choose M such that $\bar{P}^{\nu}_{*}(\frac{1}{1+M}\sum_{n=0}^{M}\prod_{i=-n}^{t-1}(1-\underline{q}(X_i)) \geq \gamma) < \gamma$; such M exists by Assumption 5. Given such M, for any $\epsilon > 0$ and any $\delta > 0$, it follows that

$$\begin{aligned}
&\bar{P}^{\nu}_{*}\Big(\sup_{\theta_0\in\Theta}\sup_{\theta\in B(\theta_0,\delta)}\big|p^{\nu}\big(X_t|X^{t-1}_{-\infty},\theta\big)-p^{\nu}\big(X_t|X^{t-1}_{-\infty},\theta_0\big)\big|>\epsilon\Big)\\
&\quad\leq \bar{P}^{\nu}_{*}\left(\sup_{\theta\in\Theta}\sum_{s\in\mathbb{S}}p_{\theta}(X_{t-1},s,X_t)\geq 0.5\epsilon/\gamma\right)+\gamma/3\\
&\qquad+\bar{P}^{\nu}_{*}\left(\frac{1}{1+M}\sum_{n=0}^{M}\sup_{\theta_0\in\Theta}\sup_{\theta\in B(\theta_0,\delta)}\big|p^{\nu}_{\theta}\big(X_t|X^{t-1}_{-n}\big)-p^{\nu}_{\theta_0}\big(X_t|X^{t-1}_{-n}\big)\big|\geq 0.5\epsilon\right).
\end{aligned}$$

Let $\delta > 0$ be such that for each $n \in \{0,\dots,M\}$, $\sup_{\theta_0\in\Theta}\sup_{\theta\in B(\theta_0,\delta)}|p^{\nu}_{\theta}(X_t|X^{t-1}_{-n}) - p^{\nu}_{\theta_0}(X_t|X^{t-1}_{-n})| < 0.5\epsilon$ a.s.-$\bar{P}^{\nu}_{*}$; such $\delta > 0$ exists by our conditions. So the third term in the RHS is 0. Also, under our conditions, it follows that $\sup_{\theta\in\Theta} p_{\theta}(X_{t-1}, s, X_t) = O_{\bar{P}^{\nu}_{*}}(1)$, and thus the first term in the RHS can be made smaller than $\epsilon/3$ by a suitably chosen $\gamma \in (0,\epsilon)$. Thus, the desired result holds.

(3) By the definition of $p^{\nu}(.|.,\theta)$, it follows that $p^{\nu}(X_1|X^0_{-\infty},\theta) \leq \max_{s\in\mathbb{S}} p_{\theta}(X_0,s,X_1)$, and for some $n \in \mathbb{N}_0$,

$$p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)\geq 0.5\sum_{s\in\mathbb{S}}p_{\theta}(X_0,s,X_1)\Pr_{\theta}\big(S_1=s\mid X^0_{-n}\big)\geq 0.5\min_{s\in\mathbb{S}}p_{\theta}(X_0,s,X_1).$$

Thus, by our conditions, for any $\delta > 0$ and $\theta_0 \in \Theta$,

$$\sup_{\theta\in B(\delta,\theta_0)}\frac{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}{p^{\nu}\big(X_1|X^0_{-\infty},\theta_0\big)}\leq\frac{\overline{p}(X_0,X_1)}{\underline{p}(X_0,X_1)},$$

and the RHS is in $L^1(\bar{P}^{\nu}_{*})$. Thus, by part (2) and the dominated convergence theorem, for any $\epsilon > 0$, there exists a $\delta > 0$ such that

$$E_{\bar{P}^{\nu}_{*}}\left[\sup_{\theta\in B(\theta_0,\delta)}\frac{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}{p^{\nu}\big(X_1|X^0_{-\infty},\theta_0\big)}\right]\leq 1+\epsilon,$$

for any $\theta_0 \in \Theta$. This readily implies Assumption 4(i). Part (ii) also follows with $(x_0, x_1) \mapsto C(x_0, x_1) = \frac{\overline{p}(x_0,x_1)}{\underline{p}(x_0,x_1)}$. Similarly, by noting that

$$1-\frac{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}{p^{\nu}_{*}\big(X_1|X^0_{-\infty}\big)}\leq\log\frac{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}{p^{\nu}_{*}\big(X_1|X^0_{-\infty}\big)}\leq\frac{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}{p^{\nu}_{*}\big(X_1|X^0_{-\infty}\big)}-1,$$

it follows that, for any $\theta \in \Theta$,

$$\left|\log\frac{p^{\nu}_{*}\big(X_1|X^0_{-\infty}\big)}{p^{\nu}\big(X_1|X^0_{-\infty},\theta\big)}\right|\leq 1+\frac{\overline{p}(X_0,X_1)}{\underline{p}(X_0,X_1)}.$$

Since the RHS is in $L^1(\bar{P}^{\nu}_{*})$, the results in part (2) and the dominated convergence theorem imply that $\theta \mapsto H(\theta)$ is continuous. *Q.E.D.*

SM.4. SUFFICIENT CONDITIONS FOR ASSUMPTIONS 5 AND 8

By exploiting the fact that $(X_t)_{t=-\infty}^{\infty}$ is β-mixing and stationary, the following lemma provides sufficient conditions for Assumption 5.

LEMMA 13: *Suppose Assumptions* 1 *and* 2 *hold.*
(1) *Suppose further there exists* $l \geq 1$ *such that*

$$E_{\bar{P}^*_\nu}\big[\big(1-\underline{q}(X_1)\big)^{l'}\big] < 1 \text{ and } E_{\bar{P}^*_\nu}\big[C(X_1, X_0)^l\big] < \infty,$$

where $1/l' + 1/l = 1$. *Then* $\lim_{T\to\infty} E_{\bar{P}^*_\nu}[T^{-1}\sum_{t=1}^{T} C(X_{t-1}, X_t)\prod_{i=0}^{t-1}(1-\underline{q}(X_i))] = 0$. [23]
(2) *Suppose further* $E_{\bar{P}^*_\nu}[(1-\underline{q}(X_1))^{\frac{2a}{1-a}}] < 1$. *Then Assumption* 8 *holds.*

PROOF: (1) By stationarity and the condition that $E[C(X_{-1}, X_0)^l] < \infty$, by Jensen's inequality (and the fact that $1/l' \leq 1$, where $1/l + 1/l' = 1$), it suffices to show that

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E_{\bar{P}^*_\nu}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{l'}\right] = 0.$$

Note that, for any $1 \leq m \leq T$,

$$T^{-1}\sum_{t=1}^{T} E_{\bar{P}^*_\nu}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{l'}\right] \leq \frac{m}{T} + T^{-1}\sum_{t=m+1}^{T} E_{\bar{P}^*_\nu}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{l'}\right].$$

By employing well-known coupling results for β-mixing sequences (see Yu (1994)), it follows that[24]

$$E_{\bar{P}^*_\nu}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{l'}\right] \leq \beta_q t/q + \big(E_{\bar{P}^*_\nu}\big[\big(1-\underline{q}(X_1)\big)^{l'}\big]\big)^{(t/q-1)/2}.$$

for any $q \in \{1, \dots, t\}$. Therefore, for any $1 \leq m \leq T$,

$$T^{-1}\sum_{t=1}^{T} E_{\bar{P}^*_\nu}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{l'}\right] \leq \frac{m}{T} + T^{-1}\sum_{t=m+1}^{T}\big\{\beta_q t/q + \big(E_{\bar{P}^*_\nu}\big[\big(1-\underline{q}(X_1)\big)^{l'}\big]\big)^{(t/q-1)/2}\big\}.$$

By Lemma 1, $\beta_q = \exp\{q\log\gamma\}$. This fact and the condition $E_{\bar{P}^*_\nu}[(1-\underline{q}(X_1))^{l'}] < 1$ imply that we can take, for instance, $q \equiv t^{1/2}$ and $m = \sqrt{T}$, so that the RHS vanishes as T diverges.
(2) By our previous calculations,

$$E_{\bar{P}^\nu_*}\left[\prod_{i=0}^{t-1}\big(1-\underline{q}(X_i)\big)^{\frac{2a}{1-a}}\right] \leq \beta_q t/q + \big(E_{\bar{P}^*_\nu}\big[\big(1-\underline{q}(X_1)\big)^{\frac{2a}{1-a}}\big]\big)^{(t/q-1)/2}.$$

So the result follows because for $q = \sqrt{t}$ the RHS is the sum of terms in ℓ_1.[25] *Q.E.D.*

[23]Clearly, by the Markov inequality, Part (1) implies that $\lim_{T\to\infty}\bar{P}^*_\nu(T^{-1}\sum_{t=1}^{T} C(X_{t-1}, X_t)\prod_{i=0}^{t-1}(1-\underline{q}(X_i)) \geq \epsilon) = 0$.

[24]For the complete derivations, we refer the reader to Pouzo, Psaradakis, and Sola (2021).

[25]For the complete derivations, we refer the reader to Pouzo, Psaradakis, and Sola (2021).

SM.5. PROOFS AND RESULTS FOR EXAMPLE 4

In what follows, $e_{\max}(M)$ and $e_{\min}(M)$ denote the maximal and minimal eigenvalues of a matrix M.

LEMMA 14: *Assumption* 2 *holds*.

PROOF: For each $s \in \mathbb{S}$, we apply Theorem 3.3 in Douc, Fort, Moulines, and Soulier (2004). To do so, we first verify their Assumptions 3.3 and 3.4. In our case, $\epsilon \sim N(0, \Sigma(s))$, so their Assumption 3.3 is satisfied for any z_0 and $\gamma_0 = 1$. In their notation, $g(x) \equiv \mu(s) + \Phi^\intercal x$. Observe that $\|g(x)\| \leq \|\mu(s)\| + e_{\max}(\Phi\Phi^\intercal)\|x\|$. By assumption, $e_{\max}(\Phi\Phi^\intercal) \equiv \gamma < 1$ and

$$\left\|\mu(s)\right\| \leq \left(1 - e_{\max}\left(\Phi\Phi^\intercal\right)\right)\|x\|\left(1 - \|x\|^{-0.5}\right)$$

for all x such that $\|x\| \geq R_0$. Such an R_0 exists because $\|\mu(s)\|$ is bounded and $e_{\max}(\Phi\Phi^\intercal) < 1$. This choice ensures that $\|g(x)\| \leq \|x\|(1 - \|x\|^{0.5})$, which in turn, ensures the validity of their Assumption 3.4 with $r = 1$ and $\rho = 0.5$. By their Theorem 3.3, Assumption 2(i) holds. Assumption 2(ii), (iii) is satisfied because $\inf_{x\in A} P_*(s, x, C) \geq \int_C \inf_{x\in A} \underline{p}(x, a)da$, and since A is bounded, it follows that $\inf_{x\in A} \underline{p}(x, a) \geq \exp\{\tilde{D} + (x')^T \tilde{F}a + \tilde{G}a\}$, so the RHS plays the role of the measure ϖ, which clearly is such that $\varpi(A) > 0$. *Q.E.D.*

Let $\overline{\kappa} \equiv \max_{s\in\mathbb{S}} \sup_{\Sigma(s)\in\Theta} e_{\max}(\Sigma(s))$ and $\underline{\kappa} \equiv \min_{s\in\mathbb{S}} \inf_{\Sigma(s)\in\Theta} e_{\min}(\Sigma(s))$; $\overline{m} = \max_{s\in\mathbb{S}} \sup_{\mu(s)\in\Theta} \|\mu(s)\|$; $\overline{M} = \sup_{\Phi\in\Theta} e_{\max}(\Phi\Phi^\intercal)$ and $\underline{M} = \min_{\Phi\in\Theta} e_{\min}(\Phi\Phi^\intercal)$; $\overline{\kappa}_* \equiv \max_{s\in\mathbb{S}} e_{\max}(\Sigma_*(s))$ and $\underline{\kappa}_* \equiv \min_{s\in\mathbb{S}} e_{\min}(\Sigma_*(s))$; $\overline{m}_* = \max_{s\in\mathbb{S}} \|\mu_*(s)\|$; $\overline{M}_* = e_{\max}(\Phi_*\Phi_*^\intercal)$ and $\underline{M}_* = e_{\min}(\Phi_*\Phi_*^\intercal)$. By the assumptions in the text, $\overline{\kappa}$, $\overline{\kappa}_*$, $\underline{\kappa}$, $\underline{\kappa}_*$, $\underline{M}$, $\overline{M}$, $\underline{M}_*$, $\overline{M}_*$ are all in $(0, \infty)$.

LEMMA 15: *There exists a* $C \in [1, \infty)$ *such that, for any* (x, y) *and any* s,

$$\begin{aligned}
&f_{\mathcal{N}}\left(\left(y - \Phi^\intercal x - \mu(s)\right)\Sigma^{-1/2}(s)\right) \\
&\quad \leq C \exp\left\{-0.5\underline{\kappa}\left(\|y\|^2 + \underline{M}\|x\|^2 - 2\sqrt{\overline{M}}\|x\|\|y\|\right) + \underline{\kappa}\left(\|y\| + \sqrt{\overline{M}}\|x\|\right)\overline{m}\right\}, \\
&f_{\mathcal{N}}\left(\left(y - \Phi^\intercal x - \mu(s)\right)\Sigma^{-1/2}(s)\right) \\
&\quad \geq C^{-1} \exp\left\{-0.5\overline{\kappa}\left(\|y\|^2 + \overline{M}\|x\|^2 + 2\sqrt{\overline{M}}\|x\|\|y\|\right) - \overline{\kappa}\left(\|y\| + \sqrt{\overline{M}}\|x\|\right)\overline{m}\right\}.
\end{aligned}$$

An analogous bound holds for $f_{\mathcal{N}}((y - \Phi_*^\intercal x - \mu_*(s))\Sigma_*^{-1/2}(s))$.

PROOF: The proof involves lengthy but straightforward derivations and is thus omitted; for details, see Pouzo, Psaradakis, and Sola (2021). *Q.E.D.*

LEMMA 16: *Suppose there exists a* $l \geq 1$ *such that* $l\overline{\kappa} < l\underline{\kappa} + \underline{\kappa}_*$ *and* $l\overline{\kappa}\overline{M} < l\underline{\kappa}\underline{M} + \underline{\kappa}_*$. *Then*

$$\begin{aligned}
&E_{\bar{P}^*_\nu}\left[\exp\left\{-0.5l\left((b_1 - a_1)\|Y\|^2 + (b_2 - a_2)\|X\|^2 - 2(b_3 + a_3)\|X\|\|Y\|\right)\right.\right. \\
&\qquad \left.\left. + l(b_4 + a_4)\|Y\| + l(b_5 + a_5)\|X\|\right\}\right] < \infty,
\end{aligned}$$

where

$$a_1 = b_4 = \overline{\kappa}, \qquad b_1 = a_4 = \underline{\kappa}, \qquad a_2 = \overline{\kappa}\overline{M}, \qquad b_2 = \underline{\kappa M},$$
$$a_3 = b_5 = \overline{\kappa}\sqrt{\overline{M}}, \qquad b_3 = a_5\underline{\kappa}\sqrt{\overline{M}}.$$

REMARK 1: Before going to the proof, we discuss the conditions in the lemma. They basically require that the "spread" of the eigenvalues of the matrices $\Sigma(\cdot)$ and $\Phi\Phi^*$ is not too large relative to the eigenvalues in $\Sigma_*(\cdot)$. This condition comes naturally since we are essentially requiring that the ratio of two exponential functions is integrable with respect to a Gaussian measure. For instance, if $\Sigma(\cdot)$, $\Sigma_*(\cdot)$ and $\Phi\Phi^\intercal$, $\Phi_*\Phi_*^\intercal$ are matrices with eigenvalues bounded between $0 < a$ and $a + \Delta$, then sufficient conditions are given by $l\Delta < a$ and $l(2a\Delta + (\Delta)^2) < a$, which is equivalent to $\frac{l\Delta^2}{1-2l\Delta} < a$. △

PROOF: It is enough to show that

$$\begin{aligned}
E_{\bar{P}_\nu^*}\big[T_1(Y)\big] &\equiv E_{\bar{P}_\nu^*}\big[\exp\big\{-0.5l(b_1 - a_1)\|Y\|^2 + l(b_4 + a_4)\|Y\|\big\}\big] < \infty,\\
E_{\bar{P}_\nu^*}\big[T_2(X)\big] &\equiv E_{\bar{P}_\nu^*}\big[\exp\big\{-0.5l(b_2 - a_2)\|X\|^2 + l(b_5 + a_5)\|X\|\big\}\big] < \infty,\\
E_{\bar{P}_\nu^*}\big[T_3(X, Y)\big] &\equiv E_{\bar{P}_\nu^*}\big[\exp\big\{l(b_3 + a_3)\|X\|\|Y\|)\big\}\big] < \infty.
\end{aligned}$$

For any $d \in \{1, 2\}$, suppose there exists φ such that $\int T_d(b)p_*(x, s, b)\,db \le \varphi(x)T_d(x)$ for any x, and, for any $\gamma > 0$, $\{x : \varphi(x) \ge \gamma\}$ is either empty or compact. Then, for any $\gamma > 0$,

$$\begin{aligned}
\int T_d(x)\nu(dx) &= \int 1\big\{\varphi(x) \le \gamma\big\}T_d(x)\nu(dx) + \int 1\big\{\varphi(x) > \gamma\big\}T_d(x)\nu(dx)\\
&= \int\int 1\big\{\varphi(x) \le \gamma\big\}T_d(b)p_*(x, s, b)\,db\nu(dx, ds)\\
&\quad + \int 1\big\{\varphi(x) > \gamma\big\}T_d(x)\nu(dx)\\
&\le \gamma \int T_d(x)\nu(dx) + \sup_{x:\varphi(x)\ge\gamma} T_d(x),
\end{aligned}$$

where the second line follows because ν is the invariant probability distribution. Since $\{x : \varphi(x) \ge \gamma\}$ is bounded and compact (if it is nonempty), $\sup_{x:\varphi(x)\ge\gamma} T_d(x) \le M < \infty$. Choosing $\gamma < 1$, it follows that $\int T_d(x)\nu(dx) \le \frac{M}{1-\gamma} < \infty$, as desired.

We now show that $\int T_1(b)p_*(x, s, b)\,db \le \varphi(x)T_1(x)$ for any x. To do this, note that

$$\int T_1(b)p_*(x, s, b)\,db \le \int \exp\big\{0.5l(\overline{\kappa} - \underline{\kappa})\|y\|^2 + l(\overline{\kappa} + \underline{\kappa})\|y\|\big\}p_*(x, s, y)\,dy,$$

and, by Lemma 15,

$$f_{\mathcal{N}}\big((y - \Phi_*^\intercal x - \mu_*(s))\Sigma_*^{-1/2}(s)\big) \le \exp\big\{-0.5\underline{\kappa}_*\big(\|y\|^2 - 2\sqrt{\overline{M}_*}\|x\|\|y\|\big) + \underline{\kappa}_*\|y\|\overline{m}_*\big\}B_*(x),$$

with $B_*(x) \equiv C_* \exp\{-0.5\underline{\kappa}_*\underline{M}_*\|x\|^2 + \underline{\kappa}_*\sqrt{\overline{M}_*}\|x\|\overline{m}_*\}$. Therefore, $\frac{\int T_1(b)p_*(x,s,y)\,dy}{B_*(x)}$ is bounded by

$$\int \exp\{0.5(\overline{\kappa}-\underline{\kappa})\|y\|^2 + l(\overline{\kappa}+\underline{\kappa})\|y\|\}$$
$$\times \exp\{-0.5\underline{\kappa}_*(\|y\|^2 - 2\sqrt{\overline{M}_*}\|x\|\|y\|) + \underline{\kappa}_*\overline{m}_*\|y\|\}\,dy$$
$$= \int \exp\{0.5(l(\overline{\kappa}-\underline{\kappa}) - \underline{\kappa}_*)\|y\|^2 + \underline{\kappa}_*\sqrt{\overline{M}_*}\|x\|\|y\| + (\underline{\kappa}_*\overline{m}_* + l(\overline{\kappa}+\underline{\kappa}))\|y\|\}\,dy.$$

By our conditions, $l(\overline{\kappa}-\underline{\kappa}) - \underline{\kappa}_* < 0$. Hence, the expression above is an integral of an exponential function of a quadratic form with negative leading coefficient, and is thus finite. Moreover, after some algebra, there exists a finite constant C, such that $\int T_1(b)p_*(x,s,y)\,dy \le CB_*(x)\exp\{\underline{\kappa}_*\sqrt{\overline{M}_*}\|x\|/(-(l(\overline{\kappa}-\underline{\kappa}) - \underline{\kappa}_*))\}$); we redefine the RHS as $CB_*(x)\exp\{D\underline{\kappa}_*\sqrt{\overline{M}_*}\|x\|\}$) with $D > 0$. Therefore, the result holds with

$$x \mapsto \varphi(x) \equiv \exp\{-0.5(\underline{\kappa}_*\underline{M}_* + l(\overline{\kappa}-\underline{\kappa}))\|x\|^2 + (\underline{\kappa}_*\sqrt{\overline{M}_*}(D+\overline{m}_*) + l(b_4+a_4))\|x\|\}.$$

As the coefficient on $\|x\|^2$ is $-0.5(\underline{\kappa}_*\underline{M}_* + l(\overline{\kappa}-\underline{\kappa}))$, which is negative, the function satisfies the required conditions.

The case for $d = 2$ is analogous and is thus omitted; for this case, we use the restriction that $l\overline{\kappa}\overline{M} < l\underline{\kappa}\underline{M} + \underline{\kappa}_*$, instead of $l(\overline{\kappa}-\underline{\kappa}) - \underline{\kappa}_* < 0$.

Finally, observe that, if $\int T_3(x,y)p_*(x,s,y)\nu(dx,ds)\,dy \le C\int \exp\{c_1\|x\|\}\nu(dx)$ for some $c_1 < \infty$, then we can follow the same approach as before to show that $\int T_3(x,y) \times p_*(x,s,y)\nu(dx,ds)\,dy < \infty$. Observe that, by the same calculations as before,

$$\int T_3(x,y)p_*(x,s,y)\,dy$$
$$\le B_*(x)\int \exp\{l(a_3+b_3)\|x\|\|y\| - 0.5\underline{\kappa}(\|y\|^2 - 2\sqrt{\overline{M}}\|x\|\|y\|) + \underline{\kappa}\|y\|\overline{m}\}\,dy$$
$$\le CB_*(x)\exp\{(l(a_3+b_3) + \underline{\kappa}\sqrt{\overline{M}})\|x\|\}.$$

The RHS is of the form $C\exp\{c_1\|x\|\}$, as desired. *Q.E.D.*

SM.6. PROOFS OF SUPPLEMENTAL LEMMAS IN APPENDIX A.2

PROOF OF LEMMA 5: For any a, b in $\mathbb{S}$,

$$\bar{P}^{\nu}_{\theta}(S_{l+1} = b \mid S_l = a, X^j_{-m}) = \frac{\bar{P}^{\nu}_{\theta}(X^j_{l+1} \mid S_{l+1} = b, S_l = a, X^l_{-m})Q_\theta(X_l, a, b)}{\sum_{c\in\mathbb{S}} \bar{P}^{\nu}_{\theta}(X^j_{l+1} \mid S_{l+1} = c, S_l = a, X^l_{-m})Q_\theta(X_l, a, c)}.$$

The expression $\bar{P}^{\nu}_{\theta}(X^j_{l+1} \mid S_{l+1} = b, S_l = a, X^l_{-m})$ equals $\bar{P}^{\nu}_{\theta}(X^j_{l+1} \mid S_{l+1} = b, X_l)$ because of the Markov property. The latter probability depends on the transitions of X_{t+1} given (X_t, S_{t+1}) and S_{t+1} given (X_t, S_t) for each $t \ge l+1$. Since these are the same for the process with $i = 1$ and $i = 2$ and the "original" process" $(S_t, X_t)^{\infty}_{t=-m}$, the last line of the previous display equals $\Pr_\theta(\eta_{1,l+1}=b \mid \eta_{1,l}=a, X^j_{-m}) = \Pr_\theta(\eta_{2,l+1}=b \mid \eta_{2,l}=a, X^j_{-m})$, as desired. *Q.E.D.*

PROOF OF LEMMA 6: Throughout this proof, we omit the dependence on θ in the probability terms and on other quantities. For any a, c in $\mathbb{S}$,

$$\begin{aligned}
&\big\|\Pr(\eta_{1,l+1}=\cdot \mid \eta_{1,l}=a, X^j_{-m}) - \Pr(\eta_{2,l+1}=\cdot \mid \eta_{1,l}=c, X^j_{-m})\big\|_1 \\
&\quad \le \big\|\Pr(\eta_{1,l+1}=\cdot, v_{1,l}=0 \mid \eta_{1,l}=a, X^j_{-m}) - \Pr(\eta_{2,l+1}=\cdot, v_{2,l}=0 \mid \eta_{2,l}=c, X^j_{-m})\big\|_1 \\
&\qquad + \big\|\Pr(\eta_{1,l+1}=\cdot, v_{1,l}=1 \mid \eta_{1,l}=a, X^j_{-m}) - \Pr(\eta_{2,l+1}=\cdot, v_{2,l}=1 \mid \eta_{2,l}=c, X^j_{-m})\big\|_1 \\
&\quad \equiv \text{Term}_1 + \text{Term}_2
\end{aligned}$$

To bound the second term, note that

$$\begin{aligned}
\Pr(\eta_{1,l+1}=\cdot, v_{1,l}=1 \mid \eta_{1,l}=a, X^j_{-m}) &= \Pr(\eta_{1,l+1}=\cdot \mid v_{1,l}=1, \eta_{1,l}=a, X^j_{-m}) \\
&\quad \times \Pr(v_{1,l}=1 \mid \eta_{1,l}=a, X^j_{-m}).
\end{aligned}$$

It follows that $\Pr(v_{1,l}=1 \mid \eta_{1,l}=a, X^j_{-m}) = \underline{q}(X_l)$, because given X^j_{-m}, $v_{1,l}$ is drawn independently according to a probability function that only depends on X_l (in particular, it does not depend on $\eta_{1,l}$), and is given by $\underline{q}(X_l)$. By some algebra, the Markov property, and the fact that, given $v_{1,l}=1$ and X^l_m, the random variable $\eta_{1,l+1}$ is independent of its past, it follows that $a \mapsto \Pr(\eta_{1,l+1}=\cdot \mid v_{1,l}=1, \eta_{1,l}=a, X^j_{-m})$ is constant (i.e., does not depend on $\eta_{1,l}=a$). Thus, $a \mapsto \Pr(\eta_{1,l+1}=\cdot, v_{1,l}=1 \mid \eta_{1,l}=a, X^j_{-m})$ is constant (i.e., does not depend on the value of a); since one can obtain the exact result for $c \mapsto \Pr(\eta_{2,l+1}=\cdot, v_{2,l}=1 \mid \eta_{2,l}=c, X^j_{-m})$ and, moreover, the laws for $i=1$ and $i=2$ coincide (see the proof of Lemma 5), it follows that $\text{Term}_2 = 0$.

To bound Term_1, it follows from the previous arguments that

$$\begin{aligned}
\text{Term}_1 = \sum_{s\in\mathbb{S}} \big|&\Pr(\eta_{1,l+1}=s \mid v_{1,l}=0, \eta_{1,l}=a, X^j_{-m}) \\
&- \Pr(\eta_{2,l+1}=s \mid v_{2,l}=0, \eta_{2,l}=c, X^j_{-m})\big| \\
&\times (1-\underline{q}(X_l)) \le 2(1-\underline{q}(X_l)),
\end{aligned}$$

and thus the desired result follows. *Q.E.D.*

SM.7. PROOFS OF SUPPLEMENTARY LEMMAS IN APPENDIX A.3

SM.7.1. *Proofs of Lemmas 7 and 8*

In this section, we provide the proofs of Lemmas 7 and 8. To do this, we use a series of lemmas, which we state below (their proofs are relegated to the end of this section).

Henceforth, for any $j \ge m$, let

$$\varrho(j,m) \equiv \left(E_{\bar{P}^\nu_*}\left[\prod_{i=m}^{j}(1-\underline{q}(X_i))^{\frac{2a}{1-a}}\right]\right)^{\frac{1-a}{2a}}, \tag{15}$$

where the constant a is the same as in Assumption 7. We also introduce the following notation: for any $\theta \in \Theta$, $(x', x, s) \mapsto \Gamma(x'|x, s; \theta) \equiv \nabla_\theta \log p_\theta(x, s, x')$ and $(s, x, s) \mapsto \Lambda(s'|s, x; \theta) \equiv \nabla_\theta \log Q_\theta(x, s, s')$. Furthermore, for any $k \geq n$ and any $l \geq m$, let

$$\Phi_\theta(k, n, l, m) \equiv E_{\bar{P}^\nu_\theta}\left[\sum_{j=n}^{k} \Gamma(X_j|X_{j-1}, S_j; \theta) \mid X^l_m\right],$$

$$\text{and} \quad \Psi_\theta(k, n, l, m) \equiv E_{\bar{P}^\nu_\theta}\left[\sum_{j=n}^{k} \Lambda(S_j|S_{j-1}, X_{j-1}; \theta) \mid X^l_m\right].$$

To state the first lemma, for any k, T, and X^k_{k-T} and any θ, let

$$\begin{aligned} &\Delta_{k,k-T}(\theta)\left(X^k_{k-T}\right) \\ &\quad \equiv \Phi_\theta(k-1, k-T, k, k-T) + \Psi_\theta(k-1, k-T-1, k, k-T) \\ &\qquad - \Phi_\theta(k-1, k-T, k-1, k-T) - \Psi_\theta(k-1, k-T-1, k-1, k-T) \\ &\qquad + \Phi_\theta(k, k, k, k-T) + \Psi_\theta(k, k, k, k-T) \end{aligned} \tag{16}$$

The next lemma is analogous to the results in Douc, Moulines, and Rydén (2004) and Bickel, Ritov, and Rydén (1998), thus the proof is omitted; it can be found in Pouzo, Psaradakis, and Sola (2021).

LEMMA 17: *Suppose Assumption* 6 *holds. Then, for any* $k, T \geq 0$ *and any* $\theta \in \Theta$,[26]

$$\nabla_\theta \log p^\nu_k\left(X_k|X^{k-1}_{k-T}; \theta\right) = \Delta_{k,k-T}(\theta)\left(X^k_{k-T}\right) \quad a.s.\text{-}\bar{P}^\nu_*.$$

This lemma characterizes the asymptotic behavior of the score functions; in particular, it shows that they are well approximated by $(\Delta_{t,-\infty}(\theta_*))_t$, which is to be defined below, but at this stage is worth pointing out that it is stationary and ergodic; this last fact is established in Lemma 19 below.

LEMMA 18: *Suppose Assumptions* 1, 2, 6, 7(i), *and* 8 *hold. Then*:
(i) *There exists a finite constant* $C > 0$ *such that for any* k *and* $T \geq 0$,

$$\begin{aligned} &\left\|\Delta_{k,k-T}(\theta_*) - \Delta_{k,-\infty}(\theta_*)\right\|_{L^2(\bar{P}^\nu_*)} \\ &\quad \leq C\left(\max\left\{\sum_{j=[k-T/2]}^{k-1} \varrho(j, k-T), \sum_{j=k-T}^{[k-T/2]-1} \varrho(k-1, j)\right\}\right); \end{aligned}$$

(ii) $\lim_{T\to\infty} \|T^{-1/2}\sum_{t=0}^{T}\{\Delta_{t,-\infty}(\theta_*) - \nabla_\theta \log p^\nu_t(\cdot|\cdot, \theta_*)\}\|_{L^2(\bar{P}^\nu_*)} = 0$.

LEMMA 19: *Suppose Assumptions* 1, 2, *and* 5 *hold. Then* $(\Delta_{t,-\infty}(\theta_*))^\infty_{t=-\infty}$ *is a stationary and ergodic* $L^2(\bar{P}^\nu_*)$ *process* (*under* $\bar{P}^\nu_*$).

LEMMA 20: *Suppose Assumption* 1 *holds. Then there exists a finite constant* $L > 0$ *such that*:

[26] When there is no risk of confusion, we will omit the dependence of $\Delta_{k,k-T}$ on the data X^k_{k-T}.

(i) *For* $-m \leq j < k$ *and any* $\theta \in \Theta$, *a.s.*-$\bar{P}_*^\nu$,

$$\left\| \bar{P}_\theta^\nu\left(S_j = \cdot | X_{-m}^k\right) - \bar{P}_\theta^\nu\left(S_j = \cdot | X_{-m}^{k-1}\right) \right\|_1 \leq L \prod_{i=j}^{k-1} \left(1 - \underline{q}(X_i)\right).$$

(ii) *For* $-n \leq -m \leq j < k$ *and any* $\theta \in \Theta$, *a.s.*-$\bar{P}_*^\nu$,

$$\left\| \bar{P}_\theta^\nu\left(S_j = \cdot | X_{-m}^k\right) - \bar{P}_\theta^\nu\left(S_j = \cdot | X_{-n}^{k-1}\right) \right\|_1 \leq L \prod_{i=-m}^{j} \left(1 - \underline{q}(X_i)\right).$$

PROOF OF LEMMA 7: Follows directly from Lemmas 18 and 19. *Q.E.D.*

PROOF OF LEMMA 8: Lemma 8 is analogous to Lemma 10 in Bickel, Ritov, and Rydén (1998). The proof follows by their Lemma 9, which in turn holds by analogous steps to theirs and by invoking Lemma 20 (which is analogous to their Lemma 7). *Q.E.D.*

SM.7.1.1. *Proofs of Lemmas*

Throughout this section, in cases where the expectations are taken with respect to $\bar{P}_*^\nu$, we omit the probability from the notation and simply use $E[\cdot]$.

The proof of Lemma 18 requires the following lemma.

LEMMA 21: *Suppose that Assumptions* 1 *and* 7(i) *hold*. *Then there exists a finite constant* $C > 0$ *such that*:

(i) *for any* $-n \leq -m \leq -m' \leq l \leq k$,

$$\left\| \Phi_{\theta_*}\left(l, -m', k, -m\right) - \Phi_{\theta_*}\left(l, -m', k, -n\right) \right\|_{L^2(\bar{P}_*^\nu)} \leq C\left(\sum_{j=-m'}^{l} \varrho(j, -m) \right);$$

(ii) *for any* $-m \leq -m' < l \leq k-1$,

$$\left\| \Phi_{\theta_*}\left(l, -m', k, -m\right) - \Phi_{\theta_*}\left(l, -m', k-1, -m\right) \right\|_{L^2(\bar{P}_*^\nu)} \leq C\left(\sum_{j=-m'}^{l} \varrho(k-1, j) \right);$$

(iii) *for any* $-n \leq -m \leq -m' < l \leq k$,

$$\left\| \Psi_{\theta_*}\left(l, -m', k, -m\right) - \Psi_{\theta_*}\left(l, -m', k, -n\right) \right\|_{L^2(\bar{P}_*^\nu)} \leq C\left(\sum_{j=-m'}^{l} \varrho(j-1, -m) \right);$$

(iv) *for any* $-m \leq -m' < l \leq k-1$,

$$\left\| \Psi_{\theta_*}\left(l, -m', k, -m\right) - \Psi_{\theta_*}\left(l, -m', k-1, -m\right) \right\|_{L^2(\bar{P}_*^\nu)} \leq C\left(\sum_{j=-m'}^{l} \varrho(k-1, j) \right).$$

PROOF OF LEMMA 21: Here, we only prove part (i) as the calculations for the rest of the parts are analogous.[27]

Throughout the proof, we omit the dependence of $E[\cdot]$ on $\bar{P}^{\nu}_{\theta_*}$. Also, let L denote the constant in Lemma 22.

Part (i). Observe that, for any $j \leq k$,

$$\begin{aligned}
&\big\| E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-m}\big] - E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-n}\big]\big\| \\
&\quad = \left\| \sum_{a\in\mathbb{S}} \Gamma(X_j|X_{j-1}, a; \theta_*)\big\{\bar{P}^{\nu}_{\theta_*}\big(S_j = a \mid X^k_{-m}\big) - \bar{P}^{\nu}_{\theta_*}\big(S_j = a \mid X^k_{-n}\big)\big\}\right\| \\
&\quad \leq \max_{a\in\mathbb{S}} \big\|\Gamma(X_j|X_{j-1}, a; \theta_*)\big\| \big\|\bar{P}^{\nu}_{\theta_*}\big(S_j = \cdot \mid X^k_{-m}\big) - \bar{P}^{\nu}_{\theta_*}\big(S_j = \cdot \mid X^k_{-n}\big)\big\|_1.
\end{aligned}$$

By Lemma 20(ii),

$$\begin{aligned}
&\big\| E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-m}\big] - E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-n}\big]\big\| \\
&\quad \leq L \max_{a\in\mathbb{S}} \big\|\Gamma(X_j|X_{j-1}, a; \theta_*)\big\| \prod_{i=-m}^{j} \big(1 - \underline{q}(X_i)\big).
\end{aligned}$$

Thus, by the Hölder inequality, for $a^{-1} + b^{-1} = 1$ (with a as in Assumption 7),

$$\begin{aligned}
&\left\| \sum_{j=-m'}^{l} \big\{E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-m}\big] - E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-n}\big]\big\}\right\|_{L^2(\bar{P}^{\nu}_*)} \\
&\quad \leq \sum_{j=-m'}^{l} \big\| E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-m}\big] - E\big[\Gamma(X_j|X_{j-1}, S_j; \theta_*) \mid X^k_{-n}\big]\big\|_{L^2(\bar{P}^{\nu}_*)} \\
&\quad \leq L\left(\sum_{a\in\mathbb{S}} E_{\bar{P}^{\nu}_*}\big[\big\|\Gamma(X_1|X_0, a; \theta_*)\big\|^{2a}\big]\right)^{1/(2a)} \sum_{j=-m'}^{l} \left(E_{\bar{P}^{\nu}_*}\left[\prod_{i=-m}^{j} \big(1 - \underline{q}(X_i)\big)^{2b}\right]\right)^{1/(2b)},
\end{aligned}$$

where the second line follows from the triangle inequality and the third follows from stationarity (Lemma 1). The fact that $\Gamma(X_1|X_0, a; \theta_*) = \nabla_\theta \log p_{\theta_*}(X_0, a, X_1)$, Assumption 7(i), and the definition of ϱ imply the desired result. *Q.E.D.*

PROOF OF LEMMA 18: Throughout the proof, we denote $\|\cdot\|_{L^2(\bar{P}^{\nu}_*)}$ as $\|\cdot\|_{L^2}$. Also, we use Φ and Ψ to denote Φ_{θ_*} and Ψ_{θ_*}, respectively.

Part (i): Observe that $\Phi(k-1, k-T, l, k-T) = \Phi(k-1, [k-T/2], l, k-T) + \Phi([k-T/2]-1, k-T, l, k-T)$ and an analogous result holds for Ψ. Therefore, by the definition of $\Delta_{k,k-T}$ and analogous calculations to those in Bickel, Ritov, and Rydén (1998, pp. 1624–1626),

$$\begin{aligned}
&\big\|\Delta_{k,k-T}(\theta_*) - \Delta_{k,-\infty}(\theta_*)\big\|_{L^2} \\
&\quad \leq \big\|\Phi\big(k-1, [k-T/2], k, k-T\big) - \Phi\big(k-1, [k-T/2], k, -\infty\big)\big\|_{L^2}
\end{aligned}$$

[27]For a complete proof, see Pouzo, Psaradakis, and Sola (2021).

$$
\begin{aligned}
&+ \big\| \Phi\big(k-1, [k-T/2], k-1, k-T\big) - \Phi\big(k-1, [k-T/2], k-1, -\infty\big) \big\|_{L^2} \\
&+ \big\| \Phi\big([k-T/2]-1, k-T, k, k-T\big) - \Phi\big([k-T/2]-1, k-T, k-1, k-T\big) \big\|_{L^2} \\
&+ \big\| \Psi\big(k-1, [k-T/2], k, k-T\big) - \Psi\big(k-1, [k-T/2], k, -\infty\big) \big\|_{L^2} \\
&+ \big\| \Psi\big(k-1, [k-T/2], k-1, k-T\big) - \Psi\big(k-1, [k-T/2], k-1, -\infty\big) \big\|_{L^2} \\
&+ \big\| \Psi\big([k-T/2]-1, k-T-1, k, k-T\big) \\
&- \Psi\big([k-T/2]-1, k-T-1, k-1, k-T\big) \big\|_{L^2} \\
&+ \big\| \Phi(k, k, k, k-T) - \Phi(k, k, k, -\infty) \big\|_{L^2} \\
&+ \big\| \Psi(k, k, k, k-T) - \Psi(k, k, k, -\infty) \big\|_{L^2} \\
&\equiv \sum_{i=1}^{8} \mathrm{Term}_i.
\end{aligned}
$$

Here, we only bound Term 1 as the bounds for the rest are analogous.[28] Observe that

$$
\begin{aligned}
&\big\| \Phi\big(k-1, [k-T/2], k, k-T\big) - \Phi\big(k-1, [k-T/2], k, -\infty\big) \big\|_{L^2} \\
&= \Bigg\| \sum_{j=[k-T/2]}^{k-1} E_{\bar{P}^{\nu}_{\theta_*}} \big[\Gamma(X_j \mid X_{j-1}, S_j; \theta_*) \mid X^m_{k-T} \big] \\
&\quad - \sum_{j=[k-T/2]}^{k-1} E_{\bar{P}^{\nu}_{\theta_*}} \big[\Gamma(X_j \mid X_{j-1}, S_j; \theta_*) \mid X^m_{-\infty} \big] \Bigg\|_{L^2}.
\end{aligned}
$$

By Lemma 21(i), for $i \in \{1, 2\}$, $\mathrm{Term}_i \precsim (\sum_{j=[k-T/2]}^{k-1} \varrho(j, k-T))$.

Part (ii). By part (i) and Lemma 17,

$$
\begin{aligned}
&\Bigg\| T^{-1/2} \sum_{t=1}^{T} \big\{ \Delta_{t,-\infty}(\theta_*) - \nabla_\theta \log p^{\nu}_t(\cdot|\cdot; \theta_*) \big\} \Bigg\|_{L^2} \\
&\leq T^{-1/2} \sum_{t=1}^{T} \big\| \big\{ \Delta_{t,-\infty}(\theta_*) - \Delta_{t,0}(\theta_*) \big\} \big\|_{L^2} \\
&\precsim \Bigg(T^{-1/2} \sum_{t=1}^{T} \sum_{j=[t/2]}^{t-1} \varrho(j, 0) + T^{-1/2} \sum_{t=1}^{T} \sum_{j=0}^{[t/2]-1} \varrho(t, j) \Bigg).
\end{aligned}
$$

By Kronecker's lemma, it suffices to show that

$$
\sum_{t=1}^{T} t^{-1/2} \sum_{j=[t/2]}^{t-1} \varrho(j, 0) \quad \text{and} \quad \sum_{t=1}^{T} t^{-1/2} \sum_{j=0}^{[t/2]-1} \varrho(t, j) \tag{17}
$$

are bounded uniformly in T, where recall, $\varrho(j, k) \equiv (E[\prod_{i=k}^{j} (1 - \underline{q}(X_i))^{\frac{2a}{1-a}}])^{\frac{1-a}{2a}}$.

[28] For the complete proof, see Pouzo, Psaradakis, and Sola (2021).

Moreover, $j \mapsto \varrho(j,k)$ is nonincreasing and $k \mapsto \varrho(j,k)$ is nondecreasing since $1-\underline{q}(\cdot) \leq 1$. By Assumption 8, $(\varrho(j,0))_j$ is p-summable with $p < 2/3$, thus $\lim_{j\to\infty} \varrho(j,0)^p j = 0$ (if not, then $\varrho(j,0) > c/j^{1/p}$ for some $c > 0$ and all j above certain point and this violates the assumption). Hence,

$$\sum_{j=[t/2]}^{t-1} \varrho(j,0) < \sum_{j=[t/2]}^{t-1} \frac{1}{j^{1/p}} \leq \int_{[t/2]+1}^{t} x^{-1/p}\, dx \leq \frac{p}{1-p}(t/2)^{1-1/p},$$

for all $t \geq \tau$ and some $\tau > 0$, and this implies that, for some constant const > 0,

$$\sum_{t=1}^{T} t^{-1/2} \sum_{j=[t/2]}^{t-1} \varrho(j,0) \leq C(\tau) + \text{const} \times \sum_{t=\tau+1}^{T} \frac{p}{1-p} t^{1-1/p-1/2} \leq C < \infty,$$

because $1 - 1/p - 1/2 < -1 \Leftrightarrow p < 2/3$ (C is a finite constant, which may depend on τ).

By stationarity of $X_{-\infty}^{\infty}$ (Lemma 1) and some simple algebra,

$$\sum_{j=0}^{[t/2]-1} \varrho(t,j) = \sum_{j=0}^{[t/2]+1} \varrho(t-j,0).$$

Thus, $\sum_{j=0}^{[t/2]+1} \varrho(t-j,0) \leq \sum_{j=0}^{[t/2]+1} \frac{1}{(t-j)^{1/p}} \leq \int_{[t/2]-1}^{t} \frac{1}{u^{1/p}} du$ and by our previous calculation the result follows. Thus, the terms in (17) are uniformly bounded. *Q.E.D.*

PROOF OF LEMMA 19: It is easy to see that $\Delta_{t,-\infty}(\theta_*)$ is adapted to the filtration associated with the σ-algebra generated by $X_{-\infty}^{t}$. Since $X_{-\infty}^{\infty}$ is stationary and ergodic (by Lemma 1), so is $(\Delta_{t,-\infty}(\theta_*))_{t=-\infty}^{\infty}$. *Q.E.D.*

To prove Lemma 20, we need the following results.

LEMMA 22: *Suppose Assumption* 1 *holds*. *Then there exists a finite constant* $L > 0$, *such that*, *for any* $-m \leq j < n \leq k$ *and any* $\theta \in \Theta$,

$$\max_{b,c} \left\| \bar{P}_\theta^\nu(S_j = \cdot | S_n = b, X_{-m}^k) - \bar{P}_\theta^\nu(S_j = \cdot | S_n = c, X_{-m}^k) \right\|_1 \leq L \prod_{i=j}^{n} (1 - q(X_i))$$

a.s.-$\bar{P}_*^\nu$.

LEMMA 23: *For any* $-m < i < l \leq r \leq n$, *let* $\mathbf{S}_l^r \equiv (S_l, \ldots, S_r)$. *Then*, *for any* $\theta \in \Theta$,

$$\bar{P}_\theta^\nu(S_i | \mathbf{S}_l^r, X_{-m+1}^n) = \bar{P}_\theta^\nu(S_i | S_l, X_{-m+1}^n),$$

that is, *the Markov property holds backward in time*.

PROOF OF LEMMA 22: Observe that, for any $b, c \in \mathbb{S}^2$,

$$\begin{aligned}
&\left\| \bar{P}_\theta^\nu(S_j = \cdot | S_n = b, X_{-m}^k) - \bar{P}_\theta^\nu(S_j = \cdot | S_n = c, X_{-m}^k) \right\|_1 \\
&\quad = \left\| \sum_{s\in\mathbb{S}} \bar{P}_\theta^\nu(S_j = \cdot | S_{j+1} = s, X_{-m}^{j+1}) \big(\bar{P}_\theta^\nu(S_{j+1} = s | S_n = b, X_{-m}^k) \right.
\end{aligned}$$

$$- \bar{P}^{\nu}_{\theta}(S_{j+1} = s | S_n = c, X^k_{-m}))\Big\|_1$$
$$\leq \alpha_{\theta,j+1,j}(X^k_{-m}) \| \bar{P}^{\nu}_{\theta}(S_{j+1} = \cdot | S_n = b, X^k_{-m}) - \bar{P}^{\nu}_{\theta}(S_{j+1} = \cdot | S_n = c, X^k_{-m}) \|_1,$$

where the second line follows from Lemma 23 with $i = j$, $r = l = j + 1$, and $n = k$, and the third follows from Lemma B.2.1 in Stachurski (2009) and the definition of $\alpha_{\theta,j+1,j}(X^k_{-m})$ in expression (12). Iterating in this fashion, it follows that

$$\| \bar{P}^{\nu}_{\theta}(S_j = \cdot | S_n = b, X^k_{-m}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot | S_n = c, X^k_{-m}) \|_1 \leq 2 \prod_{l=j}^{n} \alpha_{\theta,l+1,l}(X^k_{-m}).$$

Thus, it suffices to show that $\alpha_{\theta,l+1,l}(X^k_{-m}) \leq 1 - \underline{q}(X_l)$. Since

$$\alpha_{\theta,l+1,l}(X^k_{-m}) = 1 - \min_{a,b} \sum_{s' \in \mathbb{S}} \min\{ \bar{P}^{\nu}_{\theta}(S_l = s' | S_{l+1} = a, X^k_{-m}), \bar{P}^{\nu}_{\theta}(S_l = s' | S_{l+1} = b, X^k_{-m}) \}$$

(see Stachurski (2009, p. 344)), it suffices to show that, for any $(a, b) \in \mathbb{S}^2$,

$$\bar{P}^{\nu}_{\theta}(S_l = a | S_{l+1} = b, X^k_{-m}) \geq \underline{q}(X_l) \varpi(X^l_{-m+1}, a),$$

where $\varpi(X^{k-1}_{-m+1}, \cdot) \in \mathcal{P}(\mathbb{S})$.

To do this, first note that $\bar{P}^{\nu}_{\theta}(S_l = a | S_{l+1} = b, X^k_{-m}) = \bar{P}^{\nu}_{\theta}(S_l = a | S_{l+1} = b, X^{l+1}_{-m})$, by Lemma 23, and

$$\bar{P}^{\nu}_{\theta}(S_l = a | S_{l+1} = b, X^{l+1}_{-m+1}) = \frac{p_{\theta}(X_l, b, X_{l+1}) Q_{\theta}(X_l, a, b) \bar{P}^{\nu}_{\theta}(X^l_{-m+1}, S_l = a)}{\sum_{s \in \mathbb{S}} p_{\theta}(X_l, b, X_{l+1}) Q_{\theta}(X_l, s, b) \bar{P}^{\nu}_{\theta}(X^l_{-m+1}, S_l = s)}$$
$$\geq \underline{q}(X_l) \frac{\bar{P}^{\nu}_{\theta}(X^l_{-m+1}, S_l = a)}{\sum_{s \in \mathbb{S}} \bar{P}^{\nu}_{\theta}(X^l_{-m+1}, S_l = s)},$$

where the last line follows from Assumption 1. Letting $\varpi(\cdot, X^l_{-m+1}) \equiv \frac{\bar{P}^{\nu}_{\theta}(S_l = \cdot | X^l_{-m+1})}{\sum_{s \in \mathbb{S}} \bar{P}^{\nu}_{\theta}(S_l = s | X^l_{-m+1})}$, the desired result is obtained. *Q.E.D.*

PROOF OF LEMMA 23: Throughout the proof, we omit θ from the notation. Let $\mathbf{S}^r_{i:l} \equiv (S_i, S_l, S_{l+1}, \ldots, S_{r-1}, S_r)$ and note that, by Bayes' rule,

$$\bar{P}^{\nu}_{\theta}(S_i | \mathbf{S}^r_l, X^n_{-m+1}) = \frac{\bar{P}^{\nu}_{\theta}(X^n_r \mid \mathbf{S}^r_{i:l}, X^{r-1}_{-m+1}) \bar{P}^{\nu}_{\theta}(\mathbf{S}^r_{i:l}, X^{r-1}_{-m+1})}{\bar{P}^{\nu}_{\theta}(X^n_r \mid \mathbf{S}^r_l, X^{r-1}_{-m+1}) \bar{P}^{\nu}_{\theta}(\mathbf{S}^r_l, X^{r-1}_{-m+1})}.$$

By the Markov property, $\bar{P}^{\nu}_{\theta}(X^n_r \mid \mathbf{S}^r_{i:l}, X^{r-1}_{-m+1}) = \bar{P}^{\nu}_{\theta}(X^n_r \mid X_{r-1}, S_r)$, so

$$\bar{P}^{\nu}_{\theta}(S_i | \mathbf{S}^r_l, X^n_{-m+1}) = \frac{\bar{P}^{\nu}_{\theta}(\mathbf{S}^r_{i:l}, X^{r-1}_{-m+1})}{\bar{P}^{\nu}_{\theta}(\mathbf{S}^r_l, X^{r-1}_{-m+1})} = \frac{\bar{P}^{\nu}_{\theta}(S_r \mid \mathbf{S}^{r-1}_{i:l}, X^{r-1}_{-m+1}) \bar{P}^{\nu}_{\theta}(\mathbf{S}^{r-1}_{i:l}, X^{r-1}_{-m+1})}{\bar{P}^{\nu}_{\theta}(S_r \mid \mathbf{S}^{r-1}_l, X^{r-1}_{-m+1}) \bar{P}^{\nu}_{\theta}(\mathbf{S}^{r-1}_l, X^{r-1}_{-m+1})}.$$

Observe that $\bar{P}^{\nu}_{\theta}(S_r \mid \mathbf{S}_l^{r-1}, X_{-m+1}^{r-1}) = Q_\theta(X_{r-1}, S_{r-1}, S_r)$, and thus $\bar{P}^{\nu}_{\theta}(S_i | X^r_{-m+1}, \mathbf{S}^r_l) = \frac{\bar{P}^{\nu}_{\theta}(\mathbf{S}^{r-1}_{i:l}, X^{r-1}_{-m+1})}{\bar{P}^{\nu}_{\theta}(\mathbf{S}^{r-1}_{l}, X^{r-1}_{-m+1})}$ and, by iterating, it follows that

$$\bar{P}^{\nu}_{\theta}(S_i|\mathbf{S}^r_l, X^n_{-m+1}) = \frac{\bar{P}^{\nu}_{\theta}(S_i, S_l, X^l_{-m+1})}{\bar{P}^{\nu}_{\theta}(\mathbf{S}^l_l, X^l_{-m+1})} = \bar{P}^{\nu}_{\theta}(S_i|S_l, X^l_{-m+1}),$$

as desired. *Q.E.D.*

PROOF OF LEMMA 20: *Part* (i). By Lemma 23 with $l = r = k-1$, $n = k$,

$$\begin{aligned}\bar{P}^{\nu}_{\theta}(S_i|X^k_{-m+1}) &= \sum_{s\in\mathbb{S}} \bar{P}^{\nu}_{\theta}(S_i|S_{k-1} = s, X^k_{-m+1})\bar{P}^{\nu}_{\theta}(S_{k-1} = s|X^k_{-m+1}) \\ &= \sum_{s\in\mathbb{S}} \bar{P}^{\nu}_{\theta}(S_i|S_{k-1} = s, X^{k-1}_{-m+1})\bar{P}^{\nu}_{\theta}(S_{k-1} = s|X^k_{-m+1}),\end{aligned}$$

and similarly,

$$\bar{P}^{\nu}_{\theta}(S_i|X^{k-1}_{-m+1}) = \sum_{s\in\mathbb{S}} \bar{P}^{\nu}_{\theta}(S_i|S_{k-1} = s, X^{k-1}_{-m+1})\bar{P}^{\nu}_{\theta}(S_{k-1} = s|X^{k-1}_{-m+1}).$$

Thus, by Lemma B.2.2 in Stachurski (2009),

$$\begin{aligned}&\left\| \bar{P}^{\nu}_{\theta}(S_j = \cdot|X^k_{-m+1}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot|X^{k-1}_{-m+1}) \right\|_1 \\ &\quad\le \max_{a,b} \left\| \bar{P}^{\nu}_{\theta}(S_j = \cdot|S_{k-1} = a, X^{k-1}_{-m+1}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot|S_{k-1} = b, X^{k-1}_{-m+1}) \right\|_1 \\ &\quad\le L \prod_{l=j}^{k-1} (1 - q(X_l)),\end{aligned}$$

where the second line follows by Lemma 22 with $n = k-1$. Thus, the desired result follows.

Part (ii). The proof is analogous to that of Lemma 5 (third part) in Bickel, Ritov, and Rydén (1998). By analogous calculations to those in part (i),

$$\begin{aligned}&\left\| \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid X^k_{-m}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid X^k_{-n}) \right\|_1 \\ &\quad\le \max_{b,b'} \left\| \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid S_{-m} = b, X^k_{-m}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid S_{-m} = b', X^k_{-n}) \right\|_1 \\ &\quad= \max_{b,b'} \left\| \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid S_{-m} = b, X^k_{-m}) - \bar{P}^{\nu}_{\theta}(S_j = \cdot \mid S_{-m} = b', X^k_{-m}) \right\|_1,\end{aligned}$$

where the last line follows from the fact that, given S_{-m}, it is the same to condition on X^k_{-m} and on X^k_{-n}. The results thus follow from following the same steps as those in the proof of Theorem 2. *Q.E.D.*

SM.7.2. *Proof of Lemma 9*

PROOF OF LEMMA 9: To simplify the exposition, we present the proof for the case where $\Delta_{t,-\infty}(\theta_*)(X^t_{-\infty})$ is a scalar; since the dimension of this quantity is finite, the vector case follows readily from the results below.

Part (a) follows easily from Lemma 17 in the Supplemental Material SM.7.1.

For part (b), it follows from part (a) that $\Delta_{t+j,-\infty}(\theta_*)(X_{-\infty}^{t+j})\Delta_{t,-\infty}(\theta_*)(X_{-\infty}^{t})$ depends only on $(X_{t-\bar{L}}, \dots, X_{t+j})$. By Lemma 1, $(X_t)_{t=-\infty}^{\infty}$ is β-mixing. Since, for any fixed j, the σ-algebra generated by $(\Delta_{s+j,-\infty}(\theta_*)\Delta_{s,-\infty}(\theta_*))_{s\leq t}$ is contained in the σ-algebra generated by $(X_{s+j})_{s\leq t}$, and the σ-algebra generated by $(\Delta_{s+j,-\infty}(\theta_*)\Delta_{s,-\infty}(\theta_*))_{s\geq t}$ is contained in the σ-algebra generated by $(X_s)_{s\geq t-\bar{L}}$, it follows that, for each j, $(\Delta_{t+j,-\infty}(\theta_*)\Delta_{t,-\infty}(\theta_*))_{t=-\infty}^{\infty}$ is also β-mixing with mixing coefficients that decay at rate $O(\gamma^{n-2\bar{L}})$ as n diverges through the positive integers; as $\bar{L}$ is taken to be fixed, the decay rate is $O(\gamma^n)$. As is well known, this result implies that the corresponding α-mixing coefficients $(\alpha_n)_n$ decay at the same rate, that is, $\alpha_n = O(\gamma^n)$ for all n.

Henceforth, let $\Omega_{t+j,t,-\infty}(\theta_*) \equiv \Delta_{t+j,-\infty}(\theta_*)(X_{-\infty}^{t+j})\Delta_{t,-\infty}(\theta_*)(X_{-\infty}^{t})$ and $\bar{\Omega}_{t+j,t,-\infty}(\theta_*) \equiv \Omega_{t+j,t,-\infty}(\theta_*) - E_{\bar{P}_*^\nu}[\Omega_{t+j,t,-\infty}(\theta_*)]$, for any t, j. Observe that, for any j,

$$E_{\bar{P}_*^\nu}\left[\left(T^{-1}\sum_{t=1}^{T}\bar{\Omega}_{t+j,t,-\infty}(\theta_*)\right)^2\right] = T^{-1}E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta_*)\right)^2\right] + 2T^{-1}\sum_{t=0}^{T-1}(1-t/T)E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{t+j,t,-\infty}(\theta_*)\right)\left(\bar{\Omega}_{j,0,-\infty}(\theta_*)\right)\right],$$

where the last equality follows by stationarity. By Corollary 6.17 in White (2001), for any $m \in \mathbb{N}$,

$$\left|E_{\bar{P}_*^\nu}\left[\bar{\Omega}_{j,0,-\infty}(\theta_*)\bar{\Omega}_{j+m,m,-\infty}(\theta_*)\right]\right| \precsim (\alpha_m)^{\frac{2}{2+2\delta}}\sqrt{E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta_*)\right)^2\right]}\left(E_{\bar{P}_*^\nu}\left[\left|\bar{\Omega}_{j+m,m,-\infty}(\theta_*)\right|^{2+2\delta}\right]\right)^{\frac{1}{2+2\delta}},$$

for any j (the implicit constant in the display does not depend on j). Thus, for any j,

$$\begin{aligned}
&E_{\bar{P}_*^\nu}\left[\left(T^{-1}\sum_{t=1}^{T}\bar{\Omega}_{t+j,t,-\infty}(\theta_*)\right)^2\right] \\
&\quad\precsim T^{-1}E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta_*)\right)^2\right] \\
&\qquad + 2T^{-1}\sum_{t=0}^{T-1}(1-t/T)(\alpha_t)^{2/(2+2\delta)}\sqrt{E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta_*)\right)^2\right]} \\
&\qquad \times \left(E_{\bar{P}_*^\nu}\left[\left|\bar{\Omega}_{j+m,m,-\infty}(\theta_*)\right|^{2+2\delta}\right]\right)^{1/(2+2\delta)}.
\end{aligned}$$

By the Cauchy–Schwarz inequality and stationarity, for any j,

$$E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta^*)\right)^2\right] \precsim E_{\bar{P}_*^\nu}\left[\Delta_{0,-\infty}(\theta_*)\left(X_{-\infty}^{0}\right)^4\right],$$

which is bounded by assumption. In addition, by similar calculations,

$$E_{\bar{P}_*^\nu}\left[\left(\bar{\Omega}_{j,0,-\infty}(\theta^*)\right)^{2+2\delta}\right] \precsim E_{\bar{P}_*^\nu}\left[\Delta_{0,-\infty}(\theta_*)\left(X_{-\infty}^{0}\right)^{4+4\delta}\right] \quad \forall j,$$

which is bounded by assumption. Therefore, there exists a finite constant C (which does not depend on j) such that

$$E_{\bar{P}_*^\nu}\left[\left(T^{-1}\sum_{t=1}^{T}\bar{\Omega}_{t+j,t,-\infty}(\theta^*)\right)^2\right] \leq C\left(T^{-1} + 2T^{-1}\sum_{t=0}^{T-1}(1-t/T)(\alpha_t)^{2/(2+2\delta)}\right).$$

As $\alpha_t = O(\gamma^t)$, it follows that $\sum_{t=0}^{T-1}(1-t/T)(\alpha_t)^{2/(2+2\delta)} = O(\sum_{t=0}^{T-1}(1-t/T)(\gamma^{2/(2+2\delta)})^t)$. Since $\gamma < 1$, it follows that $\sum_{t=0}^{T-1}(1-t/T)(\alpha_t)^{2/(2+2\delta)} = O(1)$, which in turn implies that $E_{\bar{P}_*^\nu}[(T^{-1}\sum_{t=1}^{T}\bar{\Omega}_{t+j,t,-\infty}(\theta^*))^2] \leq CT^{-1}$. Hence, by the Markov inequality, for any $a > 0$,

$$\begin{aligned}
&\bar{P}_*^\nu\Bigg(\max_{j\in\{0,\dots,L\}}\Bigg\|T^{-1}\sum_{t=1}^{T}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal \\
&\qquad - E_{\bar{P}_*^\nu}\big[\Delta_{j,-\infty}(\theta_*)\big(X_{-\infty}^{j}\big)\Delta_{0,-\infty}(\theta_*)\big(X_{-\infty}^{0}\big)^\intercal\big]\Bigg\| \geq a\Bigg) \\
&\leq \sum_{j=0}^{L}\bar{P}_*^\nu\Bigg(\Bigg\|T^{-1}\sum_{t=1}^{T}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal \\
&\qquad - E_{\bar{P}_*^\nu}\big[\Delta_{j,-\infty}(\theta_*)\big(X_{-\infty}^{j}\big)\Delta_{0,-\infty}(\theta_*)\big(X_{-\infty}^{0}\big)^\intercal\big]\Bigg\| \geq a\Bigg) \\
&\leq Ca^{-2}LT^{-1},
\end{aligned}$$

which implies the desired result. *Q.E.D.*

SM.8. PROOF OF THEOREM 5

To prove Theorem 5, we use the following results, whose proofs are relegated to the end of this section. The following lemma shows that we can "quantify" convergence in probability.

LEMMA 24: *Suppose a random sequence $X_T)_{T=0}^{\infty}$ converges to zero in probability. Then there exists a deterministic positive sequence $(r_T)_{T=0}^{\infty}$ such that $r_T = o(1)$, and for any $\epsilon > 0$, there exists T_ϵ such that*

$$\Pr\big(|X_T| \geq r_T\big) \leq \epsilon,$$

for all $T \geq T_\epsilon$. In particular, $|X_T| = O_{\Pr}(r_T)$.

The next lemma presents some useful properties for the "score process."

LEMMA 25: *Under the Assumptions of Theorem 5, the following are true:*

1. $\|\sup_{\theta\in B(\delta,\theta_*)}\|\Delta_{0,-\infty}(\theta)\|\|_{L^2(\bar{P}_*^\nu)} < \infty$ ($\delta > 0$ *is as in Assumption* 7).
2. $\Delta_{t,-\infty}$ *and* $\Delta_{t,-\infty}\Delta_{0,-\infty}^\intercal$ *are continuous in* $L^1(\bar{P}_*^\nu)$*-norm, that is, for any* $\epsilon > 0$*, there exists* $\delta > 0$ *such that*

$$\max_t\Big\|\sup_{\|\theta-\theta_0\|<\delta}\big\|\Delta_{t,-\infty}(\theta)\Delta_{t,-\infty}(\theta)^\intercal - \Delta_{t,-\infty}(\theta_0)\Delta_{t,-\infty}(\theta_0)^\intercal\big\|\Big\|_{L^1(\bar{P}_*^\nu)} \leq \epsilon$$

and

$$\ddot{\varpi}(\delta) \equiv \max_t \left\| \sup_{\|\theta-\theta_0\|<\delta} \left\| \Delta_{t,-\infty}(\theta)\Delta_{0,-\infty}(\theta)^\intercal - \Delta_{t,-\infty}(\theta_0)\Delta_{0,-\infty}(\theta_0)^\intercal \right\| \right\|_{L^1(\bar{P}^\nu_*)} \leq \epsilon.$$

3. *There exists a constant $C < \infty$ such that, for any t and M,*

$$\left\| \sup_{\theta \in B(\delta,\theta_*)} \Delta_{t,-\infty}(\theta) - \Delta_{t,t-M}(\theta) \right\|_{L^2(\bar{P}^\nu_*)} \leq CM^{1-1/p}.$$

Moreover, by Assumption 8, $p \in (0, 2/3)$, *and thus the RHS vanishes as M diverges.*

PROOF OF THEOREM 5: The proof has several parts and steps.

PART (A). We show that

$$\left\| T^{-1} \sum_{t=1}^{T} \nabla^2_\theta \log p^\nu_t\big(X_t|X_0^{t-1}, \hat{\theta}_{\nu,T}\big) - E_{\bar{P}^\nu_*}\big[\xi_1(\theta_*)\big] \right\| = o_{\bar{P}^\nu_*}(1).$$

We do this by using the triangle inequality and showing the following:

$$\lim_{T\to\infty} \left\| T^{-1} \sum_{t=1}^{T} \big\{ \nabla^2_\theta \log p^\nu_t(\cdot|\cdot, \hat{\theta}_{\nu,T}) - \xi_t(\hat{\theta}_{\nu,T}) \big\} \right\|_{L^1(\bar{P}^\nu_*)} = 0$$

(which implies convergence in probability), $\|T^{-1}\sum_{t=1}^T\{\xi_t(\hat{\theta}_{\nu,T}) - \xi_t(\theta_*)\}\| = o_{\bar{P}^\nu_*}(1)$, and $\|T^{-1}\sum_{t=1}^T \xi_t(\theta_*) - E_{\bar{P}^\nu_*}[\xi_1(\theta_*)]\| = o_{\bar{P}^\nu_*}(1)$.

The first expression holds true because by Theorem 1, for any $\delta' \leq \delta$, $\hat{\theta}_{\nu,T} \in B(\delta', \theta_*)$ w.p.a.1, and hence, by Lemma 8, the desired result follows.

Regarding the second expression, again by Theorem 1, $\hat{\theta}_{\nu,T} \in B(\delta', \theta_*)$ w.p.a.1. Thus, by the Markov inequality and stationarity, it follows that, for any $\epsilon > 0$, there exists $T(\epsilon)$ such that, for any $t \geq T(\epsilon)$,

$$\bar{P}^\nu_*\left(\left\| T^{-1} \sum_{t=1}^{T} \big\{ \xi_t(\hat{\theta}_{\nu,T}) - \xi_t(\theta_*) \big\} \right\| \geq \epsilon \right) \leq \epsilon^{-1} E_{\bar{P}^\nu_*}\Big[\sup_{\theta\in B(\delta',\theta_*)} \big\| \xi_1(\theta) - \xi_1(\theta_*) \big\| \Big] + 0.5\epsilon,$$

By Lemma 8, ξ_1 is continuous, and thus uniformly continuous over compact sets. Since $\delta' > 0$ can be chosen to be any number less than δ (δ as in Assumption 7), we can choose it so that the first term in the RHS is less than 0.5ϵ. Hence, the desired follows.

Finally, ergodicity of $X_{-\infty}^{\infty}$ (Lemma 1) implies ergodicity of $(\xi_t(\theta_*))_{t=-\infty}^{\infty}$; therefore, by Lemma 8 and Birkhoff's ergodic theorem, $\|T^{-1}\sum_{t=1}^T \xi_t(\theta_*) - E_{\bar{P}^\nu_*}[\xi_1(\theta_*)]\| = o_{\bar{P}^\nu_*}(1)$. Hence,

$$\left\| T^{-1} \sum_{t=1}^{T} \nabla^2_\theta \log p^\nu_t\big(X_t|X_0^{t-1}, \hat{\theta}_{\nu,T}\big) - E_{\bar{P}^\nu_*}\big[\xi_1(\theta_*)\big] \right\| = o_{\bar{P}^\nu_*}(1).$$

With this result and the Fisher information equality (established in the proof of Corollary 1), the result of part (a) of the theorem follows.

PART (B). STEP 1 To prove part (b), it suffices to show that $\|E_{\bar{P}^\nu_*}[\xi_1(\theta_*)] - H_T(\hat{\theta}_{\nu,T})\| = o_{\bar{P}^\nu_*}(1)$ and $\|\Sigma_T(\theta_*) - J_T(\hat{\theta}_{\nu,T})\| = o_{\bar{P}^\nu_*}(1)$.

The first expression was established in Part (A). Regarding the second expression, we introduce some notation. For any $\theta \in \Theta$, let

$$\Omega_{t,l,M}(\theta) \equiv \Delta_{t,M}(\theta)\big(X_M^t\big)\Delta_{l,M}(\theta)\big(X_M^l\big)^\intercal, \quad \forall t, l, M \in \mathbb{N},$$

where it is left implicit that this quantity depends on $X_{-M}^{\max\{t,l\}}$. Also,

$$\theta \mapsto \hat{\gamma}_{T,\tau,0}(\theta) \equiv (T-\tau)^{-1} \sum_{t=1}^{T-\tau} \Omega_{t+\tau,t,0}(\theta), \quad \forall \tau \in \{1, \dots, L_T\}.$$

Recall that $\Delta_{t,0}(\theta)(X_0^t) = \nabla_\theta p_t^\nu(X_t \mid X_0^{t-1}, \theta)$, so $\hat{\gamma}_{T,\tau,0}(\theta)$ is the sample covariance of $\Omega_{t+\tau,t,0}(\theta)$.

Given this notation, observe that

$$\Sigma_T(\theta_*) = T^{-1} \sum_{t=1}^{T} E_{\bar{P}_*^\nu}\big[\Omega_{t,t,-\infty}(\theta_*)\big] + T^{-1} \sum_{t=1}^{T} \sum_{l=0}^{t-1} \big\{E_{\bar{P}_*^\nu}\big[\Omega_{t,l,-\infty}(\theta_*)\big] + E_{\bar{P}_*^\nu}\big[\Omega_{t,l,-\infty}(\theta_*)^\intercal\big]\big\}.$$

The aim is to show that each of the summands above is well approximated by its counterpart in J_T. For the first summand, we show in Step 2 below that

$$\left\| T^{-1} \sum_{t=1}^{T} \big\{\Omega_{t,t,0}(\hat{\theta}_{\nu,T}) - E_{\bar{P}_*^\nu}\big[\Omega_{t,t,-\infty}(\theta_*)\big]\big\} \right\| = o_{\bar{P}_*^\nu}(1).$$

Regarding the second summand, we observe that, for any $t \geq l$, $E_{\bar{P}_*^\nu}[\Omega_{t,l,-\infty}(\theta_*)] = E_{\bar{P}_*^\nu}[\Omega_{t-l,0,-\infty}(\theta_*)] \equiv \gamma_{t-l}(\theta_*)$ (the first equality, which follows from stationarity, can be established by analogous arguments to those presented at the beginning of Step 2). Hence,

$$T^{-1} \sum_{t=1}^{T} \sum_{l=0}^{t-1} E_{\bar{P}_*^\nu}\big[\Omega_{t,l,-\infty}(\theta_*)\big] = \sum_{j=0}^{T-1} (1 - j/T)\gamma_{j+1}(\theta_*).$$

Thus, it suffices to show that

$$\left\| \sum_{j=0}^{T-1} (1 - j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1} \omega(j, L)\hat{\gamma}_{T,j+1,0}(\hat{\theta}_{T,\nu}) \right\| = o_{\bar{P}_*^\nu}(1),$$

the proof of which is in Step 3 below.

STEP 2 We now show that

$$\left\| T^{-1} \sum_{t=1}^{T} \big\{\Delta_{t,0}(\hat{\theta}_{\nu,T})\big(X_0^t\big)\Delta_{t,0}(\hat{\theta}_{\nu,T})\big(X^t\big)^\intercal - E_{\bar{P}_*^\nu}\big[\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^t\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^t\big)^\intercal\big]\big\} \right\|$$
$$= o_{\bar{P}_*^\nu}(1),$$

where, by the definition of $\Delta_{t,M}$, $\Delta_{t,0}(\theta)(X^t) = \nabla_\theta \log p_t(X_t \mid X_0^{t-1}, \theta)$.

First, observe that

$$E_{\bar{P}_*^\nu}\big[\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^t\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^t\big)^\intercal\big] = E_{\bar{P}_*^\nu}\big[\Delta_{0,-\infty}(\theta_*)\big(X_{-\infty}^0\big)\Delta_{0,-\infty}(\theta_*)\big(X_{-\infty}^0\big)^\intercal\big].$$

This follows from stationarity (see Lemma 1) and the fact that $\Delta_{t,-\infty}(\theta_*)$ can be approximated (uniformly in t) by $\Delta_{t,t-M}(\theta_*)$ (see Lemma 18). Hence, it suffices to show that

$$\left\| T^{-1}\sum_{t=1}^{T}\Delta_{t,0}(\hat{\theta}_{\nu,T})(X^t)\Delta_{t,0}(\hat{\theta}_{\nu,T})(X^t)^\intercal - E_{\bar{P}_*^\nu}\left[\Delta_{0,-\infty}(\theta_*)(X_{-\infty}^0)\Delta_{0,-\infty}(\theta_*)(X_{-\infty}^0)^\intercal\right]\right\| \\ = o_{\bar{P}_*^\nu}(1).$$

By Lemma 1, ergodicity of $\Delta_{t,t-M}(\theta_*)$ for any M follows. This, Birkhoff's ergodic theorem, and Lemma 18 imply that

$$\left\| T^{-1}\sum_{t=1}^{T}\bar{\Delta}_\infty(\theta_*)(X_{-\infty}^t) - E_{\bar{P}_*^\nu}\left[\bar{\Delta}_\infty(\theta_*)(X_{-\infty}^0)\right]\right\| = o_{\bar{P}_*^\nu}(1),$$

where, for any $M \in \mathbb{Z}\cup\{\infty\}$, $\bar{\Delta}_M(\theta)(X_{t-M}^t) \equiv \Delta_{0,-M}(\theta)(X_{t-M}^t)\Delta_{0,-M}(\theta)(X_{t-M}^t)^\intercal$.

Hence, in order to obtain the desired result it suffices to show that

$$\left\| T^{-1}\sum_{t=1}^{T}\left\{\bar{\Delta}_t(\hat{\theta}_{\nu,T})(X_0^t) - \bar{\Delta}_\infty(\theta_*)(X_{-\infty}^t)\right\}\right\| = o_{\bar{P}_*^\nu}(1).$$

In order to do so, by the triangle inequality, it is sufficient to show that

$$\left\| T^{-1}\sum_{t=1}^{T}\left\{\bar{\Delta}_\infty(\theta_*)(X_{-\infty}^t) - \bar{\Delta}_\infty(\hat{\theta}_{\nu,T})(X_{-\infty}^t)\right\}\right\| = o_{\bar{P}_*^\nu}(1) \tag{18}$$

and

$$\left\| T^{-1}\sum_{t=1}^{T}\left\{\bar{\Delta}_t(\hat{\theta}_{\nu,T})(X_0^t) - \bar{\Delta}_\infty(\hat{\theta}_{\nu,T})(X_{-\infty}^t)\right\}\right\| = o_{\bar{P}_*^\nu}(1). \tag{19}$$

Expression (18) holds by Lemma 25, the fact that, for any $\delta > 0$, $\hat{\theta}_{T,\nu} \in B(\delta,\theta_*)$ w.p.a.1-$\bar{P}_*^\nu$ (by Theorem 1), and the Markov inequality. Regarding expression (19), by the Markov inequality and the fact that $\hat{\theta}_{T,\nu} \in B(\delta,\theta_*)$ w.p.a.1-$\bar{P}_*^\nu$ (by Theorem 1), it is sufficient to show that

$$T^{-1}\sum_{t=1}^{T}E_{\bar{P}_*^\nu}\left[\sup_{\theta\in B(\delta,\theta_*)}\left\|\bar{\Delta}_t(\theta)(X_0^t) - \bar{\Delta}_\infty(\theta_*)(X_{-\infty}^t)\right\|\right] = o(1).$$

The LHS is bounded by

$$T^{-1}\sum_{t=1}^{T}\left\|\sup_{\theta\in B(\delta,\theta_*)}\left\|\Delta_{t,0}(\theta) - \Delta_{t,-\infty}(\theta_*)\right\|\right\|_{L^2(\bar{P}_*^\nu)} \\ \times\left(\left\|\sup_{\theta\in B(\delta,\theta_*)}\left\|\Delta_{t,0}(\theta)\right\|\right\|_{L^2(\bar{P}_*^\nu)} + \left\|\sup_{\theta\in B(\delta,\theta_*)}\left\|\Delta_{t,-\infty}(\theta)\right\|\right\|_{L^2(\bar{P}_*^\nu)}\right).$$

By Lemma 25(3), the first term in the RHS is bounded (up to constants) by $T^{-1}\sum_{t=1}^{T} t^{1-1/p}$. The second term in the RHS is bounded by Lemma 25(1). Thus, under Assumption 8, the whole expression converges to zero and the desired result follows.

STEP 3. We next show that, for any $L \equiv L_T$ such that $\lim_{T\to\infty} L_T = \infty$ and $L_T(\ddot{\varpi}(T^{-1/2}\log\log T)\log\log T + r_T + T^{-1/2}) = o(1)$,

$$\left\| \sum_{j=0}^{T-1}(1-j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1}\omega(j,L)\hat{\gamma}_{T,j+1,0}(\hat{\theta}_{T,\nu})\right\| = o_{\bar{P}_*^{\nu}}(1),$$

where, for any $\tau \in \{1,\dots,L_T\}$ and any $M \le 1$,

$$\theta \mapsto \hat{\gamma}_{T,\tau,M}(\theta) \equiv T^{-1}\sum_{t=1}^{T-\tau}\Delta_{t+\tau,M}(\theta)\big(X_M^{t+\tau}\big)\Delta_{t,M}(\theta)\big(X_M^{t}\big)^{\intercal}$$

(recall that $\Delta_{t,0}(\theta)(X_0^t) = \nabla_\theta p_t^{\nu}(X_t \mid X_0^{t-1},\theta)$).

Putting $\hat{\gamma}_{T,\tau} \equiv \hat{\gamma}_{T,\tau,-\infty}$, we have, by the triangle inequality, that $\|\sum_{j=0}^{T-1}(1-j/T)\times \gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1}\omega(j,L)\hat{\gamma}_{T,j+1,0}(\hat{\theta}_{T,\nu})\|$ is bounded by

$$\begin{aligned}&\left\|\sum_{j=0}^{T-1}(1-j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1}\omega(j,L)\gamma_{j+1}(\theta_*)\right\| + \left\|\sum_{j=0}^{L_T-1}\omega(j,L)\big\{\gamma_{j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\theta_*)\big\}\right\| \\ &\quad + \left\|\sum_{j=0}^{L_T-1}\omega(j,L)\big\{\hat{\gamma}_{T,j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\hat{\theta}_{T,\nu})\big\}\right\| \\ &\quad + \left\|\sum_{j=0}^{L_T-1}\omega(j,L)\big\{\hat{\gamma}_{T,j+1,0}(\hat{\theta}_{T,\nu}) - \hat{\gamma}_{T,j+1}(\hat{\theta}_{T,\nu})\big\}\right\|.\end{aligned}$$

We now bound each term in the RHS individually.

By assumption, for any $l \ge 0$, $\|\gamma_l(\theta_*)\| \le \upsilon(l)$, and thus, for any L,

$$\begin{aligned}&\left\|\sum_{j=0}^{T-1}(1-j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L-1}\omega(j,L)\gamma_{j+1}(\theta_*)\right\| \\ &\quad \le \sum_{j=L}^{\infty}\upsilon(j) + \left\|\sum_{j=0}^{L-1}\big\{(1-j/T)-\omega(j,L)\big\}\gamma_{j+1}(\theta_*)\right\|.\end{aligned}$$

Since, υ is integrable, the first term in the RHS converges to zero as L diverges. Furthermore, since $\omega(\cdot,\cdot)$ is bounded, $\|\gamma_{j+1}(\theta_*)\| \le \upsilon(j+1)$, which is integrable, and $(1-j/T)-\omega(j,L)$ converges to zero pointwise in j as T (and thus $L = L_T$) diverges, so by the dominated convergence theorem, the second term also converges to zero as T (and thus $L = L_T$) diverges. Therefore, for any $\epsilon > 0$, there exists T_ϵ such that, for all $T \ge T_\epsilon$,

$$\left\|\sum_{j=0}^{T-1}(1-j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1}\omega(j,L_T)\gamma_{j+1}(\theta_*)\right\| \le \epsilon.$$

We now consider, for any $\delta > 0$,

$$\bar{P}_*^\nu\left(\left\|\sum_{j=0}^{L-1}\omega(j,L)\left\{T^{-1}\sum_{t=1}^{T-j}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal - \gamma_j(\theta_*)\right\}\right\| \geq \delta\right).$$

Since $\sum_{j=0}^{L-1}\omega(j,L) \leq L$, this expression is bounded above by

$$\bar{P}_*^\nu\left(\max_{j\in\{0,\ldots,L_T\}}\left\|T^{-1}\sum_{t=1}^{T-j}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal - \gamma_j(\theta_*)\right\| \geq \delta/L\right).$$

By similar arguments to those presented in Step 2 and Birkhoff's ergodic theorem, for each L,

$$\max_{j\in\{0,\ldots,L\}}\left\|T^{-1}\sum_{t=1}^{T-j}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal - \gamma_j(\theta_*)\right\| = o_{\bar{P}_*^\nu}(1).$$

By Lemma 24, for each L, there exists a positive sequence $(r_T)_T$ such that $r_T = o(1)$ and $\bar{P}_*^\nu(\max_{j\in\{0,\ldots,L\}}\|T^{-1}\sum_{t=1}^{T}\Delta_{t+j,-\infty}(\theta_*)(X_{-\infty}^{t+j})\Delta_{t,-\infty}(\theta_*)(X_{-\infty}^{t})^\intercal - \gamma_j(\theta_*)\| \geq r_T) = o(1)$. Thus, by setting $\delta = 2r_TL$, for any $\epsilon > 0$, there exists T_ϵ such that, for all $T \geq T_\epsilon$,

$$\bar{P}_*^\nu\left(\left\|\sum_{j=0}^{L-1}\omega(j,L)\left\{T^{-1}\sum_{t=1}^{T-j}\Delta_{t+j,-\infty}(\theta_*)\big(X_{-\infty}^{t+j}\big)\Delta_{t,-\infty}(\theta_*)\big(X_{-\infty}^{t}\big)^\intercal - \gamma_j(\theta_*)\right\}\right\| \geq 2r_TL\right) \leq \epsilon.$$

By Theorem 4, $\hat{\theta}_{\nu,T} \in B(T^{-1/2}\log\log T, \theta_*)$ w.p.a.1-$\bar{P}_*^\nu$. Hence, for any $\epsilon > 0$ there exists T_ϵ such that for all $T \geq T_\epsilon$,

$$\begin{aligned}
&\bar{P}_*^\nu\left(\left\|\sum_{j=0}^{L-1}\omega(j,L)\big\{\hat{\gamma}_{T,j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\hat{\theta}_{\nu,T})\big\}\right\| \geq L\ddot{\varpi}\big(T^{-1/2}\log\log T\big)\log\log T\right)\\
&\quad\leq \bar{P}_*^\nu\left(\sup_{\theta\in B(T^{-1/2}\log\log T,\theta_*)}\sum_{j=0}^{L-1}\omega(j,L)\big\|\hat{\gamma}_{T,j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\theta)\big\|\right.\\
&\qquad\left.\geq L\ddot{\varpi}\big(T^{-1/2}\log\log T\big)\log\log T\right) + \epsilon.
\end{aligned}$$

Moreover, by the Markov inequality,

$$\begin{aligned}
&\bar{P}_*^\nu\left(\sup_{\theta\in B(T^{-1/2}\log\log T,\theta_*)}\sum_{j=0}^{L-1}\omega(j,L)\big\|\hat{\gamma}_{T,j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\theta)\big\| \geq L\ddot{\varpi}\big(T^{-1/2}\log\log T\big)\log\log T\right)\\
&\quad\leq \frac{\sum_{j=0}^{L-1}\omega(j,L)\ddot{\varpi}\big(T^{-1/2}\log\log T\big)}{L\ddot{\varpi}\big(T^{-1/2}\log\log T\big)\log\log T},
\end{aligned}$$

where the last line follows from Lemma 25 and $\tilde{\Delta}_{t,-M}(\theta) \equiv \Delta_{t,-M}(\theta)\Delta_{0,-M}(\theta)^{\intercal}$ for any $t \in \{0, \dots, T, \dots\}$ and $M \in \mathbb{N} \cup \{\infty\}$. Since $\sum_{j=0}^{L-1} \omega(j, L)/L \leq 1$, the last expression is less than ϵ for sufficiently large T. Thus,

$$\bar{P}_*^{\nu}\left(\left\| \sum_{j=0}^{L-1} \omega(j, L)\{\hat{\gamma}_{T,j+1}(\theta_*) - \hat{\gamma}_{T,j+1}(\hat{\theta}_{\nu,T})\} \right\| \geq L\ddot{\varpi}(T^{-1/2}\log\log T)\log\log T\right) \leq \epsilon.$$

Finally, since, for any $\tau \in \{1, \dots, L\}$, $\theta \mapsto \hat{\gamma}_{T,\tau,0}(\theta) \equiv T^{-1}\sum_{t=1}^{T-\tau} \Delta_{t+\tau,0}(\theta)(X_0^{t+\tau}) \times \Delta_{t,0}(\theta)(X_0^t)^{\intercal}$, by Lemma 25(3) (with $M = t$),

$$\left\| \sum_{j=0}^{L-1} \omega(j, L)\{\hat{\gamma}_{T,j+1}(\hat{\theta}_{\nu,T}) - \hat{\gamma}_{T,j+1,0}(\hat{\theta}_{\nu,T})\} \right\|_{L^2(\bar{P}_*^{\nu})} \precsim \sum_{j=0}^{L-1} \omega(j, L)T^{-1}\sum_{t=1}^{T-j} t^{1-1/p}.$$

By the proof of Lemma 18, $T^{-1/2}\sum_{t=1}^{T} t^{1-1/p}$ vanishes; thus, the RHS is of order $o(LT^{-1/2})$.

Therefore, we have shown that, for any $\epsilon > 0$, there exists T_{ϵ} such that, for all $T \geq T_{\epsilon}$,

$$\bar{P}_*^{\nu}\Bigg(\left\| \sum_{j=0}^{T-1}(1 - j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1} \omega(j, L)\hat{\gamma}_{T,j+1,0}(\hat{\theta}_{T,\nu}) \right\|$$
$$\geq \epsilon + L_T\big(\ddot{\varpi}(T^{-1/2}\log\log T)\log\log T + 2r_T + T^{-1/2}\big)\Bigg) \leq \epsilon,$$

where the ϵ inside the probability arises from bounding $\| \sum_{j=0}^{T-1}(1 - j/T)\gamma_{j+1}(\theta_*) - \sum_{j=0}^{L_T-1} \omega(j, L_T)\gamma_{j+1}(\theta_*)\|$ and requires L_T to diverge. Therefore, by taking $L \equiv L_T$ such that $\lim_{T\to\infty} L_T = \infty$ and $L_T(\ddot{\varpi}(T^{-1/2}\log\log T)\log\log T + r_T + T^{-1/2}) = o(1)$, we establish the desired result. *Q.E.D.*

SM.8.1. *Proofs of Supplementary Lemmas*

PROOF OF LEMMA 24: The proof is standard, and thus omitted; it can be found in Pouzo, Psaradakis, and Sola (2021). *Q.E.D.*

PROOF OF LEMMA 25: We show that $\| \sup_{\theta \in B(\delta,\theta_*)} \|\Delta_{0,-\infty}(\theta)\| \|_{L^2(\bar{P}_*^{\nu})}$ is bounded (δ is as in Assumption 7) and that $\Delta_{t,-\infty}$ and $\Delta_{t,-\infty}\Delta_{0,-\infty}$ are continuous in $L^1(\bar{P}_*^{\nu})$ norm, that is, for any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\omega(\delta) \equiv \max_t \left\| \sup_{\|\theta-\theta_0\|<\delta} \{\Delta_{t,-\infty}(\theta)\Delta_{t,-\infty}(\theta)^{\intercal} - \Delta_{t,-\infty}(\theta_0)\Delta_{t,-\infty}(\theta_0)^{\intercal}\} \right\|_{L^1(\bar{P}_*^{\nu})} \leq \epsilon \tag{20}$$

and

$$\ddot{\varpi}(\delta) \equiv \max_t \left\| \sup_{\|\theta-\theta_0\|<\delta} \{\Delta_{t,-\infty}(\theta)\Delta_{0,-\infty}(\theta)^{\intercal} - \Delta_{t,-\infty}(\theta_0)\Delta_{0,-\infty}(\theta_0)^{\intercal}\} \right\|_{L^1(\bar{P}_*^{\nu})} \leq \epsilon. \tag{21}$$

We also show that there exist a constant $C < \infty$ such that, for any t and M,

$$\left\| \sup_{\theta \in B(\delta,\theta_*)} \Delta_{t,-\infty}(\theta) - \Delta_{t,t-M}(\theta) \right\|_{L^2(\bar{P}_*^{\nu})} \leq CM^{1-1/p}. \tag{22}$$

By Assumption 8, $p \in (0, 2/3)$ so the RHS vanishes as M diverges. We first establish (22). To do so, note that, by inspection of the proof of Lemma 18, the conclusion of that lemma holds uniformly in θ (and also in t), that is,

$$\Big\| \sup_{\theta \in B(\delta,\theta_*)} \Delta_{t,-\infty}(\theta) - \Delta_{t,t-M}(\theta) \Big\|_{L^2(\bar{P}^\nu_*)} \precsim \sum_{j=[t-M/2]}^{t-1} \varrho(j, t-M) + \sum_{j=[t-M]}^{[t-M/2]-1} \varrho(t-1, j). \tag{23}$$

By the definition of ϱ and stationarity, we have that, for any $j \geq k$, $\varrho(j,k) = \varrho(j-k, 0)$, and thus $\sum_{j=[t-M/2]}^{t-1} \varrho(j, t-M) \leq \int_{[M/2]+1}^{M} 1/(x)^{1/p}\, dx \leq \frac{p}{1-p}(M/2)^{1-1/p}$, and $\sum_{j=[t-M]}^{[t-M/2]-1} \varrho(t-1-j, 0) \leq \frac{p}{1-p}(M/2)^{1-1/p}$. Thus,

$$\Big\| \sup_{\theta \in B(\delta,\theta_*)} \Delta_{t,-\infty}(\theta) - \Delta_{t,t-M}(\theta) \Big\|_{L^2(\bar{P}^\nu_*)} \precsim M^{1-1/p},$$

as desired. Since, under Assumption 7, $\| \sup_{\theta \in B(\delta,\theta_*)} \|\Delta_{0,-M}(\theta)\| \|_{L^2(\bar{P}^\nu_*)} < \infty$ for any finite M, (22) implies that $\| \sup_{\theta \in B(\delta,\theta_*)} \|\Delta_{0,-\infty}(\theta)\| \|_{L^2(\bar{P}^\nu_*)}$ is bounded.

We show next that (21) holds (the proof of (20) is completely analogous so it is omitted). To this end, observe that, for any $t \in \mathbb{N} \cup \{0\}$ and $M \in \mathbb{N} \cup \{\infty\}$,

$$\begin{aligned}
\Big\| \sup_{\|\theta-\theta_0\|<\delta} \tilde{\Delta}_{t,-\infty}(\theta) - \tilde{\Delta}_{t,-\infty}(\theta_0) \Big\|_{L^1(\bar{P}^\nu_*)} &\leq \Big\| \sup_{\theta \in B(\delta,\theta_*)} \tilde{\Delta}_{t,-\infty}(\theta) - \tilde{\Delta}_{t,t-M}(\theta) \Big\|_{L^1(\bar{P}^\nu_*)} \\
&\quad + \Big\| \sup_{\{\|\theta-\theta_0\|<\delta\} \cap B(\delta,\theta_*)} \tilde{\Delta}_{t,t-M}(\theta) - \tilde{\Delta}_{t,t-M}(\theta_0) \Big\|_{L^1(\bar{P}^\nu_*)} \\
&\quad + \Big\| \sup_{\theta_0 \in B(\delta,\theta_*)} \tilde{\Delta}_{t,-\infty}(\theta_0) - \tilde{\Delta}_{t,t-M}(\theta_0) \Big\|_{L^1(\bar{P}^\nu_*)} \\
&\equiv \mathrm{Term}_{1,t,M} + \mathrm{Term}_{2,t,M} + \mathrm{Term}_{3,t,M},
\end{aligned}$$

where $\tilde{\Delta}_{t,-M}(\theta) \equiv \Delta_{t,-M}(\theta)\Delta_{0,-M}(\theta)^\intercal$. We now bound each of these terms.

Regarding terms 1 and 3, by simple algebra and the fact that

$$\Big\| \sup_{\theta \in B(\delta,\theta_*)} \big\|\Delta_{0,-M}(\theta)\big\| \Big\|_{L^2(\bar{P}^\nu_*)} < \infty$$

for any $M \in \mathbb{N} \cup \{\infty\}$,

$$\mathrm{Term}_{1,t,M} + \mathrm{Term}_{3,t,M} \leq C \Big\| \sup_{\theta \in B(\delta,\theta_*)} \big\|\Delta_{t,-\infty}(\theta) - \Delta_{t,t-M}(\theta)\big\| \Big\|_{L^2(\bar{P}^\nu_*)},$$

for some $C < \infty$, and, by (22), the RHS is bounded by $O(M^{1-1/p})$. Therefore, under Assumption 8, for any $\epsilon > 0$, there exists an M such that, uniformly over t, $\mathrm{Term}_{1,t,M} + \mathrm{Term}_{3,t,M} \leq \epsilon$. Henceforth, fix this M.

Regarding term 2, observe that $M < \infty$ and that $\tilde{\Delta}_{t,t-M}$ is the product of two functions that are comprised of M-term-sums of products of $\theta \mapsto \log p_\theta(x, s, x')$, $\theta \mapsto \log Q_\theta(x, s, x')$ and their derivatives, all of which are continuous functions by Assumption 6. Thus, it can be shown that $\Delta_{t,t-M}$ is continuous, thereby implying that, for any $\epsilon > 0$, there exists some $\delta_{M,\epsilon}$ (which could always be chosen to be smaller than $\delta > 0$) such that $\mathrm{Term}_{2,t,M} < \epsilon$. This completes the proof. *Q.E.D.*

REFERENCES

ATHREYA, KRISHNA B., AND SOUMENDRA N. LAHIRI (2006): *Measure Theory and Probability Theory*. New York: Springer. [1,2]

BICKEL, PETER J., YA'ACOV RITOV, AND TOBIAS RYDÉN (1998): "Asymptotic Normality of the Maximum-Likelihood Estimator for General Hidden Markov Models," *Annals of Statistics*, 26, 1614–1635. [12-14,18]

DAVYDOV, YU A. (1973): "Mixing Conditions for Markov Chains," *Theory of Probability and Its Applications*, 18, 312–328. [2]

DOUC, RANDAL, GERSENDE FORT, ERIC MOULINES, AND PHILIPPE SOULIER (2004): "Practical Drift Conditions for Subgeometric Rates of Convergence," *Annals of Applied Probability*, 14, 1353–1377. [8]

DOUC, RANDAL, ÉRIC MOULINES, AND TOBIAS RYDÉN (2004): "Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models With Markov Regime," *Annals of Statistics*, 32, 2254–2304. [12]

HAIRER, MARTIN, AND JONATHAN C. MATTINGLY (2011): "Yet Another Look at Harris' Ergodic Theorem for Markov Chains," in *Seminar on Stochastic Analysis, Random Fields and Applications VI*, ed. by Robert Dalang, Marco Dozzi, and Francesco Russo. Basel: Springer, 109–117. [1]

POUZO, DEMIAN, ZACHARIAS PSARADAKIS, AND MARTIN SOLA (2021): "Maximum Likelihood Estimation in Markov Regime-Switching Models With Covariate-Dependent Transition Probabilities," arXiv:1612.04932v3 [math.ST]. [1,7,8,12,14,15,26]

STACHURSKI, JOHN (2009): *Economic Dynamics: Theory and Computation*. Cambridge, Massachusetts: MIT Press. [3,17,18]

THIERNEY, LUKE (1996): "Introduction to General State-Space Markov Chain Theory," in *Markov Chain Monte Carlo in Practice*, ed. by Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. London: Chapman & Hall/CRC, 59–74. [1]

WHITE, HALBERT (2001): *Asymptotic Theory for Econometricians*. Revised Edition, London: Academic Press. [19]

YU, BIN (1994): "Rates of Convergence for Empirical Processes of Stationary Mixing Sequences," *Annals of Probability*, 22, 94–116. [7]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 22 April, 2019; final version accepted 15 February, 2022; available online 4 March, 2022.