Asymptotic Properties of the Maximum Likelihood Estimator in Endogenous Regime-Switching Models*

Chaojun Li **

Indiana University

Yan Liu***

Kyoto University

April 17, 2019

Abstract

This study proves the asymptotic properties of the maximum likelihood estimator (MLE) in a wide range of endogenous regime-switching models. This class of models extends the constant state transition probability in Markov-switching models to a time-varying probability that includes information from observations. A feature of importance in this proof is the mixing rate of the state process conditional on the observations, which is time varying owing to the time-varying transition probabilities. Consistency and asymptotic normality follow from the almost deterministic geometric decaying bound of the mixing rate. Relying on low-level assumptions that have been shown to hold in general, this study provides theoretical foundations for statistical inference in most endogenous regime-switching models in the literature. Monte Carlo simulation studies are conducted to examine the behavior of the MLE in finite samples.

*We are grateful to Yoosoon Chang, Yoshihiko Nishiyama, and Joon Y. Park for their guidance in writing the paper. All errors are our own.

**Department of Economics, Wylie Hall 105, 100 S. Woodlawn Ave Bloomington, IN 47405-7104, USA. Email: li469@iu.edu

***Graduate School of Economics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan. Email: liu.yan.36a@st.kyoto-u.ac.jp

1 Introduction

Regime-switching models have been applied extensively since Hamilton (1989) to study how time-series patterns change across different underlying economic states, such as boom and recession, high-volatility and low-volatility financial market environments, and active and passive monetary and fiscal policies. This class of models features a bivariate process (S_t, Y_t), where (S_t) is an unobservable Markov chain determining the regime in each period, and (Y_t) is an observable process whose conditional distribution is governed by the underlying state (S_t). The concept of Markov switching has been introduced in a broad class of time-series econometric models, including but not limited to Markov-switching vector autoregressions (Krolzig (1997)), the switching autoregressive conditional heteroskedasticity model (Cai (1994); Hamilton and Susmel (1994)), and the Markov-switching generalized autoregressive conditional heteroskedasticity model (Gray (1996); Klaassen (2002)).

In basic Markov-switching models, like Hamilton (1989), the unobservable state process is assumed to follow a homogeneous Markov chain. The implication is that the transition probability and expected duration of each regime are constant through all periods, regardless of the level of the observable and how long the regime has lasted. The setup is restrictive in application and conflicts with empirical findings, such as those of Watson (1992) and Filardo and Gordon (1998). Diebold, Lee, and Weinbach (1994) extended the state transition probability to be time varying by allowing it to depend on the predetermined variables (X_t), which are often chosen as economic covariates for predicting the regime change. This class of models, usually referred to as time-varying transition probability regime-switching models, is widely used in many empirical studies (see, e.g., Filardo (1994), Bekaert and Harvey (1995), Gray (1996), Filardo and Gordon (1998), Ang, Bekaert, and Wei (2008)). Such models, however, still assume that the evolution of the underlying state is independent of the other parts of the model. Chib and Dueker (2004), Kim, Piger, and Startz (2008), and Chang, Choi, and Park (2017) proposed endogenous regime-switching models in which the determination of states depends on the realizations of the observable process. The model of particular interest to this study is Chang et al. (2017), [1] which allows the transition of states to depend

[1] Chib and Dueker (2004) used a Bayesian estimation method and thus is beyond the scope of this study. We do not consider Kim et al. (2008), since its determination of the current state depends on the current shock to the observable process Y_t. When the regime is determined, Y_t is a determined one-point realization. We cannot interpret the model in the conventional way in which the observable process follows a particular pattern in some regime, as we

on past realizations of the observable (Y_t). In their model, accumulation of sustained positive (or negative) shocks to the observable might pressure the regime to change.

Compared to the large body of empirical work on regime-switching models, there are few studies on the asymptotic properties of their maximum likelihood estimator (MLE). Douc, Moulines, and Ryden (2004) and Kasahara and Shimotsu (2018) [2] showed the asymptotic properties of the MLE only with basic Markov-switching models. Ailliot and Pene (2015) established consistency of the MLE in models with time-inhomogeneous Markov regimes. Pouzo, Psaradakis, and Sola (2018) showed the asymptotic properties of the MLE in autoregressive models with time-inhomogeneous Markov regime switching and possible model misspecifications. Their model set-ups encompass endogenous regime switching in the sense that they allow the transition kernel of the state to depend on past realizations of the observable. Pouzo et al. (2018) showed that under certain assumptions, the MLE converges to a pseudo-true parameter set even if the model is misspecified. However, the authors restricted the observable time series to depend only on the current state, and the state process to be only a first-order inhomogeneous Markov chain. Such restrictions make their theory less applicable.

The aim of this study is to show the asymptotic properties of the MLE in a wide range of endogenous regime-switching models, including Chang et al. (2017) as a special case. The model is general enough to incorporate the time-varying transition probability regime-switching models represented by Diebold et al. (1994). The model we consider is more general than that of Douc et al. (2004) and Kasahara and Shimotsu (2018) in that it allows dependence of state transition probability on past realizations and other economic fundamentals. Our model is more general than that of Pouzo et al. (2018) in that it allows the transition of the observable and the state process to depend on more than one lag of the processes. Moreover, instead of dealing with misspecification in the transition probability as in Pouzo et al. (2018), this study assumes that the likelihood function is concentrated. The advantage is that this study relies on assumptions of lower level than Pouzo

can in basic regime-switching models.

[2]Baum and Petrie (1966), Leroux (1992), Bickel and Ritov (1996), Bickel, Ritov, and Ryden (1998), Jensen and Petersen (1999), Le Gland and Mevel (2000) and Douc and Matias (2001) contribute to the asymptotic theories with the basic Markov models less general than Douc et al. (2004) and Kasahara and Shimotsu (2018). Their models, usually referred to as hidden Markov models, do not allow autoregression and (Y_t) are conditionally independent given the current state.

et al. (2018) did. In Section 6, we show that our assumptions hold in some widely used endogenous regime-switching models. Thus, this study provides theoretical foundations for statistical inference of endogenous regime-switching models in most empirical research. Some interesting and important statistical tests can now be conducted, such as whether the observables affect transition probability and whether the effect is positive or negative.

The general difficulty in the proof of asymptotic theories with regime-switching models is that the predictive densities of the observable given past realizations do not form a stationary sequence, and thus, the ergodic theorem does not directly apply. To deal with the problem, this study, following Douc et al. (2004) and Kasahara and Shimotsu (2018), approximates the log-likelihood function by the partial sum of a stationary ergodic sequence. The cornerstone of the approximation is the almost surely geometrically decaying bound of the mixing rate of the conditional chain $S|Y, X$. The time-varying state transition probability makes it more complex to show the bound, because it enters the mixing rate and causes the rate to approach unity as the transition probability approaches zero. By contrast, the constant state transition probability is bounded away from zero in Douc et al. (2004) and Kasahara and Shimotsu (2018). The main theoretical contribution of this study is that we show the mixing rate is bounded away from zero eventually by assuming that there is a small probability that the observable takes extreme values, which is vital for our findings of consistency and asymptotic normality of the MLE.

The rest of this paper is organized as follows. Section 2 lists the main assumptions and examples of endogenous regime-switching models. Section 3 shows the mixing rate of the conditional chain and the approximation of the log-likelihood function with an ergodic stationary process. Sections 4 and 5 show the consistency and asymptotic normality of the MLE, respectively. Section 6 discusses the assumptions in detail and Section 7 shows the simulation results. Section 8 concludes.

2 Models, notations, and assumptions

2.1 Models

Endogenous regime-switching models can generally be defined by a transition equation of the observed process and a state transition probability allowed to depend on past observable $Y_{t-1}, \ldots, Y_{t-r}$:

transition equation of the observed process:

$$Y_t = f_\theta(S_t, \ldots, S_{t-r+1}, Y_{t-1}, \ldots, Y_{t-r}, X_t; U_t)^3 \tag{1}$$

state transition probability:

$$q_\theta(S_t | S_{t-1}, \ldots, S_{t-r}, Y_{t-1}, \ldots, Y_{t-r}, X_t)^4 \tag{2}$$

where X_t is a predetermined variable (vector), (U_t) is an independent and identically distributed (i.i.d.) sequence of random variables, f_θ is a family of functions indexed by θ, and q_θ is a family of probabilities indexed by θ. Pouzo et al. (2018) dealt with the model with $r = 1$. [5] The special case in which the transition probability in (2) depends only on $(S_{t-1}, \ldots, S_{t-r})$ reduces to the basic Markov-switching model. The case in which (2) depends on $(S_{t-1}, \ldots, S_{t-r}, X_t)$ is the time-varying transition probability regime-switching model represented by Diebold et al. (1994).

Some widely applied transition equations of the observed process are summarized as

$$Y_t = m(Y_{t-1}, \ldots, Y_{t-k}, S_t, \ldots, S_{t-k}, X_t) + \sigma(S_t, \ldots, S_{t-k})U_t = m_t + \sigma_t U_t. \tag{3}$$

[3] The model here is more general than it may appear by allowing for different numbers of lags of (S_t) and (Y_t) as $f_\theta(S_t, \ldots, S_{t-p+1}, Y_{t-1}, \ldots, Y_{t-q}, X_t; U_t)$. If, without loss of generality, $p \leq q$, then we can make an innocuous change by including more lags as $f_\theta(S_t, \ldots, S_{t-p+1}, S_{t-p}, \ldots, S_{t-q+1}, Y_{t-1}, \ldots, Y_{t-q}, X_t; U_t)$. Then, this is the model in (1).

[4] The transition probability here can accommodate different numbers of lags in (S_t) and (Y_t) as well as different numbers of lags from those in (1) by making similar changes to the previous comment.

[5] The state transition probability in Pouzo et al. (2018) cannot be generalized easily to accommodate more lags by defining the state variable S_t as a vector $(\tilde{S}_t, \tilde{S}_{t-1}, \ldots, \tilde{S}_{t-r})$ because their Assumption 1 requires $q_\theta(S_t | S_{t-1}, Y_{t-1}) > 0$ for all S_t and S_{t-1}. This is violated when, for $r = 2$, $q_\theta(S_t = (s_1, s_1) | S_{t-1} = (s_2, s_2), Y_{t-1}) = 0$. Similarly, the transition equation of the observed process cannot be generalized to accommodate more lags owing to their Assumption 5.

An example of (3) is the autoregressive model with switching in mean and volatility:

$$\gamma(L)(Y_t - \mu_t) = \gamma_X' X_t + \sigma_t U_t \tag{4}$$

where $\gamma(z) = 1 - \gamma_1 z - \cdots - \gamma_k z^k$, $\mu_t = \mu(S_t)$, $\sigma_t = \sigma(S_t)$, $S_t = 1, 2, 3, \ldots,$ or J.

Another example is the autoregressive model with state-dependent autoregression coefficients:

$$Y_t = \mu(S_t) + \gamma_1(S_t)Y_{t-1} + \cdots + \gamma_k(S_t)Y_{t-k} + \gamma_X(S_t)' X_t + \sigma(S_t)U_t. \tag{5}$$

For the state transition probability, we mainly consider the specifications proposed by Diebold et al. (1994) and Chang et al. (2017) as examples. Although the transition probability in Diebold et al. (1994) does not depend on past observations of Y_t and thus, is strictly not an endogenous regime-switching model, the reason we include it here is twofold. First, the model is widely used but its asymptotic properties have not been fully discussed. Second, the model can be easily extended to endogenous regime-switching models by including past realizations of the observable as the predetermined variables.

State transition 1 (Diebold et al. (1994)) Transition probabilities are functions mapping X_t to $[0, 1]$. In the case in which there are two states $S_t = 0$ or 1,

$$q_\theta(S_t = s_t | S_{t-1} = s_{t-1}, X_t) = \begin{pmatrix} p_{00}(X_t) & 1 - p_{00}(X_t) \\ 1 - p_{11}(X_t) & p_{11}(X_t) \end{pmatrix}. \tag{6}$$

Functions used most often are logistic functions and probit functions. The case in which $p_{00}(X_t)$ and $p_{11}(X_t)$ are constant functions reduces to the basic Markov-switching model.

State transition 2. (Chang et al. (2017)) Chang et al. (2017) proposed a new approach to modeling switching when there are two regimes by using an autoregressive latent factor to determine regimes, depending on whether the factor takes a value above or below a threshold τ. The latent factor follows an AR(1) process

$$W_t = \alpha W_{t-1} + V_t$$

for $t = 1, 2, \ldots$ with parameter $\alpha \in [-1, 1]$ and i.i.d. standard normal innovations (V_t). (U_t) and (V_t) are jointly i.i.d. and distributed as

$$\begin{pmatrix} U_t \\ V_{t+1} \end{pmatrix} =_d \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

The regime is decided by

$$S_t = \begin{cases} 1 & \text{if } W_t \geq \tau \\ 0 & \text{if } W_t < \tau \end{cases}.$$

The correlation ρ between innovations to the latent factor and *lagged* innovations to the observed time series connects the state transition probability to the observed time series. A positive correlation means that the information raising the level of the observed time series makes the economy more likely to be in the high regime ($S_t = 1$) in the future. A negative correlation works in the opposite direction. A zero-correlation model reduces to the basic Markov-switching model.

Theorem 3.1 in Chang et al. (2017) clarifies the state transition probability. The theorem states that with the transition equation of the observed process (3), when $|\alpha| < 1$, $|\rho| < 1$, (S_t, Y_t) together follow a $(k+1)$th-order Markov process, and the state transition probability is

$$q_\theta(S_t|S_{t-1}, \ldots, S_{t-k-1}, Y_{t-1}, \ldots, Y_{t-k-1}) = (1 - S_t)\omega_\rho + S_t(1 - \omega_\rho) \tag{7}$$

with $\omega_\rho = \omega_\rho(S_{t-1}, \ldots, S_{t-k-1}, Y_{t-1}, \ldots, Y_{t-k-1})$ defined as

$$\omega_\rho = \frac{[(1 - S_{t-1})\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} + S_{t-1}\int_{\tau\sqrt{1-\alpha^2}}^{\infty}]\Phi\left(\frac{\tau - \rho U_{t-1}}{\sqrt{1-\rho^2}} - \frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx}{(1 - S_{t-1})\Phi(\tau\sqrt{1-\alpha^2}) + S_{t-1}[1 - \Phi(\tau\sqrt{1-\alpha^2})]} \tag{8}$$

where $U_t = \frac{Y_t - m_t}{\sigma_t}$, $\Phi(\cdot)$ and $\varphi(\cdot)$ are the distribution function and the density function of the standard normal distribution,respectively. The state transition probability fits (2) with $r = k + 1$.

2.2 Notations and assumptions

Let $\triangleq$ denote "equals by defination." In the following proof, for short notation, we define

$$\mathbf{Y}_m^n \triangleq (Y_n, Y_{n-1}, \ldots, Y_m)' \text{ for } n \geq m,$$
$$\overline{\mathbf{Y}}_t \triangleq \mathbf{Y}_{t-r+1}^t = (Y_t, Y_{t-1}, \ldots, Y_{t-r+1})',$$
$$\overline{\mathbf{Y}}_m^n \triangleq (\overline{\mathbf{Y}}_n, \ldots, \overline{\mathbf{Y}}_m)' \text{ for } n \geq m,$$

and similarly for S_t, X_t and realizations s_t, y_t, and x_t.

We assume that $\{S_t\}_{t=-r+1}^{\infty}$ takes a value in a discrete set $\mathbb{S}$ with J elements. Let $\overline{\mathbb{S}} \triangleq \mathbb{S}^r$, and use $\mathcal{P}(\overline{\mathbb{S}})$ to denote the power set of $\overline{\mathbb{S}}$. For each $t \geq 1$ and given $(\mathbf{Y}_{t-r}^{t-1}, \mathbf{S}_{t-r}^{t-1}, X_t)$, S_t is conditionally independent of $(\mathbf{Y}_{-r+1}^{t-r-1}, \mathbf{S}_{-r+1}^{t-r-1}, \mathbf{X}_1^{t-1}, \mathbf{X}_{t+1}^{\infty})$. The transition probability is $q_\theta(s|\overline{\mathbf{S}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)$. $\{Y_t\}_{t=-r+1}^{\infty}$ takes a value in a set $\mathbb{Y}$, which is separable and metrizable by a complete metric. Let $\overline{\mathbb{Y}} \triangleq \mathbb{Y}^r$. For each $t \geq 1$ and given $(\mathbf{Y}_{t-r}^{t-1}, \mathbf{S}_{t-r+1}^{t}, X_t)$, Y_t is independent of $(\mathbf{Y}_{-r+1}^{t-r-1}, \mathbf{S}_{-r+1}^{t-r}, \mathbf{X}_1^{t-1}, \mathbf{X}_{t+1}^{\infty})$. The conditional law has a density $g_\theta(y|\overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{S}}_t, X_t)$ with respect to some fixed $\sigma-$finite measure ν on the Borel $\sigma-$field $\mathcal{B}(\mathbb{Y})$. $\{X_t\}_{t=1}^{\infty}$ takes a value in a set $\mathbb{X}$. Conditionally on X_t, $\{X_k\}_{k \geq t+1}$ is independent of $\{Y_k\}_{k \leq t}$ and $\{S_k\}_{k \leq t}$. Conditionally on X_t, $\{X_k\}_{k \leq t-1}$ is independent of $\{Y_k\}_{k \geq t}$ and $\{S_k\}_{k \geq t}$.

Under the setup, conditional on $\mathbf{X}_1^{\infty}$, (S_t, Y_t) follows a Markov chain of order r with transition density

$$\begin{aligned} & p_\theta(S_t, Y_t|S_{t-1}, \ldots, S_{-r+1}, Y_{t-1}, \ldots, Y_{-r+1}, X_t) \\ =& g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{S}}_t, X_t) q_\theta(S_t|\overline{\mathbf{S}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) = p_\theta(S_t, Y_t|\overline{\mathbf{S}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t). \end{aligned}$$

It also follows that for $1 \leq t \leq n$,

$$p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{S}}_0 = \overline{\mathbf{s}}_0, \mathbf{X}_1^n) = p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{S}}_0 = \overline{\mathbf{s}}_0, \mathbf{X}_1^t),$$
$$p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \mathbf{X}_1^n) = p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \mathbf{X}_1^t).$$

This study works with the conditional likelihood function given initial observations $\overline{\mathbf{Y}}_0 = (Y_0, \ldots, Y_{-r+1})$, (unobservable) initial state $\overline{\mathbf{S}}_0 = (S_0, \ldots, S_{-r+1})$, and predetermined variables

$\mathbf{X}_1^t$, owing to the difficulties obtaining the closed-form expression of the unconditional stationary likelihood function. We can write the conditional log-likelihood function as

$$\ell_n(\theta, \bar{\mathbf{s}}_0) = \log p_\theta(Y_1, \dots, Y_n | \overline{\mathbf{Y}}_0, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^n) = \sum_{t=1}^{n} \log p_\theta(Y_t | \overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^t) \tag{9}$$

with predictive density

$$\begin{aligned} & p_\theta(Y_t | \overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0, \mathbf{X}_1^t) \\ &= \sum_{s_t, \bar{\mathbf{s}}_{t-1}} g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t) q_\theta(s_t | \bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) \mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1} = \bar{\mathbf{s}}_{t-1} | \overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^{t-1}). \end{aligned} \tag{10}$$

When the number of observations is $n + r$, we condition on the first r observations, and arbitrarily choose the initial state $\bar{\mathbf{s}}_0$. The aim of this study is to show consistency and asymptotic normality of the MLE $\hat{\theta}_{n,\bar{\mathbf{s}}_0} = \arg\max_\theta \ell_n(\theta, \bar{\mathbf{s}}_0)$ with any choice of $\bar{\mathbf{s}}_0$, even when it is not the true underlying initial state.

The following are the basic assumptions.

(A1) The parameter θ belongs to Θ. Θ is compact. Let θ^* denote the true parameter. θ^* lies in the interior of Θ.

(A2) $(\overline{\mathbf{S}}_t, \overline{\mathbf{Y}}_t, X_t)$ is a strictly stationary ergodic process.

Remark. We can extend (Y_t, X_t) with doubly infinite time $\{Y_t, X_t\}_{-\infty}^{+\infty}$. According to Theorem 7.1.3 in Durrett (2013), for measurable function $f(\cdot)$, $f(Y_t, Y_{t+1}, \dots, X_t, X_{t+1}, \dots)$ is ergodic.

(A3) (a) $\sigma_-(\bar{\mathbf{y}}_0, x_1) := \inf_\theta \min_{s_1 \in \mathbb{S}, \bar{\mathbf{s}}_0 \in \overline{\mathbb{S}}} q_\theta(s_1 | \bar{\mathbf{s}}_0, \bar{\mathbf{y}}_0, x_1) > 0$, for all $\bar{\mathbf{y}}_0 \in \overline{\mathbb{Y}}$ and $x_1 \in \mathbb{X}$.

(b) $b_-(\mathbf{y}_{-r+1}^1, x_1) := \inf_\theta \min_{\bar{\mathbf{s}}_1 \in \overline{\mathbb{S}}} g_\theta(y_1 | \bar{\mathbf{y}}_0, \bar{\mathbf{s}}_1, x_1) > 0$, for all $\mathbf{y}_{-r+1}^1 \in \mathbb{Y}^{r+1}$ and $x_1 \in \mathbb{X}$.

(c) $b_+ := \sup_\theta \sup_{y_1, \bar{\mathbf{y}}_0, x_1} \max_{\bar{\mathbf{s}}_1} g_\theta(y_1 | \bar{\mathbf{y}}_0, \bar{\mathbf{s}}_1, x_1) < \infty$, and $\mathbb{E}_{\theta^*} | \log b_-(\mathbf{y}_{-r+1}^1, x_1)| < \infty$.

(A4) (a) Constants $\alpha_1 > 0$, $C_1, C_2 \in (0, +\infty)$, and $\beta_1 > 1$ exist such that, for any $\xi > 0$,

$$\mathbb{P}_{\theta^*}\left(\sigma_-(\overline{\mathbf{Y}}_0, X_1) \leq C_1 e^{-\alpha_1 \xi}\right) \leq C_2 \xi^{-\beta_1}. \tag{11}$$

(b) Constants $\alpha_2 > 0$, $C_3, C_4 \in (0, +\infty)$, and $\beta_2 > 1$ exist such that, for any $\xi > 0$,

$$\mathbb{P}_{\theta^*}\left(b_-(\mathbf{Y}_{-r+1}^1, X_1) \leq C_3 e^{-\alpha_2 \xi}\right) \leq C_4 \xi^{-\beta_2}. \tag{12}$$

Remark. Assumptions (A3b), (A3c), and (A4b) are the same as in Kasahara and Shimotsu (2018). Assumption (A3a) parallels Assumption (A1) in Douc et al. (2004) and Assumption 1(d) in Kasahara and Shimotsu (2018), but $\sigma_-(\cdot)$ here is time varying depending on the observations, whereas the value is constant in Douc et al. (2004) and Kasahara and Shimotsu (2018). Assumption (A4a) refers to the low probability of $\sigma_-(\overline{\mathbf{y}}_0, x_1)$ taking an extremely small value. This is the key assumption that we use in Lemma 3 to establish the geometric bound. It serves the same function as Assumption 4 in Pouzo et al. (2018) but is of a lower level. Their assumption parallels Lemma 2 in the following section. We will need the following additional assumptions on the continuity of q_θ and g_θ and the identifiability of θ^*:

(A5) For all $(s', \overline{\mathbf{s}}) \in \mathbb{S} \times \overline{\mathbb{S}}$, $(\overline{\mathbf{y}}, y') \in \overline{\mathbb{Y}} \times \mathbb{Y}$, and $x \in \mathbb{X}$, $\theta \to q_\theta(s'|\overline{\mathbf{s}}, \overline{\mathbf{y}}, x)$ and $\theta \to g_\theta(y'|\overline{\mathbf{y}}, \overline{\mathbf{s}}, x)$ are continuous.

(A6) θ and θ^* are identical (up to a permutation of state indexes) if and only if $\mathbb{P}_\theta(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n) = \mathbb{P}_{\theta^*}(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n)$ for all $n \geq 1$.

Remark. Section 6 shows that this assumption is satisfied in many examples with the identifiability of the mixture normal distributions. By contrast, the identifiability assumptions in Douc et al. (2004) and Kasahara and Shimotsu (2018) are difficult to verify.

3 Approximation with stationary ergodic sequence

Consistency and asymptotic normality follows if we can show the following two results:

(R1) the normalized log-likelihood $n^{-1}\ell_n(\theta, \overline{\mathbf{s}}_0)$ converges to a deterministic function $\ell(\theta)$ uniformly with respect to θ, and θ^* is a well-separated point of maximum of $\ell(\theta)$; and

(R2) a central limit theorem for the Fisher score function and a locally uniform law of large numbers for the observed Fisher information.

The updating distribution $\mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}_{t-1}|\overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0, \mathbf{X}_1^{t-1})$ in (10) is not stationary ergodic and thus, the predictive density is not stationary ergodic. Therefore, the ergodic theorem cannot be applied directly to conclude a law of large numbers or central limit theorems. Here, we follow Douc et al. (2004) and Kasahara and Shimotsu (2018) to first approximate the predictive density with a stationary ergodic sequence

$$
\begin{aligned}
& p_\theta(Y_t|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}) \\
&= \sum_{s_t, \overline{\mathbf{s}}_{t-1}} g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{s}}_t, X_t) q_\theta(s_t|\overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) \mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}_{t-1}|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}).
\end{aligned} \tag{13}
$$

(13) differs from (10) in that the updating distribution part now depends on the whole history of (Y_t, X_t) from past infinity and does not depend on the initial state. The approximation builds on the almost surely geometric decaying bound of the mixing rate of the conditional chain $(S|Y, X)$. The bound guarantees that the influence of observation and initial state far in the past quickly vanishes and thus, the difference between the exact predictive and approximated predictive log densities becomes asymptotically neglible.

The mixing rate in Corollary 1 follows from the Markov property of the conditional chain and the minorization condition in Lemma 1. Lemma 2 establishes (almost) deterministic bounds for the time-varying mixing rate. Lemma 3 shows the approximation of the log predictive density $\log p_\theta(Y_t|\overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0 = \overline{\mathbf{s}}, \mathbf{X}_1^t)$ with the stationary one $\log p_\theta(Y_t|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1})$. For short notation, we define the following predictive log densities:

$$
\begin{aligned}
\Delta_{t,m,\overline{\mathbf{s}}}(\theta) &\triangleq \log p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1}, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m+1}^t), \\
\Delta_{t,m}(\theta) &\triangleq \log p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m+1}^t).
\end{aligned}
$$

LEMMA 1 (Minorization condition)**.** *Let $m, n \in \mathbb{Z}$ with $-m \leq n$ and $\theta \in \Theta$. Conditionally on $\overline{\mathbf{Y}}_{-m}^n$ and $\mathbf{X}_{-m}^n$, $\{\overline{\mathbf{S}}_k\}_{-m \leq k \leq n}$ satisfies the Markov property. Assume (A3). Then, for all $-m + r \leq k \leq n$, a function $\mu_k(\mathbf{Y}_{k-r}^n, \mathbf{X}_k^n, A)$ exists, such that:*

(i) for any $A \in \mathcal{P}(\overline{\mathbb{S}})$, $(\mathbf{y}_{k-r}^n, \mathbf{x}_k^n) \to \mu_k(\mathbf{y}_{k-r}^n, \mathbf{x}_k^n, A)$ is a Borel function; and

(ii) for any $\mathbf{y}_{k-r}^n$ and $\mathbf{x}_k^n$, $\mu_k(\mathbf{y}_{k-r}^n, \mathbf{x}_k^n, \cdot)$ is a probability measure on $\mathcal{P}(\overline{\mathbb{S}})$. Moreover, for $A \in \mathcal{P}(\overline{\mathbb{S}})$,

the following holds:

$$\min_{\bar{\mathbf{s}}_{k-r}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in A|\overline{\mathbf{S}}_{k-r} = \bar{\mathbf{s}}_{k-r}, \overline{\mathbf{Y}}^n_{-m}, \mathbf{X}^n_{-m})$$
$$\geq \omega(\mathbf{Y}^{k-1}_{k-2r+1}, \mathbf{X}^k_{k-r+1}) \cdot \mu_k(\mathbf{Y}^n_{k-r}, \mathbf{X}^n_k, A)$$

where $\omega(\mathbf{Y}^{k-1}_{k-2r+1}, \mathbf{X}^k_{k-r+1}) := \frac{\prod_{\ell=k-r+1}^{k-1} b_-(\mathbf{Y}^\ell_{\ell-r}, X_\ell) \prod_{\ell=k-r+1}^{k} \sigma_-(\overline{\mathbf{Y}}_{\ell-1}, X_\ell)}{b_+^{r-1}}$.

The proof is in the appendix. For any probability measures μ_1 and μ_2, we define the total variation distance $\|\mu_1 - \mu_2\|_{TV} = \sup_A |\mu_1(A) - \mu_2(A)|$. For any $x \in \mathbb{R}^+$, let $\lfloor x \rfloor$ denote the largest integer that is smaller than x. The following corollary bounds the distance between the two conditional Markov chains starting from initial distributions μ_1 and μ_2.

COROLLARY 1 (Uniform ergodicity)**.** *Assume (A3). Let* $m, n \in \mathbb{Z}$, $-m \leq n$ *and* $\theta \in \Theta$*. Then, for* $-m \leq k \leq n$*, for all probability measures* μ_1 *and* μ_2 *defined on* $\mathcal{P}(\overline{\mathbb{S}})$ *and for all* $\overline{\mathbf{Y}}^n_{-m}$ *and* $\overline{\mathbf{X}}^n_{-m}$*,*

$$\left\| \sum_{\bar{\mathbf{s}}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \overline{\mathbf{Y}}^n_{-m}, \mathbf{X}^n_{-m})\mu_1(\bar{\mathbf{s}}) - \sum_{\bar{\mathbf{s}}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \overline{\mathbf{Y}}^n_{-m}, \mathbf{X}^n_{-m})\mu_2(\bar{\mathbf{s}}) \right\|_{TV}$$
$$\leq \prod_{i=1}^{\lfloor (k+m)/r \rfloor} \left(1 - \omega\left(\mathbf{Y}^{-m+ri-1}_{-m+ri-2r+1}, \mathbf{X}^{-m+ri}_{-m+ri-r+1}\right)\right) = \prod_{i=1}^{\lfloor (k+m)/r \rfloor} (1 - \omega(\mathbf{V}_{-m+ri}))$$

with $\mathbf{V}_t := (\mathbf{Y}^{t-1}_{t-2r+1}, \mathbf{X}^t_{t-r+1})$.

The proof follows from Lemma 7 in Kasahara and Shimotsu (2018). The corollary parallels Corollary 1 in Douc et al. (2004) and Corollary 1 in Kasahara and Shimotsu (2018). The main difference is that the mixing rate contains the term $\sigma_-(\cdot)$ in $\omega(\mathbf{V}_t)$, and can approach 1 as $\sigma_-(\cdot)$ approaches 0. The following lemma shows that we can bound the mixing rate almost deterministically under Assumptions (A3) and (A4), and enables us to show the approximation in Lemma 3.

LEMMA 2 (Bound for mixing rate)**.** *Assume (A2)–(A4). Then,* $\varepsilon_0 \in (0, \frac{1}{16r})$ *and* $\rho \in (0, 1)$ *exist* *such that for all* $m, n \in \mathbb{Z}^+$,

$$\mathbb{P}_{\theta^*}\left(\prod_{k=1}^{n}(1-\omega(\mathbf{V}_{t_k})) \leq \rho^{(1-2\varepsilon_0)n} \ ev.\right) = 1,^6 \tag{14}$$

$$\mathbb{P}_{\theta^*}\left(\omega(\mathbf{V}_k) \geq C_5 \rho^{\varepsilon_0 \lfloor (k-r)/r \rfloor} \ ev.\right) = 1, \tag{15}$$

$$\tag{16}$$

where $C_5 = \frac{C_1^r C_3^{r-1}}{b_+^{r-1}}$, *and* $\{t_k\}_{1 \leq k \leq n}$ *is a sequence of integers such that* $t_k \neq t_{k'}$ *for* $1 \leq k, k' \leq n$ *and* $k \neq k'$.

It also holds that

$$\mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}(1-\omega(\mathbf{V}_{t_k}))^m\right] \leq 2\rho^n, \tag{17}$$

$$\mathbb{E}_{\theta^*}\left[\prod_{k_1=1}^{n_1}(1-\omega(\mathbf{V}_{t_{k_1}}))^m \wedge \prod_{k_2=1}^{n_2}(1-\omega(\mathbf{V}_{t_{k_2}}))^m\right] \leq 2(\rho^{n_1} \wedge \rho^{n_2}), \tag{18}$$

$$\mathbb{E}_{\theta^*}\left[\prod_{k_1=1}^{n_1}(1-\omega(\mathbf{V}_{t_{k_1}}))^m \wedge \prod_{k_2=1}^{n_2}(1-\omega(\mathbf{V}_{t_{k_2}}))^m \wedge \prod_{k_3=1}^{n_3}(1-\omega(\mathbf{V}_{t_{k_3}}))^m\right] \leq 2(\rho^{n_1} \wedge \rho^{n_2} \wedge \rho^{n_3}). \tag{19}$$

LEMMA 3. *Assume (A2)–(A4). Then, for* $m' \geq m \geq 0$,

$$\mathbb{P}_{\theta^*}\left(\sup_{\theta \in \Theta} \max_{\bar{\mathbf{s}}_i, \bar{\mathbf{s}}_j \in \bar{\mathbb{S}}} |\Delta_{t,m,\bar{\mathbf{s}}_i}(\theta) - \Delta_{t,m',\bar{\mathbf{s}}_j}(\theta)| \leq \frac{1}{C_5}\rho^{\lfloor (t+m)/3r \rfloor} \ ev.\right) = 1, \tag{20}$$

$$\mathbb{P}_{\theta^*}\left(\sup_{\theta \in \Theta} \max_{\bar{\mathbf{s}}_i \in \bar{\mathbb{S}}} |\Delta_{t,m,\bar{\mathbf{s}}_i}(\theta) - \Delta_{t,m}(\theta)| \leq \frac{1}{C_5}\rho^{\lfloor (t+m)/3r \rfloor} \ ev.\right) = 1, \tag{21}$$

$$\sup_{\theta \in \Theta} \sup_{m \geq 0} \max_{\bar{\mathbf{s}} \in \bar{\mathbb{S}}} |\Delta_{t,m,\bar{\mathbf{s}}}(\theta)| \leq \max\left\{|\log b_+|, \left|\log\left(b_-(\mathbf{Y}_{t-r}^t, X_t)\right)\right|\right\}. \tag{22}$$

From (20), $\{\Delta_{t,m,\bar{\mathbf{s}}}(\theta)\}$ is a uniformly Cauchy sequence with respect to θ $\mathbb{P}_{\theta^*}-$ almost surely (a.s.).

We define $\Delta_{t,\infty}(\theta) \triangleq \lim_{m \to \infty} \Delta_{t,m,\bar{\mathbf{s}}}(\theta)$. From (22), $\{\Delta_{t,m,\bar{\mathbf{s}}}(\theta)\}$ is uniformly bounded in $L^1(\mathbb{P}_{\theta^*})$,

[6] *ev.* is an abbreviation for "eventually."

and thus, $\Delta_{t,\infty}(\theta)$ is also in $L^1(\mathbb{P}_{\theta^*})$. Let $m = 0$ and $m' \to \infty$ in (20), which yields

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} \max_{\bar{\mathbf{s}}_0\in\bar{\mathbb{S}}} |\Delta_{t,0,\bar{\mathbf{s}}_0}(\theta) - \Delta_{t,\infty}(\theta)| \to 0 \quad \mathbb{P}_{\theta^*} - a.s. \tag{23}$$

From (21), $\lim_{m\to\infty}\Delta_{t,m,\bar{\mathbf{s}}}(\theta) = \lim_{m\to\infty}\Delta_{t,m}(\theta)$.

Moreover, $\lim_{m\to\infty}\Delta_{t,m}(\theta) = \log p_\theta(Y_t|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t})$, since

$$\begin{aligned}
&\lim_{m\to\infty}\Delta_{t,m}(\theta) = \lim_{m\to\infty}\log p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m+1}^{t})\\
&=\log\sum_{s_t,\bar{\mathbf{s}}_{t-1}}\big(g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t)q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)\times\lim_{m\to\infty}\mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1}=\bar{\mathbf{s}}_{t-1}|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m+1}^{t})\big)\\
&=\log\sum_{s_t,\bar{\mathbf{s}}_{t-1}} g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t)q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)\mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1}=\bar{\mathbf{s}}_{t-1}|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t})\\
&=\log p_\theta(Y_t|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t}).
\end{aligned}$$

Then, there are two expressions for $\Delta_{t,\infty}(\theta)$. In Section 4, we use $\lim_{m\to\infty}\Delta_{t,m,\bar{\mathbf{s}}}(\theta)$ to derive continuity, and use $\lim_{m\to\infty}\Delta_{t,m}(\theta)$ to show convergence. Since $\log p_\theta(Y_t|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t})$ is strictly stationary ergodic, it holds that

$$n^{-1}\sum_{t=1}^{n}\Delta_{t,\infty}(\theta) \to \ell(\theta) \triangleq \mathbb{E}_{\theta^*}[\Delta_{0,\infty}(\theta)] \quad \mathbb{P}_{\theta^*} - a.s. \tag{24}$$

4 Consistency

This section shows the consistency in Theorem 1 based on uniform convergence (Proposition 1) and identification (Proposition 2).

PROPOSITION 1 (Uniform convergence)**.** *Assume (A1)–(A5). Then,*

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|n^{-1}\ell_n(\theta, \bar{\mathbf{s}}_0) - \ell(\theta)| = 0 \quad \mathbb{P}_{\theta^*} - a.s.$$

Proof. By the compactness of Θ, it suffices to show $\forall\theta\in\Theta$,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\sup_{\theta':|\theta'-\theta|\le\delta}|n^{-1}\ell_n(\theta', \bar{\mathbf{s}}_0) - \ell(\theta)| = 0, \quad \mathbb{P}_{\theta^*} - a.s.$$

We decompose the difference as

$$\begin{aligned}&\limsup_{\delta\to 0}\limsup_{n\to\infty}\sup_{\theta':|\theta'-\theta|\le\delta}|n^{-1}\ell_n(\theta',\bar{\mathbf{s}}_0)-\ell(\theta)|\\&=\limsup_{\delta\to 0}\limsup_{n\to\infty}\sup_{\theta':|\theta'-\theta|\le\delta}|n^{-1}\ell_n(\theta',\bar{\mathbf{s}}_0)-\ell(\theta')|+\limsup_{\delta\to 0}\sup_{\theta':|\theta'-\theta|\le\delta}|\ell(\theta')-\ell(\theta)|.\end{aligned}$$

On the one hand, the first term

$$\begin{aligned}&\limsup_{\delta\to 0}\limsup_{n\to\infty}\sup_{\theta':|\theta'-\theta|\le\delta}|n^{-1}\ell_n(\theta',\bar{\mathbf{s}}_0)-\ell(\theta')|\\&\le\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta'\in\Theta}|\Delta_{t,0,\bar{\mathbf{s}}_0}(\theta')-\Delta_{t,\infty}(\theta')|+\limsup_{n\to\infty}\sup_{\theta'\in\Theta}|\frac{1}{n}\sum_{t=1}^{n}\Delta_{t,\infty}(\theta')-\ell(\theta')|\\&=0,\quad \mathbb{P}_{\theta^*}-a.s.\end{aligned}$$

follows from the approximation (23) and is a consequence of the ergodic theorem (24). On the other hand, note that $\ell(\theta)=\mathbb{E}_{\theta^*}[\Delta_{0,\infty}(\theta)]$ and $\Delta_{0,\infty}(\theta)$ is continuous if $\Delta_{0,m,\bar{\mathbf{s}}}(\theta)$ is continuous (see appendix) by uniform convergence. Then, the second term is bounded by

$$\begin{aligned}\limsup_{\delta\to 0}\sup_{\theta':|\theta'-\theta|\le\delta}|\ell(\theta')-\ell(\theta)|&\le\limsup_{\delta\to 0}\sup_{\theta':|\theta'-\theta|\le\delta}\mathbb{E}_{\theta^*}|\Delta_{0,\infty}(\theta')-\Delta_{0,\infty}(\theta)|\\&\le\limsup_{\delta\to 0}\mathbb{E}_{\theta^*}\left[\sup_{\theta':|\theta'-\theta|\le\delta}|\Delta_{0,\infty}(\theta')-\Delta_{0,\infty}(\theta)|\right]\\&=\mathbb{E}_{\theta^*}\left[\limsup_{\delta\to 0}\sup_{\theta':|\theta'-\theta|\le\delta}|\Delta_{0,\infty}(\theta')-\Delta_{0,\infty}(\theta)|\right]=0\end{aligned}$$

where the equality follows by the dominated convergence theorem. ■

PROPOSITION 2 (Identification)**.** *Under (A2) and (A6),* $\ell(\theta)\le\ell(\theta^*)$ *and* $\ell(\theta)=\ell(\theta^*)$ *if and only if* $\theta=\theta^*$*.*

The proof is in the appendix. The following theorem summarizes the finding and establishes consistency. The proof is similar to the scheme of Wald (1949), with slight modifications to include conditioning on $\bar{\mathbf{s}}_0$.

THEOREM 1. *Assume (A1)–(A6). Then, for any* $\bar{\mathbf{s}}_0\in\bar{\mathbb{S}}$*,* $\lim_{n\to\infty}\hat{\theta}_{n,\bar{\mathbf{s}}_0}=\theta^*$*,* $\mathbb{P}_{\theta^*}-a.s.$

Proof. As θ^* is a well-separated maximum of $\ell(\theta)$, we have $\forall\varepsilon>0$, $\sup_{\theta:|\theta-\theta^*|\ge\varepsilon}\ell(\theta)<\ell(\theta^*)$. For

ε, $\delta > 0$ exists such that $|\theta - \theta^*| > \varepsilon$ implies $\ell(\theta) < \ell(\theta^*) - \delta$, and thus,

$$\mathbb{P}_{\theta^*}(|\hat{\theta}_{n,\bar{\mathbf{s}}_0} - \theta^*| > \varepsilon) \leq \mathbb{P}_{\theta^*}(\ell(\hat{\theta}_{n,\bar{\mathbf{s}}_0}) < \ell(\theta^*) - \delta) = \mathbb{P}_{\theta^*}(\ell(\theta^*) - \ell(\hat{\theta}_{n,\bar{\mathbf{s}}_0}) > \delta).$$

The proof is completed by Proposition 1 and

$$\begin{aligned}
&\ell(\theta^*) - \ell(\hat{\theta}_{n,\bar{\mathbf{s}}_0}) = n^{-1}\ell_n(\theta^*, \bar{\mathbf{s}}_0) - \ell(\hat{\theta}_{n,\bar{\mathbf{s}}_0}) + \ell(\theta^*) - n^{-1}\ell_n(\theta^*, \bar{\mathbf{s}}_0) \\
\leq & n^{-1}\ell_n(\hat{\theta}_{n,\bar{\mathbf{s}}_0}, \bar{\mathbf{s}}_0) - \ell(\hat{\theta}_{n,\bar{\mathbf{s}}_0}) + \ell(\theta^*) - n^{-1}\ell_n(\theta^*, \bar{\mathbf{s}}_0) \\
\leq & 2\sup_{\theta} |n^{-1}\ell_n(\theta, \bar{\mathbf{s}}_0) - \ell(\theta)|.
\end{aligned}$$

■

5 Asymptotic normality

Let ∇_θ be the gradient and ∇^2_θ the Hessian operator with respect to the parameter θ. For any matrix or vector A, $\|A\| = \sum |A_{ij}|$. Here, we list additional regularity assumptions required for the asymptotic normality proof. Assume a positive real δ exists such that on $G \triangleq \{\theta \in \Theta : |\theta - \theta^*| < \delta\}$, the following conditions hold:

(A7) For all $s' \in \mathbb{S}, \bar{\mathbf{s}} \in \overline{\mathbb{S}}$ and $y' \in \mathbb{Y}, \bar{\mathbf{y}} \in \overline{\mathbb{Y}}$, the functions $\theta \to q_\theta(s'|\bar{\mathbf{s}}, \bar{\mathbf{y}}, x)$ and $\theta \to g_\theta(y'|\bar{\mathbf{y}}, \bar{\mathbf{s}}, x)$ are twice continuously differentiable.

(A8) (a) $\mathbb{E}_{\theta^*}[\sup_{\theta\in G} \max_{s',\bar{\mathbf{s}}} \|\nabla_\theta \log q_\theta(s'|\bar{\mathbf{s}}, \overline{\mathbf{Y}}_0, X_0)\|^4] < \infty$,
$\mathbb{E}_{\theta^*}[\sup_{\theta\in G} \max_{s',\bar{\mathbf{s}}} \|\nabla^2_\theta \log q_\theta(s'|\bar{\mathbf{s}}, \overline{\mathbf{Y}}_0, X_0)\|^2] < \infty$;

(b) $\mathbb{E}_{\theta^*}[\sup_{\theta\in G} \max_{\bar{\mathbf{s}}} \|\nabla_\theta \log g_\theta(Y_0|\overline{\mathbf{Y}}_{-1}, \bar{\mathbf{s}}, X_0)\|^4] < \infty$,
$\mathbb{E}_{\theta^*}[\sup_{\theta\in G} \max_{\bar{\mathbf{s}}} \|\nabla^2_\theta \log g_\theta(Y_0|\overline{\mathbf{Y}}_{-1}, \bar{\mathbf{s}}, X_0)\|^2] < \infty$.

(A9) (a) For almost all $(\bar{\mathbf{y}}, y', x)$, a finite function $f_{\bar{\mathbf{y}},y',x} : \overline{\mathbb{S}} \to \mathbb{R}^+$ exists such that $\sup_{\theta\in G} g_\theta(y'|\bar{\mathbf{y}}, \bar{\mathbf{s}}, x) \leq f_{\bar{\mathbf{y}},y',x}(\bar{\mathbf{s}})$.

(b) For almost all $(\bar{\mathbf{y}}, \bar{\mathbf{s}}, x)$, functions $f^1_{\bar{\mathbf{y}},\bar{\mathbf{s}},x} : \mathbb{Y} \to \mathbb{R}^+$ and $f^2_{\bar{\mathbf{y}},\bar{\mathbf{s}},x} : \mathbb{Y} \to \mathbb{R}^+$ exist in $L^1(\nu)$ such that $\|\nabla_\theta g_\theta(y'|\bar{\mathbf{y}}, \bar{\mathbf{s}}, x)\| \leq f^1_{\bar{\mathbf{y}},\bar{\mathbf{s}},x}(y')$ and $\|\nabla^2_\theta g_\theta(y'|\bar{\mathbf{y}}, \bar{\mathbf{s}}, x)\| \leq f^2_{\bar{\mathbf{y}},\bar{\mathbf{s}},x}(y')$ for all $\theta \in G$.

Assumptions (A7)–(A9) parallel Assumption 7 in Kasahara and Shimotsu (2018). We do not need any high-level assumptions on the finite moments of $\sup_{m\geq 0}\sup_{\theta\in G}\sup_{\bar{\mathbf{s}}}\nabla^i_\theta \log p_\theta(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{X}^1_{-m})$ with $i=1,2$, as Assumption 8 in Kasahara and Shimotsu (2018). Kasahara and Shimotsu (2018) needed the high-level assumptions to apply the dominated convergence theorem to show the the convergence in moments of the gradiant and hessian of the period log-likelihood functions. In this study, we can derive the convergence in moments with (17) and (18) in Lemma 2 and without any further high-level assumptions.

5.1 A central limit theorem for the score function

We define the period score function as

$$\dot{\Delta}_{t,m,\bar{\mathbf{s}}}(\theta) \triangleq \nabla_\theta \Delta_{t,m,\bar{\mathbf{s}}}(\theta) = \nabla_\theta \log p_\theta(Y_t|\overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}, \overline{\mathbf{X}}^t_{-m+1}),$$
$$\dot{\Delta}_{t,m}(\theta) \triangleq \nabla_\theta \Delta_{t,m,\bar{\mathbf{s}}}(\theta) = \nabla_\theta \log p_\theta(Y_t|\overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{X}}^t_{-m+1}).$$

This subsection shows the asymptotic normality of the score function. The proof follows a similar method to that in the proof for consistency. First, we show that the score function can be approximated by a sequence of integrable martingale increments. A useful expression in the approximation is the Fisher identity[7] , which states that the score function in a model with missing data can be obtained by the expectation of the complete score conditional on the observed data.

$$\begin{aligned}\dot{\Delta}_{t,m,\bar{\mathbf{s}}}(\theta) = \mathbb{E}_\theta\left[\sum_{k=-m+1}^{t}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}^t_{-m}, \overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}, \mathbf{X}^t_{-m}\right]\\ - \mathbb{E}_\theta\left[\sum_{k=-m+1}^{t-1}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}, \mathbf{X}^{t-1}_{-m}\right],\end{aligned} \tag{25}$$

$$\begin{aligned}\dot{\Delta}_{t,m}(\theta) = \mathbb{E}_\theta\left[\sum_{k=-m+1}^{t}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}^t_{-m}, \mathbf{X}^t_{-m}\right]\\ - \mathbb{E}_\theta\left[\sum_{k=-m+1}^{t-1}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}^{t-1}_{-m}, \mathbf{X}^{t-1}_{-m}\right]\end{aligned} \tag{26}$$

where $Z_k \triangleq (Y_k, S_k)'$, $\overline{\mathbf{Z}}_k \triangleq (\overline{\mathbf{Y}}_k, \overline{\mathbf{S}}_k)'$, and

$$\phi(\theta, \mathbf{Z}^k_{k-r}, X_k) = \nabla_\theta \log p_\theta(Y_k, S_k|\overline{\mathbf{Y}}_{k-1}, \overline{\mathbf{S}}_{k-1}, X_k)$$

[7] For details, see Cappé, Moulines, and Rydén (2009, p.353).

$$= \nabla_\theta \log \big(g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, \overline{\mathbf{S}}_k, X_k) q_\theta(S_k|\overline{\mathbf{S}}_{k-1}, \overline{\mathbf{Y}}_{k-1}, X_k) \big).$$

As in the previous section, the stationary conditional score is that conditional on the whole history of (Y_t, X_t) starting from past infinity:

$$\begin{aligned}
&\dot{\Delta}_{t,\infty}(\theta^*) \triangleq \lim_{m\to\infty} \dot{\Delta}_{t,m}(\theta^*) \\
=& \lim_{m\to\infty} \Big[\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t] \\
&+ \sum_{k=-m+1}^{t-1} \big(\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1}] \big) \Big] \\
=& \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{-\infty}^t, \mathbf{X}_{-\infty}^t] \\
&+ \sum_{k=-\infty}^{t-1} \big(\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^t, \mathbf{X}_{-\infty}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}] \big).
\end{aligned}$$

(A.23) and (A.24) in the appendix show that $\dot{\Delta}_{k,\infty}(\theta^*)$ is well defined in $L^2(\mathbb{P}_{\theta^*})$. The next lemma shows that $\dot{\Delta}_{t,0,\overline{\mathbf{s}}}(\theta^*)$ converges to $\dot{\Delta}_{t,\infty}(\theta^*)$ in $L^2(\mathbb{P}_{\theta^*})$.

LEMMA 4. *Assume (A2)–(A4) and (A7)-(A8). Then, for all $\overline{\mathbf{s}} \in \overline{\mathbb{S}}$,*

$$\lim_{n\to\infty} \mathbb{E}_{\theta^*} \Big\| \frac{1}{\sqrt{n}} \sum_{t=1}^n \big(\dot{\Delta}_{t,0,\overline{\mathbf{s}}}(\theta^*) - \dot{\Delta}_{t,\infty}(\theta^*) \big) \Big\|^2 = 0.$$

Define the filtration $\mathcal{F}_t = \sigma\big((\overline{\mathbf{Y}}_k, X_{k+1}), -\infty < k \leq t \big)$ for $t \in \mathbb{Z}$. $\mathbb{E}_{\theta^*}[\dot{\Delta}_{t,\infty}|\mathcal{F}_{t-1}] = 0$, since

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\Bigg[\sum_{k=-\infty}^{t-1} \Big(\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^t, \mathbf{X}_{-\infty}^t] \\
&\qquad - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}] \Big) \Big| \overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1} \Bigg] = 0, \\
&\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}] \\
&\qquad = \mathbb{E}_{\theta^*}\Big[\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \overline{\mathbf{S}}_{t-1}, \mathbf{X}_{-\infty}^{t-1}] \Big| \overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1} \Big] = 0.
\end{aligned}$$

$\{\dot{\Delta}_{t,\infty}(\theta^*)\}_{k=-\infty}^{\infty}$ is an $(\mathcal{F}, \mathbb{P}_{\theta^*})$-adapted stationary ergodic and square integrable martingale increment sequence. The central limit theorem for the sums of such a sequence shows that

$$n^{-\frac{1}{2}}\sum_{t=1}^{n}\dot{\Delta}_{t,\infty}(\theta^*) \to \mathbb{N}(0, I(\theta^*)), \qquad \mathbb{P}_{\theta^*} - weakly$$

where $I(\theta^*) \triangleq \mathbb{E}_{\theta^*}[\dot{\Delta}_{0,\infty}(\theta^*)\dot{\Delta}_{0,\infty}(\theta^*)^T]$ is the asymptotic Fisher information matrix. By Lemma 4, $n^{-\frac{1}{2}}\sum_{t=1}^{n}\dot{\Delta}_{t,0,\bar{\mathbf{s}}}(\theta^*)$ has the same limiting distribution.

THEOREM 2. *Assume (A2)–(A4) and (A7)–(A9). Then, for any* $\bar{\mathbf{s}} \in \overline{\mathbb{S}}$,

$$n^{-\frac{1}{2}}\nabla_\theta \ell_n(\theta^*, \bar{\mathbf{s}}) \to \mathbb{N}(0, I(\theta^*)), \qquad \mathbb{P}_{\theta^*} - weakly.$$

5.2 Law of large numbers for the observed Fisher information.

This subsection presents the law of large numbers for the observed Fisher information. The proof is similar to that of the uniform convergence of the log-likelihood function in Section 4 and is based on the approximation of the observed Hessian with a stationary ergodic sequence. A useful expression is the Louis missing information principle (Louis (1982)), which expresses the observed Fisher information in terms of the Hessian of the complete log-likelihood function:

$$\begin{aligned}
&\nabla^2_\theta \log p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^n) \\
=&\mathbb{E}_\theta\Big[\sum_{t=1}^{n}\dot{\phi}_\theta(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)\Big|\overline{\mathbf{Y}}_0^n, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^n\Big] \\
&\qquad\qquad + \mathrm{Var}_\theta\Big[\sum_{t=1}^{n}\phi_\theta(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)\Big|\overline{\mathbf{Y}}_0^n, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^n\Big] \\
=&\sum_{t=1}^{n}\Bigg(\mathbb{E}_\theta\Big[\sum_{k=1}^{t}\dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_0^t, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^t\Big] \\
&\qquad\qquad - \mathbb{E}_\theta\Big[\sum_{k=1}^{t-1}\dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^{t-1}\Big]\Bigg) \\
&+\sum_{t=1}^{n}\Bigg(\mathrm{Var}_\theta\Big[\sum_{k=1}^{t}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_0^t, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^t\Big] \\
&\qquad\qquad - \mathrm{Var}_\theta\Big[\sum_{k=1}^{t-1}\phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_0^{t-1}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}_0, \mathbf{X}_1^{t-1}\Big]\Bigg)
\end{aligned}$$

where

$$\dot{\phi}_\theta(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t) \triangleq \nabla_\theta \phi_\theta(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t) = \nabla^2_\theta \log p_\theta(Y_t, S_t | \overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{S}}_{t-1}, X_t).$$

We define

$$\begin{aligned}
\Gamma_{t,m,\bar{\mathbf{s}}}(\theta) \triangleq \mathbb{E}_\theta\Big[& \sum_{k=-m+1}^{t} \dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^t_{-m}, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}^t_{-m}\Big] \\
& - \mathbb{E}_\theta\Big[\sum_{k=-m+1}^{t-1} \dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}^{t-1}_{-m}\Big], \\
\Phi_{t,m,\bar{\mathbf{s}}}(\theta) \triangleq \mathrm{Var}_\theta\Big[& \sum_{k=-m+1}^{t} \phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^t_{-m}, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}^t_{-m}\Big] \\
& - \mathrm{Var}_\theta\Big[\sum_{k=-m+1}^{t-1} \phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}^{t-1}_{-m}\Big].
\end{aligned}$$

Similarly, we define

$$\begin{aligned}
\Gamma_{t,m}(\theta) \triangleq \mathbb{E}_\theta\Big[& \sum_{k=-m+1}^{t} \dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^t_{-m}, \mathbf{X}^t_{-m}\Big] \\
& - \mathbb{E}_\theta\Big[\sum_{k=-m+1}^{t-1} \dot{\phi}_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^{t-1}_{-m}, \mathbf{X}^{t-1}_{-m}\Big], \\
\Phi_{k,m}(\theta) \triangleq \mathrm{Var}_\theta\Big[& \sum_{k=-m+1}^{t} \phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^t_{-m}, \mathbf{X}^t_{-m}\Big] \\
& - \mathrm{Var}_\theta\Big[\sum_{k=-m+1}^{t-1} \phi_\theta(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}^{t-1}_{-m}, \mathbf{X}^{t-1}_{-m}\Big].
\end{aligned}$$

As in Section 4, we construct the stationary ergodic sequence by conditioning the observed Hessian on the whole history of $(\overline{\mathbf{Y}}_t, X_t)$ from past infinity. Lemmas 11 to 15 in the appendix establish the approximation using the deterministic bounds in Lemma 2. Lemmas 11 and 12 show that $\{\Gamma_{t,m,\bar{\mathbf{s}}}(\theta)\}_{m\geq 0}$ converges uniformly with respect to $\theta \in G$ $\mathbb{P}_{\theta^*}-$a.s. and in $L^1(\mathbb{P}_{\theta^*})$ to a random variable that we denote by $\Gamma_{t,\infty}(\theta)$, and the limit does not depend on $\bar{\mathbf{s}}$. Lemmas 14 and 15 show similar results for $\{\Phi_{t,m,\bar{\mathbf{s}}}(\theta)\}_{m\geq 0}$. Following a similar procedure to Douc et al. (2004), Propositions 3 and 4 show the uniform convergence. We relegate the details of the proof to the appendix.

PROPOSITION 3. *Assume (A2)–(A5) and (A7)–(A8). Then, for all* $\bar{\mathbf{s}} \in \overline{\mathbb{S}}$ *and* $\theta \in G$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\Gamma_{t,0,\bar{\mathbf{s}}}(\theta') - \mathbb{E}_{\theta^*}[\Gamma_{0,\infty}(\theta)]\right| = 0, \qquad \mathbb{P}_{\theta^*} - a.s.$$

PROPOSITION 4. *Assume (A2)–(A5) and (A7)–(A8). Then, for all* $\bar{\mathbf{s}} \in \overline{\mathbb{S}}$ *and* $\theta \in G$,

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\Phi_{t,0,\bar{\mathbf{s}}}(\theta') - \mathbb{E}_{\theta^*}[\Phi_{0,\infty}(\theta)]\right| = 0, \qquad \mathbb{P}_{\theta^*} - a.s.$$

Using the Louis missing information principle and Assumption (A9),

$$\mathbb{E}_{\theta^*}[\Gamma_{0,\infty}(\theta^*) + \Phi_{0,\infty}(\theta^*)] = -\mathbb{E}_{\theta^*}[\dot{\Delta}_{0,\infty}(\theta^*)\dot{\Delta}_{0,\infty}(\theta^*)^T] = -I(\theta^*).$$

Propositions 3 and 4 yield the following theorem.

THEOREM 3. *Assume (A2)–(A5) and (A7)–(A9). Let* $\{\theta_n^*\}$ *be any, possibly stochastic, sequence in* Θ *such that* $\theta_n^* \to \theta^*$ $\mathbb{P}_{\theta^*}-a.s.$ *Then, for all* $\bar{\mathbf{s}}_0 \in \overline{\mathbb{S}}$, $-n^{-1}\nabla_\theta^2 \ell_n(\theta_n^*, \bar{\mathbf{s}}_0) \to I(\theta^*)$, $\mathbb{P}_{\theta^*} - a.s.$

Theorems 2 and 3 together yield the following theorem of asymptotic normality.

THEOREM 4. *Assume (A1)–(A9). Then, for any* $\bar{\mathbf{s}}_0 \in \overline{\mathbb{S}}$,

$$n^{1/2}(\hat{\theta}_{n,\bar{\mathbf{s}}_0} - \theta^*) \to \mathbb{N}(0, I(\theta^*)^{-1}), \qquad \mathbb{P}_{\theta^*} - weakly.$$

From Theorems 1, 3, and 4, the negative of the inverse of the Hessian matrix at the MLE value is a consistent estimate of the asymptotic variance.

6 Discussion of assumptions

This section shows that the Assumptions (A4) and (A6) hold in models with a transition equation of the observed process (3) and a transition probability of logistic form (6) or of the Chang et al.

(2017) type (7). In addition, we derive explicit conditions for Assumption (A2) to hold in models with a transition equation of the observed process (5) and transition probability (7).[8] The models here can cover a wide range of empirical research. The other assumptions are either standard in the literature or easy to verify, and are not discussed here.

PROPOSITION 5. *Assumptions (A4) and (A6) hold in a two-regime endogenous regime-switching model with a transition probability (6) of logistic form:* $q_\theta(s_t = s | s_{t-1} = s, s_{t-2}, \ldots, s_{t-r}, X_t) = \frac{\exp(\beta_s' X_t)}{1+\exp(\beta_s' X_t)}$ *and a transition equation of the observed process (3) satisfying*

1. $\sigma(0) \neq \sigma(1)$ *and* $0 < \sigma(0), \sigma(1) < \infty$*;*

2. $U_t \sim \mathbb{N}(0, 1)$*;*

3. $m_\theta(\overline{\mathbf{s}}_i) \neq m_\theta(\overline{\mathbf{s}}_j)$*, for* $\overline{\mathbf{s}}_i \neq \overline{\mathbf{s}}_j$*; and*

4. *for all* $s \in \mathbb{S}$*,* $\exp(\beta_s' X_t)$ *and* $\exp(\beta_s' X_t)^{-1}$ *have finite second moments.*

PROPOSITION 6. *Assumptions (A4) and (A6) hold in a two-regime endogenous regime-switching model with a transition equation of the observed process (3) satisfying conditions 1–3 in Proposition 5 and a Chang et al. (2017)-type transition probability (7).*

The following lemmas are helpful for showing the propositions.

LEMMA 5. *A sufficient condition for Assumption (A4b) is that for some* $\delta > 0$*,* $\mathbb{E}_{\theta^*}[|\log b_-(\mathbf{Y}_{-r+1}^1, X_1)|^{1+\delta}] < \infty$*.*

The proof follows from Kasahara and Shimotsu (2018, p. 9).

LEMMA 6. *A sufficient condition for Assumption (A4a) is that for some* $\delta > 0$*,* $\mathbb{E}_{\theta^*}[|\log \sigma_-(\overline{\mathbf{y}}_0, x_1)|^{1+\delta}] < \infty$*.*

Proof. Set $C_1 = 1$.

$$\mathbb{P}_{\theta^*}(\sigma_-(\overline{\mathbf{Y}}_0, X_1) \leq e^{-\alpha_1 \xi}) = \mathbb{P}_{\theta^*}(|\log \sigma_-(\overline{\mathbf{Y}}_0, X_1)|^{1+\delta} \geq (\alpha_1 \xi)^{1+\delta})$$

[8] Pouzo et al. (2018) discussed the stationary ergodicity assumption in models with a transition probability (6), and thus, we do not repeat it here.

$$\leq \mathbb{E}_{\theta^*}\left[|\log \sigma_-(\overline{\mathbf{Y}}_0, X_1)|^{1+\delta}\right]/(\alpha_1\xi)^{1+\delta} = C_2\xi^{-(1+\delta)}$$

where $C_2 = \mathbb{E}_{\theta^*}[|\log \sigma_-(\overline{\mathbf{Y}}_0, X_1))|^{1+\delta}]/\alpha_1^{1+\delta}$, and the second inequality follows from Markov's inequality. ■

LEMMA 7 (Teicher, 1967)**.** *Assume that the class of finite mixtures of the family* (f_ϕ) *of densities on* $y \in \mathbb{Y}$ *with parameter* $\phi \in \Phi$ *is identifiable. Then, the class of finite mixtures of the* $n-$*fold product densities* $f_\phi^{(n)}(y) = f_{\phi_1}(y_1)\cdots f_{\phi_n}(y_n)$ *on* $y \in \mathbb{Y}^n$ *with parameter* $\phi \in \Phi^n$ *is identifiable.*

This lemma was proved by Teicher (1967). Write the conditional likelihood function as

$$\begin{aligned} &p_\theta(Y_1, \ldots, Y_n|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n) \\ =&\sum_{s_{-r+2}^n}\left(\prod_{t=1}^n g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t)\prod_{t=2}^n q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)\mathbb{P}_\theta(\overline{\mathbf{S}}_1 = \bar{\mathbf{s}}_1|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n)\right). \end{aligned} \tag{27}$$

The conditional likelihood function is a finite mixture of the $n-$fold product densities $\prod_{t=1}^n g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t)$ with mixing distributions $\prod_{t=2}^n q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)\mathbb{P}_\theta(\overline{\mathbf{S}}_1 = \bar{\mathbf{s}}_1|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n)$. Under condition 2 of Propositions 5, $g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, \bar{\mathbf{s}}_t, X_t) =_d \mathbb{N}(m_\theta(\bar{\mathbf{s}}_t), \sigma(s_t)^2)$. Teicher (1960) stated that finite joint mixtures of normal distributions over both mean and variance are identifiable. It follows that the conditional likelihood function is identifiable.

Proof of Proposition 5. The transition equation of the observed process (3) satisfies the condition in Lemma 5 with $\delta = 1$, since

$$\mathbb{E}_{\theta^*}[|\log g_\theta(Y_1|\overline{\mathbf{Y}}_0, \bar{\mathbf{s}}_1, X_1)|^2] = \mathbb{E}_{\theta^*}\left[\left|-\frac{1}{2}\log(2\pi) - \log\sigma(s_t) - \frac{1}{2}U_t\right|^2\right] < \infty \tag{28}$$

Next, we show that the logistic-type Diebold et al. (1994) transition probability satisfies the condition in Lemma 6 with $\delta = 1$. For $s \in \mathbb{S}$, we use $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$ to have

$$\begin{aligned} &\mathbb{E}_{\theta^*}[|\log q_\theta(s_1 = s|s_0 = s, s_{-1}, \ldots, s_{-r+1}, X_1)|^2] \\ =&\mathbb{E}_{\theta^*}\left[\left|\log\frac{\exp(\beta_s' X_1)}{1+\exp(\beta_s' X_1)}\right|^2\right] \leq \mathbb{E}_{\theta^*}[\exp(\beta_s' X_1)^{-2}] < \infty. \end{aligned}$$

For $s, s' \in \mathbb{S}$ and $s' \neq s$,

$$\begin{aligned}
&\mathbb{E}_{\theta^*}[|\log q_\theta(s_1 = s'|s_0 = s, s_{-1}, \ldots, s_{-r+1}, X_1)|^2] \\
=&\mathbb{E}_{\theta^*}\left[\left|\log \frac{1}{1+\exp(\beta_s' X_1)}\right|^2\right] \leq \mathbb{E}_{\theta^*}[\exp(\beta_s' X_1)^2] < \infty.
\end{aligned}$$

Next, we show (A6). The "only if" part is obvious. Here, we show only the "if" part. $\theta = (\theta_1', \theta_2')'$, where θ_1 is the parameter vector in (3), and $\theta_2 = (\beta(s_1), \ldots, \beta(s_J))'$. If $\mathbb{P}_\theta(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n) = \mathbb{P}_{\theta^*}(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n)$, by Lemma 7, the law of the process $\{m_\theta(\bar{\mathbf{s}}_t), \sigma(s_t)^2\}$ is in accordance with $\{m_{\theta^*}(\bar{\mathbf{s}}_t), \sigma^*(s_t)^2\}$. Since $m_{\theta^*}(\bar{\mathbf{s}}_i) \neq m_{\theta^*}(\bar{\mathbf{s}}_j)$ for $\bar{\mathbf{s}}_i \neq \bar{\mathbf{s}}_j$, and $\sigma(s_i) \neq \sigma(s_j)$ for $s_i \neq s_j$, the sets $\{m_{\theta^*}(\bar{\mathbf{s}}_i)\}$ and $\{m_\theta(\bar{\mathbf{s}}_i)\}$, $\{\sigma^*(s_i)\}$ and $\{\sigma(s_i)\}$ are identical. Under appropriate permutation, $m_{\theta^*}(\bar{\mathbf{s}}_i) = m_\theta(\bar{\mathbf{s}}_i)$ and $\sigma^*(s_i) = \sigma(s_i)$. Thus, $\theta_1^* = \theta_1$. Again, since the law of $\{m_\theta(\bar{\mathbf{s}}_t), \sigma(s_t)^2\}$ is in accordance with $\{m_{\theta^*}(\bar{\mathbf{s}}_t), \sigma^*(s_t)^2\}$, $q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, X_t) = q_{\theta^*}(s_t|\bar{\mathbf{s}}_{t-1}, X_t)$. By the monotonicity of q_θ in $\beta(s)$, $\theta_2^* = \theta_2$. ■

Proof of Proposition 6. (A4a) can be shown as (28) and is omitted here for brevity. We can prove (A4b) by showing that the condition in Lemma 6 holds with $\delta = 1$. The proof is in the appendix. Then, we show the "if" part in (A6). When the number of lags is k in (4) and $\theta = (\theta_1', \theta_2')'$, where θ_1 is the parameter vector in (3) and $\theta_2 = (\alpha, \tau, \rho)'$. When $\mathbb{P}_\theta(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n) = \mathbb{P}_{\theta^*}(\mathbf{Y}_1^n \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^n)$, similarly to the previous analysis, we can show $\theta_1^*=\theta_1$ (with permutation of states if necessary). Since the laws of the processes $\{m_\theta(\bar{\mathbf{s}}_t), \sigma(s_t)^2\}$ and $\{m_{\theta^*}(\bar{\mathbf{s}}_t), \sigma^*(s_t)^2\}$ are in accordance, $\mathbb{P}_\theta(S_t = 0) = \Phi(\tau\sqrt{1-\alpha^2})$ and $\mathbb{P}_\theta(S_t = 0) = \mathbb{P}_{\theta^*}(S_t = 0)$ imply $\tau\sqrt{1-\alpha^2} = \tau^*\sqrt{1-\alpha^{*2}}$. Likewise, $q_\theta(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) = q_{\theta^*}(s_t|\bar{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t)$, (7), and (8) imply $\frac{\tau-\rho u_{t-1}-\alpha x/\sqrt{1-\alpha^2}}{\sqrt{1-\rho^2}} = \frac{\tau^*-\rho^* u_{t-1}-\alpha^* x/\sqrt{1-\alpha^{*2}}}{\sqrt{1-\rho^{*2}}}$. It follows that $\theta_2^* = \theta_2$. ■

In order to investigate Assumption (A2), we use the following vectorial form of the transition equation of the observed process (5):

$$\mathbf{Y}_{t-k+1}^t = A(S_t)\mathbf{Y}_{t-k}^{t-1} + B(S_t, X_t) \tag{29}$$

where $B(S_t, X_t) = (\mu(S_t) + \gamma_X(S_t)'X_t + \sigma(S_t)U_t, 0, \ldots, 0)'$ and

$$A(S_t) = \begin{pmatrix} \gamma_1(S_t) & \gamma_2(S_t) & \ldots & \gamma_k(S_t) \\ 1 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

In the following proof, we simply write $B(S_t)$ instead of $B(S_t, X_t)$. Results on ergodicity for autoregressive processes with basic Markov switching have been shown in Yao (2001) and Francq and Zakoıan (2001). We extend their proofs to Chang et al. (2017)-type transition probability, and summarize our findings in Proposition 7. Note that the Chang et al. (2017)-type transition probability (7) is essentially a function of (S_t, S_{t-1}, U_{t-1}), which allows us to use another notation $q_\theta(S_t|S_{t-1}, U_{t-1})$. Since W_{t-1} is independent of U_{t-1}, and so is S_{t-1}, integration with respect to U_{t-1} yields the unconditional transition probabilities

$$\begin{aligned} p_{ij} &\triangleq \mathbb{E}[q_\theta(S_t = j|S_{t-1} = i, U_{t-1})|S_{t-1} = i] \\ &= (1-j)\omega + j(1-\omega) \end{aligned}$$

where

$$\omega = \frac{[(1-i)\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} + i\int_{\tau\sqrt{1-\alpha^2}}^{\infty}]\Phi\left(\tau - \frac{\alpha x}{\sqrt{1-\alpha^2}}\right)\varphi(x)dx}{(1-i)\Phi(\tau\sqrt{1-\alpha^2}) + i[1-\Phi(\tau\sqrt{1-\alpha^2})]}.$$

We denote by $\otimes$ the Kronecker tensor product. Let the $2k^2 \times 2k^2$ matrix

$$M = \begin{pmatrix} (A(0)\otimes A(0))p_{00} & (A(0)\otimes A(0))p_{10} \\ (A(1)\otimes A(1))p_{01} & (A(1)\otimes A(1))p_{11} \end{pmatrix}. \tag{30}$$

We denote the spectral radius of a real matrix $A = (a_{ij})$ by $\rho(A)$. Then, we obtain the following result.

PROPOSITION 7. *Assumption (A2) holds in a two-regime endogenous regime-switching model with a transition equation of the observed process (5) satisfying that X_t is strictly stationary and ergodic and has finite second moment, and Chang et al. (2017)-type transition probability (7) satisfying $|\alpha| < 1$, $W_0 \sim N(0, \frac{1}{1-\alpha^2})$ and $\rho(M) < 1$.*

Proof. In the proof, we omit subscript θ for simplicity. Note that the stationarity and ergodicity of $\{S_t\}$ are implied by those of $\{W_t\}$. Under the assumptions that $|\alpha| < 1$ and $W_0 \sim N(0, \frac{1}{1-\alpha^2})$, $\{W_t\}$ becomes a strictly stationary process. In addition, the dynamics of $\{W_t\}$ can be equivalently defined as

$$\begin{aligned} W_t &= \alpha W_{t-1} + \rho U_{t-1} + \sqrt{1-\rho^2} V_t \\ &= \sum_{l=0}^{\infty} \rho \alpha^l U_{t-1-l} + \sum_{l=0}^{\infty} \sqrt{1-\rho^2} \alpha^l V_{t-l}. \end{aligned}$$

According to Theorem 7.1.3 in Durrett (2013), the ergodicity of $\{W_t\}$ follows from the ergodicity of $\{U_t\}$ and $\{V_t\}$. Consequently, $\{A(S_t), B(S_t)\}$ is a strictly stationary and ergodic sequence.

We show that

$$H_{t,p} \triangleq A(S_t)A(S_{t-1}) \dots A(S_{t-p+1})B(S_{t-p})$$

converges to zero in L^2 at an exponential rate, as p goes to infinity. We obtain

$$\begin{aligned} &\mathbb{E}[vec(H_{t,p}H'_{t,p})] \\ =&\mathbb{E}[(A(S_t) \otimes A(S_t)) \dots (A(S_{t-p+1}) \otimes A(S_{t-p+1}))(B(S_{t-p}) \otimes B(S_{t-p}))] \end{aligned} \tag{31}$$

$$\begin{aligned} =&\int \cdots \int \sum_{s_t,\dots,s_{t-p}} (A(s_t) \otimes A(s_t)) \dots (A(s_{t-p+1}) \otimes A(s_{t-p+1}))(B(s_{t-p}) \otimes B(s_{t-p})) \\ &\times p(s_t, s_{t-1}, u_{t-1}, s_{t-2}, u_{t-2}, \dots, s_{t-p}, u_{t-p}, x_{t-p}) du_{t-1} \dots du_{t-p} dx_{t-p} \end{aligned} \tag{32}$$

with

$$\begin{aligned} &p(s_t, s_{t-1}, u_{t-1}, s_{t-2}, u_{t-2}, \dots, s_{t-p}, u_{t-p}, x_{t-p}) \\ =&p(s_t|s_{t-1}, u_{t-1})p(u_{t-1})p(s_{t-1}|s_{t-2}, u_{t-2}) \dots p(s_{t-p+1}|s_{t-p}, u_{t-p})p(u_{t-p})p(s_{t-p}, x_{t-p}) \\ =&p(s_t, u_{t-1}|s_{t-1})p(s_{t-1}, u_{t-2}|s_{t-2}) \dots p(s_{t-p+1}, u_{t-p}|s_{t-p})p(s_{t-p}, x_{t-p}) \end{aligned} \tag{33}$$

where the first equality follows because U_t is independent of W_t, W_{t-1}, and U_{t-1}.

Plugging (33) back into (32) yields

$$\mathbb{E}[vec(H_{t,p}H'_{t,p})] = \mathbb{E}[(A(S_t) \otimes A(S_t)) \dots (A(S_{t-p+1}) \otimes A(S_{t-p+1}))(B(S_{t-p}) \otimes B(S_{t-p}))]$$

$$= \int \cdots \int \sum_{s_t,\dots,s_{t-p}} (A(s_t) \otimes A(s_t)) \dots (A(s_{t-p+1}) \otimes A(s_{t-p+1}))(B(s_{t-p}) \otimes B(s_{t-p}))$$

$$\times p(s_t, u_{t-1}|s_{t-1})p(s_{t-1}, u_{t-2}|s_{t-2}) \dots p(s_{t-p+1}, u_{t-p}|s_{t-p})p(s_{t-p}, x_{t-p})du_{t-1} \dots du_{t-p}dx_{t-p}$$

$$= \int \int \sum_{s_t,\dots,s_{t-p}} (A(s_t) \otimes A(s_t)) \dots (A(s_{t-p+1}) \otimes A(s_{t-p+1}))(B(s_{t-p}) \otimes B(s_{t-p}))$$

$$\times p(s_t|s_{t-1})p(s_{t-1}|s_{t-2}) \dots p(s_{t-p+2}|s_{t-p+1})p(s_{t-p+1}, u_{t-p}|s_{t-p})p(s_{t-p}, x_{t-p})du_{t-p}dx_{t-p}.$$

We define the $2k^2 \times 1$ matrix as

$$N_{t-p} =$$

$$\begin{pmatrix} \int \int \sum_{s_{t-p}} \big(A(1) \otimes A(1)\big)\big(B(s_{t-p}) \otimes B(s_{t-p})\big)p(s_{t-p+1} = 1, s_{t-p}, u_{t-p}, x_{t-p})du_{t-p}dx_{t-p} \\ \int \int \sum_{s_{t-p}} \big(A(2) \otimes A(2)\big)\big(B(s_{t-p}) \otimes B(s_{t-p})\big)p(s_{t-p+1} = 2, s_{t-p}, u_{t-p}, x_{t-p})du_{t-p}dx_{t-p} \end{pmatrix}.$$

Then, we obtain $\mathbb{E}[vec(H_{t,p}H'_{t,p})] = \mathbb{I}'M^{p-1}N_{t-p}$ where $\mathbb{I} = (I_{k^2}, I_{k^2})'$ and I_{k^2} are a $2k^2 \times k^2$ matrix and a $k^2 \times k^2$ identity matrix, respectively. It follows that

$$\|H_{t,p}\|_{L^2} \le \|\mathbb{I}'M^{p-1}N_{t-p}\|^{1/2} \le \|\mathbb{I}'\|^{1/2}\|M^{p-1}\|^{1/2}\|N_{t-p}\|^{1/2}.$$

Then, the process

$$\begin{aligned} \mathbf{Y}_{t-k+1}^t &= B(S_t) + A(S_t)B(S_{t-1}) + A(S_t)A(S_{t-1})B(S_{t-2}) + \dots \\ &= H_{t,0} + H_{t,1} + H_{t,2} + \dots \end{aligned}$$

is well defined in L^2 if $\rho(M) < 1$, since $\|\mathbf{Y}_{t-k+1}^t\|_{L^2} \le \sum_{p=0}^{\infty} \|H_{t,p}\|_{L^2}$ and $\|H_{t,p}\|_{L^2}$ converges to zero at an exponential rate as p goes to infinity. Since $\{A(S_t), B(S_t)\}$ is strictly stationary ergodic, again by Theorem 7.1.3 in Durrett (2013), $\mathbf{Y}_{t-k+1}^t$ is strictly stationary ergodic. ■

7 Simulation

This section presents the results of Monte Carlo simulation of the sample properties of the MLE in Diebold et al. (1994)-type and Chang et al. (2017)-type models. This section lists the frequency with which the true parameter falls in the 95% confidence interval centered at the MLE. If the distribution of the sample estimator can be well approximated by the asymptotic normal distribution, then the frequency should be close to 95%. Moreover, as the sample size increases, the coverage error should decrease.

The first experiment is based on a two-regime volatility-switching model with a Chang et al. (2017)-type transition probability:

$$Y_t = \sigma_{S_t} U_t$$
$$S_t = \mathbb{1}\{W_t \geq \tau\}.$$

The parameter $\theta = (\sigma_0, \sigma_1, \alpha, \tau, \rho)'$. We use the parameter in the simulation model of Chang et al. (2017), $\theta^* = (0.04, 0.12, 0.4, 0.4, 0.5)$. The result is listed in Table 1.

Table 1: Coverage frequency of asymptotic 95% confidence interval of volatility model with Chang, Choi, and Park (2017) transition probability (1000 replications)

Observations	σ_0	σ_1	ρ	α	τ
$n = 200$	0.908	0.916	0.875	0.887	0.934
$n = 400$	0.940	0.944	0.910	0.908	0.945
$n = 600$	0.948	0.937	0.919	0.901	0.945
$n = 800$	0.940	0.955	0.938	0.897	0.963
$n = 1000$	0.957	0.941	0.926	0.901	0.948

The second experiment is based on a two-regime mean-switching model with a Diebold et al. (1994)-type transition probability:

$$Y_t = \mu_{S_t} + U_t, \qquad U_t \sim i.i.d.\ \mathbb{N}(0,1)$$

with transition matrix

$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} \frac{\exp(\beta_{q0}+\beta_{q1}X_t)}{1+\exp(\beta_{q0}+\beta_{q1}X_t)} & 1-\frac{\exp(\beta_{q0}+\beta_{q1}X_t)}{1+\exp(\beta_{q0}+\beta_{q0}X_t)} \\ 1-\frac{\exp(\beta_{p0}+\beta_{p1}X_t)}{1+\exp(\beta_{p0}+\beta_{p0}X_t)} & \frac{\exp(\beta_{p0}+\beta_{p0}X_t)}{1+\exp(\beta_{p0}+\beta_{p0}X_t)} \end{pmatrix} \end{array}$$

The parameter $\theta = (\mu_0, \mu_1, \beta_{q0}, \beta_{q1}, \beta_{p0}, \beta_{p1})'$. We choose parameters from the simulation example in Diebold et al. (1994). $\theta^* = (-0.8, 0.5, 0.79, -2, 1, 2)'$. The process of the predetermined

Table 2: Coverage frequency of asymptotic 95% confidence interval of mean model with Diebold, Lee, and Weinbach (1994) transition probability (1000 replications)

Observations	μ_0	μ_1	β_{q0}	β_{q1}	β_{p0}	β_{p1}
$n = 200$	0.889	0.907	0.970	0.938	0.956	0.927
$n = 400$	0.925	0.949	0.979	0.934	0.976	0.956
$n = 600$	0.945	0.935	0.958	0.954	0.972	0.952
$n = 800$	0.941	0.939	0.970	0.963	0.968	0.945
$n = 1000$	0.940	0.935	0.953	0.956	0.955	0.958

variable is $X_{t+1} = \alpha_X X_t + \tilde{U}_t$ with $\alpha_X = 0.4$ and $\tilde{U}_t \sim i.i.d.\mathbb{N}(0, 1)$. The result is listed in Table 2.

8 Conclusion

This study proves consistency and asymptotic normality of the MLE in endogenous regime-switching models represented by Chang et al. (2017). The model we consider is general enough to cover time-varying transition probability regime-switching models represented by Diebold et al. (1994).

Our proof follows the method of Douc et al. (2004) and Kasahara and Shimotsu (2018). The key feature is to approximate the sequence of non-ergodic period predictive densities with a stationary ergodic process using the mixing rate of the unobservable state process conditional on the observations. The time-varying transition probabilities, however, heighten the difficulty of proving the approximation, because the mixing rate can approach one as the transition probabilities takes extremely small values. In this study, we provide almost deterministic geometric decaying bounds for the time-varying mixing rate. The assumptions in the proof are low-level ones and have been shown to hold in general. Thus, this study provides theoretical foundations for a wide range of endogenous regime-switching models in the literature. Some interesting and important statistical tests in empirical work can now be conducted, such as whether the observables affect transition probability and whether the effect is positive or negative.

Throughout the paper, we assume that the number of regimes is precisely known. One fails to identify all of the parameters if the number of regimes is set greater than the true value. There are few studies about criteria for selecting the proper number of regimes in endogenous regime-switching models. We leave this topic for future research.

References

Ailliot, P., & Pene, F. (2015). Consistency of the maximum likelihood estimate for non-homogeneous markov–switching models. *ESAIM: Probability and Statistics*, *19*, 268–292.

Ang, A., Bekaert, G., & Wei, M. (2008). The term structure of real rates and expected inflation. *The Journal of Finance*, *63*(2), 797–849.

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Bekaert, G., & Harvey, C. R. (1995). Time-varying world market integration. *The Journal of Finance*, *50*(2), 403–444.

Bickel, P. J., & Ritov, Y. (1996). Inference in hidden markov models i: Local asymptotic normality in the stationary case. *Bernoulli*, *2*(3), 199–228.

Bickel, P. J., Ritov, Y., & Ryden, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. *Annals of Statistics*, 1614–1635.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, *12*(3), 309–316.

Cappé, O., Moulines, E., & Rydén, T. (2009). Inference in hidden markov models. In *Proceedings of eusflat conference* (pp. 14–16).

Chang, Y., Choi, Y., & Park, J. Y. (2017). A new approach to model regime switching. *Journal of Econometrics*, *196*(1), 127–143.

Chib, S., & Dueker, M. (2004). Non-markovian regime switching with endogenous states and time-varying state strengths.

Diebold, F. X., Lee, J.-H., & Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. *Business Cycles: Durations, Dynamics, and Forecasting*, *1*, 144–165.

Douc, R., & Matias, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden markov models. *Bernoulli*, *7*(3), 381–420.

Douc, R., Moulines, E., & Ryden, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime. *The Annals of statistics*, *32*(5), 2254–2304.

Durrett, R. (2013). *Probability: Theory and examples.*

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, *12*(3), 299–308.

Filardo, A. J., & Gordon, S. F. (1998). Business cycle durations. *Journal of econometrics*, *85*(1), 99–123.

Francq, C., & Zakoıan, J.-M. (2001). Stationarity of multivariate markov–switching arma models. *Journal of Econometrics*, *102*(2), 339–364.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1), 307–333.

Jensen, J. L., & Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Annals of Statistics*, 514–535.

Kasahara, H., & Shimotsu, K. (2018). Asymptotic properties of the maximum likelihood estimator in regime switching econometric models. *Journal of Econometrics*.

Kim, C.-J., Piger, J., & Startz, R. (2008). Estimation of markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, *143*(2), 263–273.

Klaassen, F. (2002). Improving garch volatility forecasts with regime-switching garch. *Empirical Economics*, *27*(2), 363–394.

Krolzig, H.-M. (1997). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*. Springer-Verlag Berlin Heidelberg.

Le Gland, F., & Mevel, L. (2000). Exponential forgetting and geometric ergodicity in hidden markov models. *Mathematics of Control, Signals and Systems*, *13*(1), 63–93.

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden markov models. *Stochastic processes and their applications*, *40*(1), 127–143.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 226–233.

Pouzo, D., Psaradakis, Z., & Sola, M. (2018). Maximum likelihood estimation in possibly misspecified dynamic models with time inhomogeneous markov regimes.

Shiryaev, A. N. (1996). *Probability* (2nd). Springer, New York.

Teicher, H. (1960). On the mixture of distributions. *The Annals of Mathematical Statistics*, 55–73.

Teicher, H. (1967). Identifiability of mixtures of product measures. *The Annals of Mathematical Statistics*, *38*(4), 1300–1302.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, *20*(4), 595–601.

Watson, M. W. (1992). *Business cycle durations and postwar stabilization of the us economy*. National Bureau of economic research.

Yao, J. (2001). On square-integrability of an AR process with Markov switching. *Statistics & Probability Letters*, *52*(3), 265–270.

A Appendix

A.1 Detailed proof of lemmas and corollaries in Section 3

Proof of Lemma 1. For $-m \leq k \leq n$, $\{S_k\}$ is a Markov chain conditionally, since

$$p_\theta(\overline{\mathbf{S}}_k|\overline{\mathbf{S}}_{-m}^{k-1}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n) = p_\theta(\overline{\mathbf{S}}_k|\overline{\mathbf{S}}_{k-1}, \overline{\mathbf{Y}}_{k-1}^n, \mathbf{X}_k^n).$$

The proof follows from Lemma 1 in Kasahara and Shimotsu (2018).

To see the minorization condition, observe that

$$\begin{aligned}
&\mathbb{P}_\theta(\overline{\mathbf{S}}_k \in A|\overline{\mathbf{S}}_{k-r}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n) = \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in A|\overline{\mathbf{S}}_{k-r}, \overline{\mathbf{Y}}_{k-r}^n, \mathbf{X}_{k-r+1}^n) \\
=&\frac{\sum_{\overline{\mathbf{s}}_k \in A} p_\theta(\mathbf{Y}_k^n|\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k, \overline{\mathbf{Y}}_{k-1}, \mathbf{X}_k^n)\mathbb{P}_\theta(\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k|\overline{\mathbf{S}}_{k-r}, \overline{\mathbf{Y}}_{k-r}^{k-1}, \mathbf{X}_{k-r+1}^k)}{\sum_{\overline{\mathbf{s}}_k \in \overline{\mathbb{S}}} p_\theta(\mathbf{Y}_k^n|\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k, \overline{\mathbf{Y}}_{k-1}, \mathbf{X}_k^n)\mathbb{P}_\theta(\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k|\overline{\mathbf{S}}_{k-r}, \overline{\mathbf{Y}}_{k-r}^{k-1}, , \mathbf{X}_{k-r+1}^k)}.
\end{aligned}$$

Since $\mathbb{P}_\theta(\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k|\overline{\mathbf{s}}_{k-r}, \overline{\mathbf{Y}}_{k-r}^{k-1}, X_{k-r+1}^k) \leq 1$ and

$$\begin{aligned}
&\mathbb{P}_\theta(\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k|\overline{\mathbf{s}}_{k-r}, \overline{\mathbf{Y}}_{k-r}^{k-1}, \mathbf{X}_{k-r+1}^k) \\
=&\frac{\prod_{\ell=k-r+1}^{k-1} g_\theta(Y_\ell|\overline{\mathbf{Y}}_{\ell-1}, \overline{\mathbf{s}}_\ell, X_\ell) \prod_{\ell=k-r+1}^{k} q_\theta(s_\ell|\overline{\mathbf{s}}_{\ell-1}, \overline{\mathbf{Y}}_{\ell-1}, X_\ell)}{\sum_{\overline{\mathbf{s}}_k \in \overline{\mathbb{S}}} \prod_{\ell=k-r+1}^{k-1} g_\theta(Y_\ell|\overline{\mathbf{Y}}_{\ell-1}, \overline{\mathbf{s}}_\ell, X_\ell) \prod_{\ell=k-r+1}^{k} q_\theta(s_\ell|\overline{\mathbf{s}}_{\ell-1}, \overline{\mathbf{Y}}_{\ell-1}, X_\ell)} \\
\geq&\frac{\prod_{\ell=k-r+1}^{k-1} b_-(\mathbf{Y}_{\ell-r}^\ell, X_\ell) \prod_{\ell=k-r+1}^{k} \sigma_-(\overline{\mathbf{Y}}_{\ell-1}, X_\ell)}{b_+^{r-1}} = \omega(\mathbf{Y}_{k-2r+1}^{k-1}, \mathbf{X}_{k-r+1}^k),
\end{aligned}$$

it readily follows that

$$\mathbb{P}_\theta(\overline{\mathbf{S}}_k \in A|\overline{\mathbf{S}}_{k-r}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n) \geq \omega(\mathbf{Y}_{k-2r+1}^{k-1}, \mathbf{X}_{k-r+1}^k)\mu_k(\mathbf{Y}_{k-r}^n, \mathbf{X}_k^n, A)$$

with

$$\mu_k(\mathbf{Y}_{k-r}^n, \mathbf{X}_k^n, A) \triangleq \frac{\sum_{\overline{\mathbf{s}}_k \in A} p_\theta(\mathbf{Y}_k^n|\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k, \overline{\mathbf{Y}}_{k-1}, \mathbf{X}_k^n)}{\sum_{\overline{\mathbf{s}}_k \in \overline{\mathbb{S}}} p_\theta(\mathbf{Y}_k^n|\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k, \overline{\mathbf{Y}}_{k-1}, \mathbf{X}_k^n)}. \tag{A.1}$$

In order for $\mu_k(\mathbf{Y}_{k-r}^n, \mathbf{X}_k^n, A)$ to be a well-defined probability measure, we need show only that the denominator is strictly positive. The summand term in the denominator of (A.1) is

$$\begin{aligned} p_\theta(\mathbf{Y}_k^n|\overline{\mathbf{S}}_k = \overline{\mathbf{s}}_k, \overline{\mathbf{Y}}_{k-1}, \mathbf{X}_k^n) &= \sum_{s_{k+1}^n} \prod_{\ell=k}^n g_\theta(Y_\ell|\overline{\mathbf{Y}}_{\ell-1}, \overline{\mathbf{s}}_\ell, X_\ell) \prod_{\ell=k+1}^n q_\theta(s_\ell|\overline{\mathbf{s}}_{\ell-1}, \overline{\mathbf{Y}}_{\ell-1}, X_\ell) \\ &\geq \prod_{\ell=k}^n b_-(\mathbf{Y}_{\ell-r}^\ell, X_\ell) \prod_{\ell=k+1}^n \sigma_-(\overline{\mathbf{Y}}_{\ell-1}, X_\ell) > 0. \end{aligned}$$

■

Proof of Lemma 2. First, we show (17). By Assumption (A4), $C_6 < 1$ exists such that for any $\xi > 0$,

$$\mathbb{P}_{\theta^*}\left(1 - \omega(\mathbf{V}_t) \geq 1 - C_6 e^{-[\alpha_1 r + \alpha_2(r-1)]\xi}\right) \leq rC_2\xi^{-\beta_1} + (r-1)C_4\xi^{-\beta_2}. \tag{A.2}$$

We choose ξ_0 such that it satisfies

$$\frac{1}{16r}(1 - C_6 e^{-[\alpha_1 r + \alpha_2(r-1)]\xi_0})^{mn} = rC_2\xi_0^{-\beta_1} + (r-1)C_4\xi_0^{-\beta_2}. \tag{A.3}$$

The existence of such ξ_0 is guaranteed by monotone increasing of $\frac{1}{16r}(1 - C_6 e^{-[\alpha_1 r + \alpha_2(r-1)]\xi})^{mn}$ in ξ from $\frac{1}{16r}(1 - C_6)^{mn}$ to $\frac{1}{16r}$ and monotone decreasing of $rC_2\xi^{-\beta_1} + (r-1)C_4\xi^{-\beta_2}$ in ξ from $+\infty$ to 0. We define

$$\rho = 1 - C_6 e^{-[\alpha_1 r + \alpha_2(r-1)]\xi_0} \tag{A.4}$$

and

$$I_{t_k} \triangleq \mathbb{1}\{1-\omega(\mathbf{V}_{t_k}) \geq \rho\} \tag{A.5}$$

Notice that $\rho \in (0,1)$. Using $(1-\omega(\mathbf{V}_{t_k}))^m \leq \rho^{(1-I_{t_k})m}$ and $mn \geq n$,

$$\mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}(1-\omega(\mathbf{V}_{t_k}))^m\right] \leq \mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}\rho^{(1-I_{t_k})m}\right] = \rho^{mn}\mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}\rho^{-mI_{t_k}}\right] \leq \rho^{n}\mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}\rho^{-mI_{t_k}}\right].$$

Next, we find an upper bound for $\mathbb{E}_{\theta^*}[\prod_{k=1}^{n}\rho^{-mI_{t_k}}]$. Using the generalized Hölder's inequality,

$$\mathbb{E}_{\theta^*}\left[\prod_{k=1}^{n}\rho^{-mI_{t_k}}\right] \leq \prod_{k=1}^{n}\left(\mathbb{E}_{\theta^*}\left[\rho^{-mnI_{t_k}}\right]\right)^{\frac{1}{n}}. \tag{A.6}$$

Given (A.2) and (A.3),

$$\begin{aligned}\mathbb{E}_{\theta^*}[\rho^{-mnI_{t_k}}] &= \rho^{-mn}\mathbb{P}_{\theta^*}\{I_{t_k}=1\} + \mathbb{P}_{\theta^*}\{I_{t_k}=0\} \leq \rho^{-mn}\mathbb{P}_{\theta^*}\{1-\omega(\mathbf{V}_{t_k}) \geq \rho\} + 1 \\ &\leq \rho^{-mn}[rC_2\xi_0^{-\beta_1} + (r-1)C_4\xi_0^{-\beta_2}] + 1 = \rho^{-mn}\cdot\frac{1}{16r}\rho^{mn} + 1 \leq 2.\end{aligned} \tag{A.7}$$

The proof is completed by plugging (A.7) back into (A.6).

(18) follows from (17) given $\mathbb{E}_{\theta^*}\big[\prod_{k_1=1}^{n_1}(1-\omega(\mathbf{V}_{t_{k_1}}))^m \wedge \prod_{k_2=1}^{n_2}(1-\omega(\mathbf{V}_{t_{k_2}}))^m\big] \leq \min\{\mathbb{E}_{\theta^*}[\prod_{k_1=1}^{n_1}(1-\omega(\mathbf{V}_{t_{k_1}}))^m], \mathbb{E}_{\theta^*}[\prod_{k_2=1}^{n_2}(1-\omega(\mathbf{V}_{t_{k_2}}))^m]\}$. (19) similarly follows.

Next, we show (14). The proof follows a similar procedure to Kasahara and Shimotsu (2018, pp. 19–20). Let ρ and I_{t_k} be defined as in (A.4) and (A.5), respectively. We define

$$\varepsilon_0 = rC_2\xi_0^{-\beta_1} + (r-1)C_4\xi_0^{-\beta_2}. \tag{A.8}$$

Then, $\varepsilon_0 \in (0, \frac{1}{16r})$, and

$$\mathbb{P}_{\theta^*}(1-\omega(\mathbf{V}_t) \geq \rho) \leq \varepsilon_0.$$

Using $1-\omega(\mathbf{V}_{t_k}) \leq \rho^{1-I_{t_k}}$,

$$\prod_{k=1}^{n}(1-\omega(\mathbf{V}_{t_k})) \leq \rho^{n-\sum_{k=1}^{n}I_{t_k}} = \rho^n a_n \tag{A.9}$$

with $a_n := \rho^{-\sum_{k=1}^n I_{t_k}}$. Since $\mathbf{V}_{t_k}$ is stationary and ergodic, it follows from the strong law of large numbers that $n^{-1}\sum_{k=1}^n I_{t_k} \to \mathbb{E}_{\theta^*}[I_{t_k}] < \varepsilon_0$ $\mathbb{P}_{\theta^*}-$a.s. Hence,

$$\mathbb{P}_{\theta^*}(a_n \leq \rho^{-2\varepsilon_0 n}\ ev.) = 1. \tag{A.10}$$

The proof is completed by combining (A.9) and (A.10).

Next, we show (15). Let ρ and ε_0 be defined as in (A.4) and (A.5) respectively. For each $t \geq 2r$, set ξ_t such that it satisfies $\rho^{\varepsilon_0\lfloor (t-r)/r\rfloor} = e^{-[\alpha_1 r+\alpha_2(r-1)]\xi_t}$. The existence of ξ_t is guaranteed, since $\rho^{\varepsilon_0\lfloor (t-r)/r\rfloor} \in (0,1)$, and $e^{-[\alpha_1 r+\alpha_2(r-1)]\xi_t}$ is monotone decreasing in ξ_t from 1 to 0. Then,

$$\begin{aligned}\sum_{t=1}^{\infty}\mathbb{P}_{\theta^*}(\omega(\mathbf{V}_t)\leq C_5\rho^{\varepsilon_0\lfloor (t-r)/r\rfloor}) &\leq (2r-1)+\sum_{t=2r}^{\infty}\mathbb{P}_{\theta^*}\left(\omega(\mathbf{V}_t)\leq C_5 e^{-[\alpha_1 r+\alpha_2(r-1)]\xi_t}\right)\\ \leq&(2r-1)+r\sum_{t=2r}^{\infty}\mathbb{P}_{\theta^*}(\sigma_-(\overline{\mathbf{Y}}_{t-1},X_t)\leq C_1 e^{-\alpha_1\xi_t})\\ &+(r-1)\sum_{t=2r}^{\infty}\mathbb{P}_{\theta^*}(b_-(\mathbf{Y}_{t-r}^t,X_t)\leq C_3 e^{-\alpha_2\xi_t})<\infty.\end{aligned}$$

By the Borel–Cantelli lemma, $\mathbb{P}_{\theta^*}(\omega(\mathbf{V}_t)\leq C_5\rho^{\varepsilon_0\lfloor (t-r)/r\rfloor}\ i.o.) = 0.$[9] ■

Proof of Lemma 3. First, we show (20). Write

$$\begin{aligned}&p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_i,\mathbf{X}_{-m+1}^t)\\ &=\sum_{s_t,\overline{\mathbf{s}}_{t-1},\overline{\mathbf{s}}_{-m}} g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1},\overline{\mathbf{s}}_t,X_t)q_\theta(s_t|\overline{\mathbf{s}}_{t-1},\overline{\mathbf{Y}}_{t-1},X_t)\\ &\qquad\times\mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1}=\overline{\mathbf{s}}_{t-1}|\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_{-m},\overline{\mathbf{Y}}_{-m}^{t-1},\mathbf{X}_{-m+1}^{t-1})\delta_{\overline{\mathbf{s}}_i}(\overline{\mathbf{s}}_{-m})\end{aligned}$$

and similarly for $p_\theta(Y_t|\overline{\mathbf{Y}}_{-m'}^{t-1},\overline{\mathbf{S}}_{-m'}=\overline{\mathbf{s}}_j,\mathbf{X}_{-m'+1}^t)$. It follows that

$$\begin{aligned}&\left|p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_i,\mathbf{X}_{-m+1}^t)-p_\theta(Y_t|\overline{\mathbf{Y}}_{-m'}^{t-1},\overline{\mathbf{S}}_{-m'}=\overline{\mathbf{s}}_j,\mathbf{X}_{-m'+1}^t)\right|\\ =&\Bigg|\sum_{s_t,\overline{\mathbf{s}}_{t-1},\overline{\mathbf{s}}_{-m}}\Big[g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1},\overline{\mathbf{s}}_t,X_t)q_\theta(s_t|\overline{\mathbf{s}}_{t-1},\overline{\mathbf{Y}}_{t-1},X_t)\\ &\qquad\times\mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1}=\overline{\mathbf{s}}_{t-1}|\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_{-m},\overline{\mathbf{Y}}_{-m}^{t-1},\mathbf{X}_{-m+1}^{t-1})\end{aligned}$$

[9] $i.o.$ is an abbreviation for "infinitely often."

$$\times \left(\delta_{\overline{\mathbf{s}}_i}(\overline{\mathbf{s}}_{-m}) - \mathbb{P}_\theta(\overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_{-m} | \overline{\mathbf{S}}_{-m'} = \overline{\mathbf{s}}_j, \overline{\mathbf{Y}}^{t-1}_{-m'}, \mathbf{X}^{t-1}_{-m'+1})\right)\Big]\Big|$$

$$\leq \prod_{i=1}^{\lfloor (t-1+m)/r \rfloor} \left(1 - \omega(\mathbf{V}_{-m+ri})\right) \sum_{s_t, \overline{\mathbf{s}}_{t-1}} g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{s}}_t, X_t) q_\theta(s_t | \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t).$$

Moreover, we rewrite the period predictive density as

$$\begin{aligned}
&p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_i, \mathbf{X}^t_{-m+1}) \\
&= \sum_{s_t, \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{s}}_{t-r-1}, \overline{\mathbf{s}}_{-m}} \big(g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{s}}_t, X_t) q_\theta(s_t | \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) \\
&\qquad \times \mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}_{t-1} | \overline{\mathbf{S}}_{t-r-1} = \overline{\mathbf{s}}_{t-r-1}, \overline{\mathbf{Y}}^{t-1}_{t-r-1}, \mathbf{X}^{t-1}_{t-r}) \\
&\qquad \times \mathbb{P}_\theta(\overline{\mathbf{S}}_{t-r-1} = \overline{\mathbf{s}}_{t-r-1} | \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_{-m}, \overline{\mathbf{Y}}^{t-1}_{-m}, \mathbf{X}^{t-1}_{-m+1}) \delta_{\overline{\mathbf{s}}_i}(\overline{\mathbf{s}}_{-m})\big)
\end{aligned}$$

and similarly for $p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m'}, \overline{\mathbf{S}}_{-m'} = \overline{\mathbf{s}}_j, \mathbf{X}^t_{-m'+1})$. It follows that

$$\begin{aligned}
&p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_i, \mathbf{X}^t_{-m+1}) \wedge p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m'} = \overline{\mathbf{s}}_j, \mathbf{X}^t_{-m'+1}) \\
&\geq \min_{\overline{\mathbf{s}}'_{t-1}, \overline{\mathbf{s}}_{t-r-1}} \mathbb{P}_\theta(\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}'_{t-1} | \overline{\mathbf{S}}_{t-r-1} = \overline{\mathbf{s}}_{t-r-1}, \overline{\mathbf{Y}}^{t-1}_{t-r-1}, \mathbf{X}^{t-1}_{t-r}) \\
&\qquad \times \sum_{s_t, \overline{\mathbf{s}}_{t-1}} g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{s}}_t, X_t) q_\theta(s_t | \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t) \\
&= \omega(\mathbf{V}_{t-1}) \sum_{s_t, \overline{\mathbf{s}}_{t-1}} g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, \overline{\mathbf{s}}_t, X_t) q_\theta(s_t | \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}, X_t).
\end{aligned}$$

By $|\log x - \log y| \leq |x - y| / (x \wedge y)$,

$$\begin{aligned}
&|\log p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m}, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_i, \mathbf{X}^t_{-m+1}) - \log p_\theta(Y_t | \overline{\mathbf{Y}}^{t-1}_{-m'}, \overline{\mathbf{S}}_{-m'} = \overline{\mathbf{s}}_j, \mathbf{X}^t_{-m'+1})| \\
&\qquad \leq \frac{\prod_{i=1}^{\lfloor (t-1+m)/r \rfloor} (1 - \omega(\mathbf{V}_{-m+ri}))}{\omega(\mathbf{V}_{t-1})}.
\end{aligned} \tag{A.11}$$

By (14) and (15) in Lemma (3),

$$\mathbb{P}_{\theta^*}\left(|\Delta_{t,m,\overline{\mathbf{s}}_i}(\theta) - \Delta_{t,m,\overline{\mathbf{s}}_j}(\theta)| \leq \frac{1}{C_5} \rho^{(1-3\varepsilon_0)\lfloor (t-1+m)/r \rfloor} \ ev.\right) = 1.$$

Since $(1-3\varepsilon_0)\lfloor (t-1+m)/r\rfloor \geq \lfloor (t-1+m)/r\rfloor/2 \geq \lfloor (t-1+m)/2r\rfloor \geq \lfloor (t+m)/3r\rfloor$,

$$\mathbb{P}_{\theta^*}\left(|\Delta_{t,m,\bar{\mathbf{s}}_i}(\theta)-\Delta_{t,m,\bar{\mathbf{s}}_j}(\theta)| \leq \frac{1}{C_5}\rho^{\lfloor (t+m)/3r\rfloor}\ ev.\right)=1.$$

(21) follows by replacing $\mathbb{P}_\theta(\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m}|\overline{\mathbf{S}}_{-m'}=\bar{\mathbf{s}}_j,\overline{\mathbf{Y}}_{-m'}^{t-1},\mathbf{X}_{-m'+1}^{t})$ with $\mathbb{P}_\theta(\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m}|\overline{\mathbf{Y}}_{-m}^{t-1},\mathbf{X}_{-m+1}^{t})$. (22) follows from

$$b_-(\mathbf{Y}_{t-r}^t,X_t)\leq p_\theta(Y_t|\overline{\mathbf{Y}}_{-m}^{t-1},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_i,\mathbf{X}_{-m+1}^t)\leq b_+.$$

■

A.2 Detailed proof of propositions and theorems in Section 4

Proof of continuity in Proposition 1. Show that $\Delta_{0,m,\bar{\mathbf{s}}_{-m}}(\theta)$ is continuous with respect to θ.

$$p_\theta(Y_0|\overline{\mathbf{Y}}_{-m}^{-1},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m+1}^0)=\frac{p_\theta(\mathbf{Y}_{-m+1}^0|\overline{\mathbf{Y}}_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m+1}^0)}{p_\theta(\mathbf{Y}_{-m+1}^{-1}|\overline{\mathbf{Y}}_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m+1}^{-1})},$$

$$\begin{aligned}&p_\theta(\mathbf{Y}_{-m+1}^j|\overline{\mathbf{Y}}_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m+1}^j)\\&=\sum_{\mathbf{s}_{-m+1}^j}\prod_{\ell=-m+1}^{j}q_\theta(s_\ell|\bar{\mathbf{s}}_{\ell-1},\overline{\mathbf{Y}}_{\ell-1},X_\ell)\prod_{\ell=-m+1}^{j}g_\theta(Y_\ell|\overline{\mathbf{Y}}_{\ell-1},\bar{\mathbf{s}}_\ell,X_\ell)\end{aligned}$$

for $j=0,-1$. Continuity follows from Assumption (A5). ■

Proof of Proposition 2.

$$\begin{aligned}\ell(\theta)&=\mathbb{E}_{\theta^*}\left[\lim_{m\to\infty}\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0,\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m+1}^1)\right]\\&=\lim_{m\to\infty}\mathbb{E}_{\theta^*}\left[\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0,\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m+1}^1)\right]\\&=\lim_{m\to\infty}\mathbb{E}_{\theta^*}\left[\mathbb{E}_{\theta^*}[\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0,\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m+1}^1)|\overline{\mathbf{Y}}_{-m}^0,\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m+1}^1]\right].\end{aligned}\tag{A.12}$$

It follows that

$$\ell(\theta^*)-\ell(\theta)$$

$$= \lim_{m\to\infty} \mathbb{E}_{\theta^*}\left[\mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^1)}{p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^1)}\middle|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^1\right]\right] \geq 0.$$

The nonnegativity follows because Kullback–Leibler divergence is nonnegative, and thus, the limit of its expectation is nonnegative. Next, we show that θ^* is the unique maximizer. The proof closely follows Douc et al. (2004, pp. 2269–2270). Assume $\ell(\theta) = \ell(\theta^*)$. For any $t \geq 1$ and $m \geq 0$,

$$\mathbb{E}_{\theta^*}[\log p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)] = \sum_{k=1}^t \mathbb{E}_{\theta^*}[\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^k)].$$

By (A.12) and stationarity, $\lim_{m\to\infty} \mathbb{E}_{\theta^*}[\log p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)] = t\ell(\theta)$. For $1 \leq k \leq t - r + 1$,

$$\begin{aligned}
0 = t(\ell(\theta^*) - \ell(\theta)) &= \lim_{m\to\infty} \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}{p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}\right] \\
&= \limsup_{m\to\infty} \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_{t-k+1}^t|\overline{\mathbf{Y}}_{t-k}, \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}{p_\theta(\overline{\mathbf{Y}}_{t-k+1}^t|\overline{\mathbf{Y}}_{t-k}, \mathbf{Y}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}\right. \\
&\quad \left. + \log \frac{p_{\theta^*}(\overline{\mathbf{Y}}_{t-k}|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}{p_\theta(\overline{\mathbf{Y}}_{t-k}|\overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)} + \log \frac{p_{\theta^*}(\mathbf{Y}_1^{t-k-r}|\mathbf{Y}_{t-k-r+1}^t, \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}{p_\theta(\mathbf{Y}_1^{t-k-r}|\mathbf{Y}_{t-k-r+1}^t, \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}\right] \\
&\geq \limsup_{m\to\infty} \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_{t-k+1}^t|\overline{\mathbf{Y}}_{t-k}, \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}{p_\theta(\overline{\mathbf{Y}}_{t-k+1}^t|\overline{\mathbf{Y}}_{t-k}, \mathbf{Y}_{-m}^0, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m+1}^t)}\right] \\
&= \limsup_{m\to\infty} \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m-t+k}^{-t+k}, \overline{\mathbf{S}}_{-m-t+k}, \mathbf{X}_{-m+1-t+k}^k)}{p_{\theta^*}(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m-t+k}^{-t+k}, \overline{\mathbf{S}}_{-m-t+k}, \mathbf{X}_{-m+1-t+k}^k)}\right].
\end{aligned}$$

This holds when we let $t \to \infty$. It suffices to show that for all $t \geq 0$ and all $\theta \in \Theta$,

$$\lim_{k\to\infty} \sup_{m\geq k} \left|\mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t)}{p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t)}\right] - \mathbb{E}_{\theta^*}\left[\log \frac{p_{\theta^*}(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)}{p_\theta(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)}\right]\right| = 0. \tag{A.13}$$

From (A.13) and the previous inequality, if $\ell(\theta^*) = \ell(\theta)$, then $\mathbb{E}_{\theta^*}\big[\log\big(p_{\theta^*}(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)/p_\theta(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)\big)\big] = 0$. The laws $\mathbb{P}_{\theta^*}(\mathbf{Y}_1^t \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)$ and $\mathbb{P}_\theta(\mathbf{Y}_1^t \in \cdot|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)$ agree. From Assumption (A6), $\theta = \theta^*$.

Next we show (A.13). Define $U_{k,m}(\theta) = \log p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t)$, and $U(\theta) = \log p_\theta(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)$.

It is enough to show that for all $\theta \in \Theta$,

$$\lim_{k\to\infty} \mathbb{E}_{\theta^*}\left[\sup_{m\geq k} |U_{k,m}(\theta) - U(\theta)|\right] = 0.$$

Put

$$A_{k,m} = p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t), \quad A = p_\theta(\mathbf{Y}_{-r+1}^t | \mathbf{X}_{-r+1}^t),$$
$$B_{k,m} = p_\theta(\mathbf{Y}_{-r+1}^0 | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^0), \quad B = p_\theta(\mathbf{Y}_{-r+1}^0 | \mathbf{X}_{-r+1}^0).$$

Then

$$\begin{aligned} |p_\theta(\mathbf{Y}_1^t | \overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t) - p_\theta(\mathbf{Y}_1^t | \mathbf{Y}_{-r+1}^0, \mathbf{X}_{-r+1}^t)| &= \left|\frac{A_{k,m}}{B_{k,m}} - \frac{A}{B}\right| \\ &\leq \frac{B|A_{k,m} - A| + A|B_{k,m} - B|}{BB_{k,m}}. \end{aligned} \tag{A.14}$$

For all $t \geq 0$ and $k \geq r$, write

$$\begin{aligned} &p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t) \\ &= \int\int p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Z}}_{-r} = \overline{\mathbf{z}}_{-r}, \mathbf{X}_{-r+1}^t) \mathbb{P}_\theta(d\overline{\mathbf{z}}_{-r} | \overline{\mathbf{Z}}_{-k+1} = \overline{\mathbf{z}}_{-k+1}, \mathbf{X}_{-k+2}^{-r}) \\ &\qquad \times \mathbb{P}_\theta(d\overline{\mathbf{z}}_{-k+1} | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^{-k+1}) \\ &p_\theta(\mathbf{Y}_{-r+1}^t | \mathbf{X}_{-r+1}^t) \\ &= \int\int p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Z}}_{-r} = \overline{\mathbf{z}}_{-r}, \mathbf{X}_{-r+1}^t) \mathbb{P}_\theta(d\overline{\mathbf{z}}_{-r} | \mathbf{X}_{-2r+1}^{-r}) \\ &\qquad \times \mathbb{P}_\theta(d\overline{\mathbf{z}}_{-k+1} | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^{-k+1}), \end{aligned}$$

where $\overline{\mathbf{Z}}_t$ is defined as in (26). The second expression holds because conditionally on X_t, $\{X_k\}_{k\geq t+1}$ is independent of $\{Z_k\}_{k\leq t}$, and conditionally on X_t, $\{X_k\}_{k\leq t-1}$ is independent of $\{Z_k\}_{k\geq t}$. An upper bound of their difference is

$$\begin{aligned} &|p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t) - p_\theta(\mathbf{Y}_{-r+1}^t | \mathbf{X}_{-r+1}^t)| \\ &\leq \int\int p_\theta(\mathbf{Y}_{-r+1}^t | \overline{\mathbf{Z}}_{-r} = \overline{\mathbf{z}}_{-r}, \mathbf{X}_{-r+1}^t) \end{aligned}$$

$$\times\left|\mathbb{P}_\theta(d\overline{\mathbf{z}}_{-r}|\overline{\mathbf{Z}}_{-k+1}=\overline{\mathbf{z}}_{-k+1},\mathbf{X}_{-k+2}^{-r})-\mathbb{P}_\theta(d\overline{\mathbf{z}}_{-r}|\mathbf{X}_{-2r+1}^{-r})\right|\mathbb{P}_\theta(d\overline{\mathbf{z}}_{-k+1}|\overline{\mathbf{Y}}_{-m}^{-k},\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})$$
$$\leq b_+^{t+r}\int\|\mathbb{P}_\theta(\overline{\mathbf{z}}_{-r}\in\cdot|\overline{\mathbf{Z}}_{-k+1}=\overline{\mathbf{z}}_{-k+1},\mathbf{X}_{-k+2}^{-r})-\mathbb{P}_\theta(\overline{\mathbf{z}}_{-r}\in\cdot|\mathbf{X}_{-2r+1}^{-r})\|_{TV}$$
$$\times\,\mathbb{P}_\theta(d\overline{\mathbf{z}}_{-k+1}|\overline{\mathbf{Y}}_{-m}^{-k},\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})$$

$\|\mathbb{P}_\theta(\overline{\mathbf{z}}_{-r}\in\cdot|\overline{\mathbf{Z}}_{-k+1}=\overline{\mathbf{z}}_{-k+1},\mathbf{X}_{-k+2}^{-r})-\mathbb{P}_\theta(\overline{\mathbf{z}}_{-r}\in\cdot|\mathbf{X}_{-2r+1}^{-r})\|_{TV}$ goes to zero as k goes to infinity, owing to the Markovian property of $\overline{\mathbf{Z}}_t$ conditional on $\mathbf{X}_{-\infty}^{+\infty}$ and Assumption (A2). Thus

$$\sup_{m\geq k}|p_\theta(\mathbf{Y}_{-r+1}^{t}|\overline{\mathbf{Y}}_{-m}^{-k},\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})-p_\theta(\mathbf{Y}_{-r+1}^{t}|\mathbf{X}_{-r+1}^{t})|\to 0,\mathbb{P}_{\theta^*}-a.s.\text{ as }k\to\infty. \tag{A.15}$$

Using Assumption (A3),

$$\begin{aligned}B&=p_\theta(\mathbf{Y}_{-r+1}^{0}|\mathbf{X}_{-r+1}^{0})\\&=\int\sum_{\overline{\mathbf{s}}}\left(\prod_{k=-r+1}^{0}p_\theta(Y_k|\overline{\mathbf{Y}}_{-r}^{k-1},\overline{\mathbf{S}}_{-r}=\overline{\mathbf{s}},\mathbf{X}_{-r+1}^{k})\right)\\&\qquad\qquad\times\mathbb{P}_\theta(\overline{\mathbf{S}}_{-r}=\overline{\mathbf{s}}|\overline{\mathbf{Y}}_{-r},\mathbf{X}_{-r+1}^{0})\mathbb{P}_\theta(\overline{\mathbf{Y}}_{-r}\in d\overline{\mathbf{y}}_{-r}|\mathbf{X}_{-r+1}^{0})\\&\geq\int\prod_{k=-r+1}^{0}b_-(\mathbf{Y}_{k-r}^{k},X_k)\mathbb{P}_\theta(\overline{\mathbf{Y}}_{-r}\in d\overline{\mathbf{y}}_{-r}|\mathbf{X}_{-r+1}^{0})>0\end{aligned}$$

By (A.15), with $\mathbb{P}_{\theta^*}$-probability arbitrarily close to 1, $B_{k,m}$ is uniformly bounded away from zero for $m\geq k$ and k sufficiently large. By (A.14) and (A.15),

$$\lim_{k\to\infty}\sup_{m\geq k}|p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0,\overline{\mathbf{Y}}_{-m}^{-k},\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})-p_\theta(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^{0},\mathbf{X}_{-r+1}^{t})|=0,$$
$$\text{in }\mathbb{P}_\theta^*-\text{probability}.$$

Since $p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0,\overline{\mathbf{Y}}_{-m}^{-k},\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})=\sum_{\overline{\mathbf{s}}}p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0,\overline{\mathbf{S}}_0=\overline{\mathbf{s}},\mathbf{X}_1^t)\mathbb{P}_\theta(\overline{\mathbf{S}}_0=\overline{\mathbf{s}}|\overline{\mathbf{Y}}_0,\overline{\mathbf{S}}_{-m},\mathbf{X}_{-m}^{t})$, it is bounded below by $\prod_{k=1}^{t}b_-(\mathbf{Y}_{k-r}^{k},X_k)$ and bounded above by b_+^t. The same lower bound can be attained for $p_\theta(\mathbf{Y}_1^t|\mathbf{Y}_{-r+1}^{0},\mathbf{X}_{-r+1}^{t})$. Use the inequality $|\log x-\log y|\leq|x-y|/(x\wedge y)$,

$$\lim_{k\to\infty}\sup_{m\geq k}|U_{k,m}(\theta)-U(\theta)|=0,\ \text{in }\mathbb{P}_\theta^*-\text{probability}$$

Using the bounds of $p_\theta(\mathbf{Y}_1^t|\overline{\mathbf{Y}}_0, \overline{\mathbf{Y}}_{-m}^{-k}, \overline{\mathbf{S}}_{-m}, \mathbf{X}_{-m}^t)$,

$$\mathbb{E}_{\theta^*}\left[\sup_k \sup_{m\geq k} |U_{k,m}(\theta)|\right] < \infty.$$

The proof is completed by applying the bounded convergence theorem.

■

A.3 Detailed proof of lemmas and theorems in Subsection 5.1

To show the approximation of the score function, we first need to show the mixing rate of the time-reversed Markov process.

LEMMA 8 (Minorization condition of the time-reversed Markov process)**.** *Let $m, n \in \mathbb{Z}$ with $-m \leq n$ and $\theta \in \Theta$. Conditionally on $\overline{\mathbf{Y}}_{-m}^n$ and $\mathbf{X}_{-m}^n$, $\{\overline{\mathbf{S}}_{n-k}\}_{0\leq k\leq n+m}$ satisfies the Markov property. Assume (A3). Then, for all $r \leq k \leq n+m$, a function $\tilde{\mu}_k(\overline{\mathbf{Y}}_{-m}^{n-k+r}, \mathbf{X}_{-m}^{n-k+r}, A)$ exists such that:*

- *(i) for any $A \in \mathcal{P}(\overline{\mathbb{S}})$, $(\overline{\mathbf{y}}_{-m}^{n-k+r}, \mathbf{x}_{-m}^{n-k+r}) \to \tilde{\mu}_k(\overline{\mathbf{y}}_{-m}^{n-k+r}, \mathbf{x}_{-m}^{n-k+r}, A)$ is a Borel function; and*
- *(ii) for any $\overline{\mathbf{y}}_{-m}^{n-k+r}$ and $\mathbf{x}_{-m}^{n-k+r}$, $\tilde{\mu}_k(\overline{\mathbf{y}}_{-m}^{n-k+r}, \mathbf{x}_{-m}^{n-k+r}, \cdot)$ is a probability measure on $\mathcal{P}(\overline{\mathbb{S}})$. Moreover, for $A \in \mathcal{P}(\overline{\mathbb{S}})$, the following holds:*

$$\min_{\overline{\mathbf{s}}_{n-k+r}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_{n-k} \in A|\overline{\mathbf{S}}_{n-k+r} = \overline{\mathbf{s}}_{n-k+r}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n) \\ \geq \omega(\mathbf{Y}_{n-k-r+1}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r}) \cdot \tilde{\mu}_k(\overline{\mathbf{Y}}_{-m}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}, A)$$

where $\omega(\mathbf{Y}_{n-k-r+1}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r}) \triangleq \frac{\prod_{\ell=n-k+1}^{n-k+r-1} b_-(\mathbf{Y}_{\ell-r}^{\ell}, X_\ell) \prod_{\ell=n-k+1}^{n-k+r} \sigma_-(\overline{\mathbf{Y}}_{\ell-1}, X_\ell)}{{b_+}^{r-1}}$.

Proof. To observe the Markovian property, for $2 \leq k \leq m+n$,

$$p_\theta(\overline{\mathbf{S}}_{n-k}|\overline{\mathbf{S}}_{n-k+1}^n, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n) \tag{A.16}$$

$$= \frac{p_\theta(\overline{\mathbf{S}}_{n-k+2}^n, \overline{\mathbf{Y}}_{n-k+1}^n|\overline{\mathbf{S}}_{n-k+1}, \overline{\mathbf{Y}}_{n-k}, \mathbf{X}_{-m}^n) p_\theta(\overline{\mathbf{S}}_{n-k}^{n-k+1}, \overline{\mathbf{Y}}_{-m}^{n-k}|\mathbf{X}_{-m}^n)}{p_\theta(\overline{\mathbf{S}}_{n-k+2}^n, \overline{\mathbf{Y}}_{n-k+1}^n|\overline{\mathbf{S}}_{n-k+1}, \overline{\mathbf{Y}}_{n-k}, \mathbf{X}_{-m}^n) p_\theta(\overline{\mathbf{S}}_{n-k+1}, \overline{\mathbf{Y}}_{-m}^{n-k}|\mathbf{X}_{-m}^n)} \tag{A.17}$$

$$= p_\theta(\overline{\mathbf{S}}_{n-k}|\overline{\mathbf{S}}_{n-k+1}, \overline{\mathbf{Y}}_{-m}^{n-k}, \mathbf{X}_{-m}^{n-k+1}). \tag{A.18}$$

To observe the minorization condition, note that

$$\mathbb{P}_\theta(\overline{\mathbf{S}}_{n-k} \in A|\overline{\mathbf{S}}_{n-k+r}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{X}_{-m}^{n}) = \mathbb{P}_\theta(\overline{\mathbf{S}}_{n-k} \in A|\overline{\mathbf{S}}_{n-k+r}, \overline{\mathbf{Y}}_{-m}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r})$$
$$= \frac{\sum_{\overline{\mathbf{s}}_{n-k}\in A} p_\theta(\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{-m}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}) p_\theta(\overline{\mathbf{S}}_{n-k+r}|\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{n-k}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r})}{\sum_{\overline{\mathbf{s}}_{n-k}\in \overline{\mathbb{S}}} p_\theta(\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{-m}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}) p_\theta(\overline{\mathbf{S}}_{n-k+r}|\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{n-k}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r})}.$$

Since $p_\theta(\overline{\mathbf{S}}_{n-k+r}|\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{n-k}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r}) \leq 1$, and for all $\overline{\mathbf{s}}_{n-k+r}, \overline{\mathbf{s}}_{n-k} \in \overline{\mathbb{S}}$,

$$\mathbb{P}_\theta(\overline{\mathbf{S}}_{n-k+r}=\overline{\mathbf{s}}_{n-k+r}|\overline{\mathbf{S}}_{n-k}=\overline{\mathbf{s}}_{n-k}, \overline{\mathbf{Y}}_{n-k}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r}) \geq \omega(\mathbf{Y}_{n-k-r+1}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r}),$$

it readily follows that

$$\mathbb{P}_\theta(\overline{\mathbf{S}}_{n-k} \in A|\overline{\mathbf{S}}_{n-k+r}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{X}_{-m}^{n}) \geq \omega(\mathbf{Y}_{n-k-r+1}^{n-k+r-1}, \mathbf{X}_{n-k+1}^{n-k+r})\tilde{\mu}_k(\mathbf{Y}_{-m-r+1}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}, A)$$

with

$$\tilde{\mu}_k(\mathbf{Y}_{-m-r+1}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}, A) \triangleq p_\theta(\overline{\mathbf{S}}_{n-k} \in A|\overline{\mathbf{Y}}_{-m}^{n-k+r-1}, \mathbf{X}_{-m}^{n-k+r}).$$

■

COROLLARY 2 (Uniform ergodicity of the time-reversed Markov process)**.** *Assume (A3). Let $m, n \in \mathbb{Z}$, $-m \leq n$, and $\theta \in \Theta$. Then, for $-m \leq k \leq n$, for all probability measures μ_1 and μ_2 defined on $\mathcal{P}(\overline{\mathbb{S}})$ and all $\overline{\mathbf{Y}}_{-m}^{n}$,*

$$\left\| \sum_{\overline{\mathbf{s}}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_n=\overline{\mathbf{s}}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{X}_{-m}^{n})\mu_1(\overline{\mathbf{s}}) - \sum_{\overline{\mathbf{s}}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_n=\overline{\mathbf{s}}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{X}_{-m}^{n})\mu_2(\overline{\mathbf{s}}) \right\|_{TV}$$
$$\leq \prod_{i=1}^{\lfloor (n-k)/r \rfloor} \left(1-\omega\left(\mathbf{Y}_{n-ri-r+1}^{n-ri+r-1}, \mathbf{X}_{n-ri+1}^{n-ri+r}\right)\right) = \prod_{i=1}^{\lfloor (n-k)/r \rfloor} (1-\omega(\mathbf{V}_{n+r-ri})).$$

We follow almost the same procedure to obtain

$$\left\| \sum_{\overline{\mathbf{s}}\in\overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_n=\overline{\mathbf{s}}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{X}_{-m}^{n}, \overline{\mathbf{S}}_{-m})\mu_1(\overline{\mathbf{s}}) \right.$$

$$
- \sum_{\overline{\mathbf{s}} \in \overline{\mathbb{S}}} \mathbb{P}_\theta(\overline{\mathbf{S}}_k \in \cdot | \overline{\mathbf{S}}_n = \overline{\mathbf{s}}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{X}_{-m}^n, \overline{\mathbf{S}}_{-m}) \mu_2(\overline{\mathbf{s}}) \Bigg\|_{TV}
$$

$$
\leq \prod_{i=1}^{\lfloor (n-k)/r \rfloor} \left(1 - \omega\left(\mathbf{Y}_{n-ri-r+1}^{n-ri+r-1}, \mathbf{X}_{n-ri+1}^{n-ri+r}\right)\right) = \prod_{i=1}^{\lfloor (n-k)/r \rfloor} (1 - \omega(\mathbf{V}_{n+r-ri})).
$$

First, we show that the initial state $\overline{\mathbf{s}}$ does not show up in the limit. We define $\|\phi_{\theta^*,t}\|_\infty = \max_{s_t, \overline{\mathbf{s}}_{t-1}} \|\phi_{\theta^*}((s_t, Y_t), (\overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{t-1}), X_t)\|$.

LEMMA 9. *Assume (A2)–(A4) and (A7)–(A8). Then, for $-m' < -m < k \leq t$,*

$$
\mathbb{E}_{\theta^*}\left[\left\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]\right\|^2\right] \leq 8\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}} \rho^{\lfloor (k+m-1)/2r \rfloor}, \tag{A.19}
$$

$$
\mathbb{E}_{\theta^*}\left[\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m'}^t, \mathbf{X}_{-m'}^t]\|^2\right] \leq 8\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}} \rho^{\lfloor (k+m-1)/2r \rfloor}, \tag{A.20}
$$

$$
\mathbb{E}_{\theta^*}\left[\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1}]\|^2\right] \leq 8\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}} \rho^{\lfloor (t-1-k)/2r \rfloor}. \tag{A.21}
$$

Proof. First, show (A.19).

$$
\begin{aligned}
&\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]\| \\
\leq & 2\|\phi_{\theta^*,k}\|_\infty \Bigg\| \sum_{\overline{\mathbf{s}}_{-m}} \mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_{k-1} \in \cdot | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_{-m}, \mathbf{X}_{-m}^t) \delta_{\overline{\mathbf{s}}}(\overline{\mathbf{s}}_{-m}) \\
& - \sum_{\overline{\mathbf{s}}_{-m}} \mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_{k-1} \in \cdot | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_{-m}, \mathbf{X}_{-m}^t) \mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}_{-m} | \overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t) \Bigg\| \\
\leq & 2\|\phi_{\theta^*,k}\|_\infty \prod_{i=1}^{\lfloor (k+m-1)/r \rfloor} (1 - \omega(\mathbf{V}_{-m+ri}))
\end{aligned}
$$

Using Lemma 2 and Hölder's inequality, the second moment is bounded by

$$
4\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,k}\|_\infty^2 \prod_{i=1}^{\lfloor (k+m-1)/r \rfloor} (1 - \omega(\mathbf{V}_{-m+ri}))^2\right]
$$

$$
\begin{aligned}
&\leq 4\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}}\left\{\mathbb{E}_{\theta^*}\left[\prod_{i=1}^{\lfloor (k+m-1)/r \rfloor}(1-\omega(\mathbf{V}_{-m+ri}))^4\right]\right\}^{\frac{1}{2}} \\
&\leq 8\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}}\rho^{\lfloor (k+m-1)/2r \rfloor}.
\end{aligned}
$$

We can show (A.20) by replacing $\mathbb{P}(\overline{\mathbf{S}}_{k-1} \in \cdot|\overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^t)$ with $\mathbb{P}(\overline{\mathbf{S}}_{k-1} \in \cdot|\overline{\mathbf{Y}}_{-m'}^t, \mathbf{X}_{-m'}^t)$. For (A.21),

$$
\begin{aligned}
&\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1}]\| \\
\leq & 2\|\phi_{\theta^*,k}\|_\infty \Big\| \sum_{\overline{\mathbf{s}}_{t-1}} \mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1})\mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_{t-1}|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t) \\
&\qquad - \sum_{\overline{\mathbf{s}}_{t-1}} \mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_k \in \cdot|\overline{\mathbf{S}}_{t-1} = \overline{\mathbf{s}}_{t-1}, \overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1})\mathbb{P}_{\theta^*}(\overline{\mathbf{S}}_{t-1}|\overline{\mathbf{Y}}_{-m}^{t-1}, \mathbf{X}_{-m}^{t-1}) \Big\| \\
\leq & 2\|\phi_{\theta^*,k}\|_\infty \prod_{i=1}^{\lfloor (t-1-k)/r \rfloor}(1-\omega(\mathbf{V}_{t-1+r-ri})).
\end{aligned}
$$

The bound for its second moment follows similarly to the proof for (A.19). ∎

Since $\{\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]\}_{m\geq 0}$ is a martingale, by Jensen's inequality, $\{\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}$ is a submartingale. Moreover, for any m,

$$
\begin{aligned}
\mathbb{E}_{\theta^*}\left[\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]\|^2\right] &\leq \mathbb{E}_{\theta^*}[\mathbb{E}_{\theta^*}[\|\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)\|^2|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]] \\
&= \mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)\|^2\right] < \infty
\end{aligned}
$$

under A8. Then by the martingale convergence theorem (see, e.g., Shiryaev (1996, p.508)),

$$
\begin{aligned}
&\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-m}^t, \mathbf{X}_{-m}^t]\|^2 \\
&\to \|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^t, \mathbf{X}_{-\infty}^t]\|^2, \quad \mathbb{P}_{\theta^*} - a.s.
\end{aligned} \tag{A.22}
$$

as $m \to \infty$ and

$$
\mathbb{E}_{\theta^*}\left[\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^t, \mathbf{X}_{-\infty}^t]\|^2\right] < \infty. \tag{A.23}
$$

On the other hand, by setting $m = \infty$ in (A.21),

$$\mathbb{E}_{\theta^*}\left[\sum_{k=-\infty}^{t-1} \|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^{t}, \mathbf{X}_{-\infty}^{t}] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k)|\overline{\mathbf{Y}}_{-\infty}^{t-1}, \mathbf{X}_{-\infty}^{t-1}]\|^2\right] < \infty. \tag{A.24}$$

Combining (A.23) and (A.24), $\dot{\Delta}_{t,\infty}(\theta^*)$ is well defined in $L^2(\mathbb{P}_{\theta^*})$.

Proof of Lemma 4. It suffices to show $\lim_{n\to\infty} \mathbb{E}_{\theta^*}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\dot{\Delta}_{t,0,\bar{\mathbf{s}}}(\theta^*) - \dot{\Delta}_{t,0}(\theta^*)\right)\right\|^2\right] = 0$ and $\lim_{n\to\infty} \mathbb{E}_{\theta^*}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\dot{\Delta}_{t,0}(\theta^*) - \dot{\Delta}_{t,\infty}(\theta^*)\right)\right\|^2\right] = 0$.

First, we show the first term. From (25),

$$\sum_{t=1}^{n} \dot{\Delta}_{t,0,\bar{\mathbf{s}}}(\theta^*) = \sum_{t=1}^{n} \mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}, \mathbf{X}_{0}^{t}\right]$$

and similarly, from (26),

$$\sum_{t=1}^{n} \dot{\Delta}_{t,0}(\theta^*) = \sum_{t=1}^{n} \mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \mathbf{X}_{0}^{t}\right].$$

Using the Minkowski inequality and (A.19),

$$\begin{aligned}
&\left\{\mathbb{E}_{\theta^*}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\dot{\Delta}_{t,0,\bar{\mathbf{s}}}(\theta^*) - \dot{\Delta}_{t,0}(\theta^*)\right)\right\|^2\right]\right\}^{\frac{1}{2}} \\
=&\left\{\mathbb{E}_{\theta^*}\left[\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}, \mathbf{X}_{0}^{t}\right] - \mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \mathbf{X}_{0}^{t}\right]\right)\right\|^2\right]\right\}^{\frac{1}{2}} \\
\leq&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\mathbb{E}_{\theta^*}\left[\left\|\mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \overline{\mathbf{S}}_0 = \bar{\mathbf{s}}, \mathbf{X}_{0}^{t}\right] - \mathbb{E}_{\theta^*}\left[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \mathbf{X}_{0}^{t}\right]\right\|^2\right]\right)^{\frac{1}{2}} \\
\leq&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[8\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{2}}\rho^{\lfloor(t-1)/2r\rfloor}\right]^{\frac{1}{2}} \leq \frac{1}{\sqrt{n}}4\left(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\right)^{\frac{1}{4}}\frac{1}{\rho(1-\rho^{\frac{1}{4r}})}
\end{aligned}$$

which converges to zero as $n \to \infty$. Next, consider the second term.

$$\begin{aligned}
&\left(\mathbb{E}_{\theta^*}\left[\left\|\dot{\Delta}_{t,0}(\theta^*) - \dot{\Delta}_{t,\infty}(\theta^*)\right\|^2\right]\right)^{\frac{1}{2}} \\
\leq&\left(\mathbb{E}_{\theta^*}\left[\left\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{-\infty}^{t}, \mathbf{X}_{-\infty}^{t}] - \mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_t, \overline{\mathbf{Z}}_{t-1}, X_t)|\overline{\mathbf{Y}}_{0}^{t}, \mathbf{X}_{0}^{t}]\right\|^2\right]\right)^{\frac{1}{2}}
\end{aligned}$$

$$+\sum_{k=1}^{t-1}\Big(\mathbb{E}_{\theta^*}\Big[\Big\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_{-\infty}^t,\mathbf{X}_{-\infty}^t]-\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_{-\infty}^{t-1},\mathbf{X}_{-\infty}^{t-1}]$$

$$-\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_0^t,\mathbf{X}_0^t]-\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_0^{t-1},\mathbf{X}_0^{t-1}]\Big\|^2\Big]\Big)^{\frac{1}{2}}$$

$$+\sum_{k=-\infty}^{0}\Big(\mathbb{E}_{\theta^*}\Big[\Big\|\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_{-\infty}^t,\mathbf{X}_{-\infty}^t]-\mathbb{E}_{\theta^*}[\phi_{\theta^*}(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}_{-\infty}^{t-1},\mathbf{X}_{-\infty}^{t-1}]\Big\|^2\Big]\Big)^{\frac{1}{2}}$$

$$\le 4\big(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\big)^{\frac{1}{4}}\left(\rho^{\lfloor(t-1)/4r\rfloor}+2\sum_{k=1}^{t-1}\rho^{\lfloor(k-1)/4r\rfloor}\wedge\rho^{\lfloor(t-1-k)/4r\rfloor}+\sum_{k=-\infty}^{0}\rho^{\lfloor(t-1-k)/4r\rfloor}\right)$$

$$\le 4\big(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\big)^{\frac{1}{4}}\rho^{-1}\left(\rho^{\frac{t-1}{4r}}+2\sum_{k\ge\frac{t}{2}}\rho^{\frac{k-1}{4r}}+2\sum_{k\le\frac{t}{2}}\rho^{\frac{t-1-k}{4r}}+\sum_{k=-\infty}^{0}\rho^{\frac{t-1-k}{4r}}\right)$$

$$\le 4\big(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\big)^{\frac{1}{4}}\frac{6-\rho^{\frac{1}{4r}}}{1-\rho^{\frac{1}{4r}}}\rho^{\frac{\frac{t}{2}-4r-1}{4r}}.$$

Using Minkowski inequality,

$$\left\{\mathbb{E}_{\theta^*}\left[\Big\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\dot{\Delta}_{t,0}(\theta^*)-\dot{\Delta}_{t,\infty}(\theta^*)\big)\Big\|^2\right]\right\}^{\frac{1}{2}}\le\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\mathbb{E}_{\theta^*}\left[\Big\|\dot{\Delta}_{t,0}(\theta^*)-\dot{\Delta}_{t,\infty}(\theta^*)\Big\|^2\right]\right)^{\frac{1}{2}}$$

$$\le\frac{1}{\sqrt{n}}\sum_{t=1}^{n}4\big(\mathbb{E}_{\theta^*}\left[\|\phi_{\theta^*,0}\|_\infty^4\right]\big)^{\frac{1}{4}}\frac{6-\rho^{\frac{1}{4r}}}{1-\rho^{\frac{1}{4r}}}\rho^{\frac{\frac{t}{2}-4r-1}{4r}}$$

which converges to 0 as $n\to\infty$. ■

A.4 Detailed proof of theorems in Subsection 5.2

The following Lemma 10 extends (14).

LEMMA 10. *Assume (A2)–(A4). For $1\le m\le n$, $\{t_{k_{i_1}}\}_{1\le i_1\le a_m}$ and $\{t_{h_{i_2}}\}_{1\le i_2\le b_m}$ are two sequences of integers satisfying (i) $t_{k_{i_1}}\ne t_{k_{i_1'}}$ for $1\le i_1,i_1'\le a_m$ and $i_1\ne i_1'$, (ii) $t_{h_{i_2}}\ne t_{h_{i_2'}}$ for $1\le i_2,i_2'\le b_m$ and $i_2\ne i_2'$, and (iii) $t_{k_{i_1}}\ne t_{h_{i_2}}$ for all $1\le i_1\le a_m$ and $1\le i_2\le b_m$, then it holds that for the same ε_0 and ρ defined as in Lemma 2,*

$$\mathbb{P}_{\theta^*}\left\{\sum_{m=1}^{n}\Big[\prod_{i_1=1}^{a_m}(1-\omega(\mathbf{V}_{t_{k_{i_1}}}))\wedge\prod_{i_2=1}^{b_m}(1-\omega(\mathbf{V}_{t_{h_{i_2}}}))\Big]\le\rho^{-2\varepsilon_0(\bar{t}-\underline{t}+1)}\sum_{m=1}^{n}\rho^{a_m\vee b_m}\ ev.\right\}=1,\tag{A.25}$$

where $\underline{t}$ and $\bar{t}$ are integers such that $\underline{t} \le t_{k_{i_1}}, t_{h_{i_2}} \le \bar{t}$, for all $1 \le i_1 \le a_m$ and $1 \le i_2 \le b_m$, and all $1 \le m \le n$.

For $1 \le m \le n$, $\{t_{k_{i_1}}\}_{1 \le i_1 \le a_m}$, $\{t_{h_{i_2}}\}_{1 \le i_2 \le b_m}$, and $\{t_{\ell_{i_3}}\}_{1 \le i_3 \le c_m}$ are three sequences of integers satisfying (i) $t_{k_{i_1}} \ne t_{k_{i_1'}}$ for $1 \le i_1, i_1' \le a_m$ and $i_1 \ne i_1'$, (ii) $t_{h_{i_2}} \ne t_{h_{i_2'}}$ for $1 \le i_2, i_2' \le b_m$ and $i_2 \ne i_2'$, (iii) $t_{\ell_{i_3}} \ne t_{\ell_{i_3'}}$ for $1 \le i_3, i_3' \le c_m$ and $i_3 \ne i_3'$, and (iv) $t_{k_{i_1}} \ne t_{h_{i_2}}$, $t_{k_{i_1}} \ne t_{\ell_{i_3}}$ and $t_{h_{i_2}} \ne t_{\ell_{i_3}}$ for all $1 \le i_1 \le a_m$, $1 \le i_2 \le b_m$, and $1 \le i_3 \le c_m$, then it holds that for the same ε_0 and ρ defined as in Lemma 2,

$$\mathbb{P}_{\theta^*}\Bigg\{ \sum_{m=1}^{n} \Bigg[\prod_{i_1=1}^{a_m} (1 - \omega(\mathbf{V}_{t_{k_{i_1}}})) \wedge \prod_{i_2=1}^{b_m} (1 - \omega(\mathbf{V}_{t_{h_{i_2}}})) \wedge \prod_{i_3=1}^{c_m} (1 - \omega(\mathbf{V}_{t_{\ell_{i_3}}})) \Bigg] \le \rho^{-2\varepsilon_0(\bar{t} - \underline{t} + 1)} \sum_{m=1}^{n} \rho^{a_m \vee b_m \vee c_m} \ ev. \Bigg\} = 1, \tag{A.26}$$

where $\underline{t}$ and $\bar{t}$ are integers such that $\underline{t} \le t_{k_{i_1}}, t_{h_{i_2}}, t_{\ell_{i_3}} \le \bar{t}$, for all $1 \le i_1 \le a_m$, $1 \le i_2 \le b_m$, $1 \le i_3 \le c_m$ and all $1 \le m \le n$.

Proof. First, we show (A.25). We define ρ, I_t and ε_0 as in (A.4), (A.5) and (A.8), respectvely. Notice that $\varepsilon_0 \in (0, \frac{1}{16r})$. Using $1 - \omega(\mathbf{V}_t) \le \rho^{1 - I_t}$, it follows that

$$\prod_{i_1=1}^{a_m} (1 - \omega(\mathbf{V}_{t_{k_{i_1}}})) \le \rho^{a_m - \sum_{i_1=1}^{a_m} I_{t_{k_{i_1}}}} \le \rho^{a_m - \sum_{t=\underline{t}}^{\bar{t}} I_t},$$
$$\prod_{i_2=1}^{b_m} (1 - \omega(\mathbf{V}_{t_{\ell_{i_2}}})) \le \rho^{b_m - \sum_{i_2=1}^{b_m} I_{t_{\ell_{i_2}}}} \le \rho^{b_m - \sum_{t=\underline{t}}^{\bar{t}} I_t},$$

and

$$\sum_{m=1}^{n} \prod_{i_1=1}^{a_m} (1 - \omega(\mathbf{V}_{t_{k_{i_1}}})) \wedge \prod_{i_2=1}^{b_m} (1 - \omega(\mathbf{V}_{t_{\ell_{i_2}}})) \le \sum_{m=1}^{n} \rho^{a_m - \sum_{t=\underline{t}}^{\bar{t}} I_t} \wedge \rho^{b_m - \sum_{t=\underline{t}}^{\bar{t}} I_t} = \rho^{-\sum_{t=\underline{t}}^{\bar{t}} I_t} \sum_{m=1}^{n} \rho^{a_m \vee b_m}. \tag{A.27}$$

Since $\mathbf{V}_t$ forms a stationary and ergodic sequence for $\underline{t} \le t \le \bar{t}$, the strong law of large numbers yields $(\bar{t} - \underline{t} + 1)^{-1} (\sum_{t=\underline{t}}^{\bar{t}} I_t) \to \mathbb{E}_{\theta^*}[I_1] < \varepsilon_0$, $\mathbb{P}_{\theta^*} - a.s.$ Hence,

$$\mathbb{P}_{\theta^*}\left(\rho^{-\sum_{t=\underline{t}}^{\bar{t}} I_t} \le \rho^{-2\varepsilon_0(\bar{t} - \underline{t} + 1)} \ ev. \right) = 1. \tag{A.28}$$

The proof is completed by combining (A.27) and (A.28). (A.26) similarly follows. ∎

LEMMA 11. *Assume (A2)–(A4) and (A7)–(A8). A random variable* $K \in L^1(\mathbb{P}_{\theta^*})$ *exists such that, for all* $t \geq 1$ *and* $0 \leq m \leq m'$,

$$\mathbb{P}_{\theta^*}\left(\max_{\bar{\mathbf{s}}}\sup_{\theta\in G}\|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta)-\Gamma_{t,m',\bar{\mathbf{s}}}(\theta)\| \leq K(t\vee m)^2\rho^{(t+m)/4r}\ ev.\right)=1, \tag{A.29}$$

$$\mathbb{P}_{\theta^*}\left(\max_{\bar{\mathbf{s}}}\sup_{\theta\in G}\|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta)-\Gamma_{t,m}(\theta)\| \leq K(t\vee m)^2\rho^{(t+m)/4r}\ ev.\right)=1. \tag{A.30}$$

Proof. Put $\|\dot{\phi}_k\|_\infty = \max_{s_k,\bar{\mathbf{s}}_{k-1}}\sup_{\theta\in G}\|\dot{\phi}_\theta((s_k,Y_k),(\bar{\mathbf{s}}_{k-1},\overline{\mathbf{Y}}_{k-1}),X_k)\|$. For $-m' < -m < k \leq t$,

$$\begin{aligned}
&\left\|\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}},\mathbf{X}^t_{-m}\right]-\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m},\mathbf{X}^t_{-m}\right]\right\|\\
&\qquad\qquad\leq 2\|\dot{\phi}_k\|_\infty\prod_{i=1}^{\lfloor(k+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri})),\\
&\left\|\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}},\mathbf{X}^t_{-m}\right]-\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m'},\overline{\mathbf{S}}_{-m'}=\bar{\mathbf{s}},\mathbf{X}^t_{-m'}\right]\right\|\\
&\qquad\qquad\leq 2\|\dot{\phi}_k\|_\infty\prod_{i=1}^{\lfloor(k+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri})),\\
&\left\|\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}},\mathbf{X}^t_{-m}\right]-\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^{t-1}_{-m},\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}},\mathbf{X}^{t-1}_{-m}\right]\right\|\\
&\qquad\qquad\leq 2\|\dot{\phi}_k\|_\infty\prod_{i=1}^{\lfloor(t-1-k)/r\rfloor}(1-\omega(\mathbf{V}_{t+r-ri})),\\
&\left\|\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^t_{-m},\mathbf{X}^t_{-m}\right]-\mathbb{E}_\theta\left[\dot{\phi}_\theta(Z_k,\overline{\mathbf{Z}}_{k-1},X_k)|\overline{\mathbf{Y}}^{t-1}_{-m},\mathbf{X}^{t-1}_{-m}\right]\right\|\\
&\qquad\qquad\leq 2\|\dot{\phi}_k\|_\infty\prod_{i=1}^{\lfloor(t-1-k)/r\rfloor}(1-\omega(\mathbf{V}_{t+r-ri})).
\end{aligned}$$

First, we show (A.30).

$$\begin{aligned}
&\|\Gamma_{t,m,\bar{\mathbf{s}}}-\Gamma_{t,m}\| \leq 2\|\dot{\phi}_t\|_\infty\prod_{i=1}^{\lfloor(t+m-1)/r\rfloor}\big(1-\omega(\mathbf{V}_{-m+ri})\big)\\
&+4\sum_{k=-m+1}^{t-1}\|\dot{\phi}_k\|_\infty\Big[\prod_{i=1}^{\lfloor(k+m-1)/r\rfloor}\big(1-\omega(\mathbf{V}_{-m+ri})\big)\wedge\prod_{i=1}^{\lfloor(t-k-1)/r\rfloor}\big(1-\omega(\mathbf{V}_{t+r-ri})\big)\Big]\\
&\leq 4\max_{-m+1\leq k\leq t}\|\dot{\phi}_k\|_\infty\sum_{k=-m+1}^{t}\Big[\prod_{i=1}^{\lfloor(k+m-1)/r\rfloor}\big(1-\omega(\mathbf{V}_{-m+ri})\big)\wedge\prod_{i=1}^{\lfloor(t-k-1)/r\rfloor}\big(1-\omega(\mathbf{V}_{t+r-ri})\big)\Big].
\end{aligned} \tag{A.31}$$

The first part of (A.31) is bounded by

$$
\begin{aligned}
4 \max_{-m+1\le k\le t} \|\dot{\phi}_k\|_\infty &\le 4 \sum_{k=-m+1}^{t} (|k|\vee 1)^2 \frac{1}{(|k|\vee 1)^2} \|\dot{\phi}_k\|_\infty \\
&\le 4(t\vee m)^2 \sum_{k=-\infty}^{\infty} \frac{1}{(|k|\vee 1)^2} \|\dot{\phi}_k\|_\infty \qquad \text{(A.32)}
\end{aligned}
$$

We proceed to bound the second part of (A.31). Since $-m + r\lfloor (k+m-1)/r \rfloor < t + r - r\lfloor (t-k-1)/r \rfloor$, we can apply Lemma 10. Using $\rho^{\lfloor (k+m-1)/r\rfloor} \wedge \rho^{\lfloor (t-k-1)/r\rfloor} = \rho^{\lfloor (t-k-1)/r\rfloor}$ for $k \le \frac{t-m}{2}$ and $\rho^{\lfloor (k+m-1)/r\rfloor}$ for $k \ge \frac{t-m}{2}$,

$$
\begin{aligned}
&\rho^{-2\varepsilon_0(t+m+r-1)} \sum_{k=-m+1}^{t} \left(\rho^{\lfloor \frac{k+m-1}{r} \rfloor} \wedge \rho^{\lfloor \frac{t-k-1}{r}\rfloor}\right) \\
\le& \rho^{-2\varepsilon_0(t+m+r-1)} \left(\sum_{k\le (t-m)/2} \rho^{\lfloor \frac{t-k-1}{r}\rfloor} + \sum_{k\ge (t-m)/2} \rho^{\lfloor \frac{k+m-1}{r}\rfloor} \right) \\
\le& 2 \frac{\rho^{\frac{t+m}{4r}}}{\rho^{r-1}(1-\rho^{\frac{1}{r}})}. \qquad \text{(A.33)}
\end{aligned}
$$

Because of (A.32), (A.33), and Lemma 10, (A.30) follows.

Next, we follow a similar procedure to show (A.29).

$$
\begin{aligned}
&\|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta) - \Gamma_{t,m',\bar{\mathbf{s}}}(\theta)\| \le 2\|\dot{\phi}_t\|_\infty \prod_{i=1}^{\lfloor (t+m-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{-m+ri})\big) \\
&+ 4 \sum_{k=-m+1}^{t-1} \|\dot{\phi}_k\|_\infty \Big[\prod_{i=1}^{\lfloor (k+m-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{-m+ri})\big) \wedge \prod_{i=1}^{\lfloor (t-k-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{t+r-ri})\big) \Big] \\
&+ 2 \sum_{k=-m'+1}^{-m} \|\dot{\phi}_k\|_\infty \prod_{i=1}^{\lfloor (t-k-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{t+r-ri})\big) \qquad \text{(A.34)}
\end{aligned}
$$

The first two terms on the right-hand side can be bounded as above. We proceed to show the bound for the third term. Define ρ, I_t and ε_0 as in (A.4), (A.5) and (A.8), respectvely. Notice that the ρ and ε_0 are the same as the ones in Lemma 2. Using $1 - \mathbf{V}_t \le \rho^{1-I_t}$,

$$
\sum_{k=-m'+1}^{-m} \|\dot{\phi}_k\|_\infty \prod_{i=1}^{\lfloor (t-k-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{t+r-ri})\big)
$$

$$
\begin{aligned}
&\le \sum_{k=-m'+1}^{-m} \|\dot\phi_k\|_\infty \rho^{\lfloor (t-k-1)/r\rfloor - \sum_{i=1}^{\lfloor (t-k-1)/r\rfloor} I_{t+r-ri}} \\
&\le \sum_{k=-m'+1}^{-m} \|\dot\phi_k\|_\infty \rho^{\lfloor (t-k-1)/r\rfloor - \sum_{i=-m'+r+1}^{t} I_i} \\
&\le \rho^{-\sum_{i=-m'+r+1}^{t} I_i} \sum_{k=-m'+1}^{-m} \|\dot\phi_k\|_\infty \rho^{\lfloor (t-k-1)/r\rfloor}.
\end{aligned}
$$

Using (A.6) and (A.7), $\mathbb{E}_{\theta^*}[\rho^{-\sum_{i=-m'+r+1}^{t} I_i}] \le 2$ for all m' and t. Then $\rho^{-\sum_{i=-m'+r+1}^{t} I_i} \le \rho^{-\sum_{i=-\infty}^{+\infty} I_i}$, which is in $L^1(\mathbb{P}_{\theta^*})$.

Since $-\frac{k}{2r} \le \frac{t-m-2k}{2r}$ for $k \le m$, we can follow a similar procedure to Douc et al. (2004, p. 2295) to obtain

$$
\begin{aligned}
&2 \sum_{k=-m'+1}^{-m} \|\dot\phi_k\|_\infty \rho^{\lfloor (t-k-1)/r\rfloor} \le 2\rho^{\frac{t+m}{2r}} \sum_{k=-m'+1}^{-m} \|\dot\phi_k\|_\infty \rho^{\frac{t-m-2k}{2r}-1} \\
&\le 2\rho^{\frac{t+m}{2r}} \sum_{k=-\infty}^{\infty} \|\dot\phi_k\|_\infty \rho^{\frac{|k|}{2r}-1} \le 2\rho^{\frac{t+m}{4r}} \sum_{k=-\infty}^{\infty} \|\dot\phi_k\|_\infty \rho^{\frac{|k|}{2r}-1}.
\end{aligned}
$$

Then (A.29) follows.

■

LEMMA 12. *Assume (A2)–(A4) and (A7)–(A8). Then, for all $t \ge 1$ and $0 \le m \le m'$,*

$$\lim_{m\to\infty} \mathbb{E}_{\theta^*}[\max_{\bar{\mathbf{s}}} \sup_{\theta\in G} \|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta) - \Gamma_{t,m',\bar{\mathbf{s}}}(\theta)\|] = 0, \tag{A.35}$$

$$\lim_{m\to\infty} \mathbb{E}_{\theta^*}[\max_{\bar{\mathbf{s}}} \sup_{\theta\in G} \|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta) - \Gamma_{t,m}(\theta)\|] = 0. \tag{A.36}$$

Proof. (A.35) follows from (A.34) and

$$
\begin{aligned}
&\mathbb{E}_{\theta^*}\big[\sup_{\theta\in G} \max_{\bar{\mathbf{s}}\in\bar{\mathbb{S}}} \|\Gamma_{t,m,\bar{\mathbf{s}}}(\theta) - \Gamma_{t,m',\bar{\mathbf{s}}}(\theta)\|\big] \\
&\le 2\big(\mathbb{E}_{\theta^*}\|\dot\phi_0\|_\infty^2\big)^{\frac12} \Bigg\{ \Bigg[\mathbb{E}_\theta\Big[\prod_{i=1}^{\lfloor (t+m-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{-m+ri})\big)^2\Big]\Bigg]^{\frac12} \\
&+ 2\sum_{k=-m+1}^{t-1} \Bigg[\mathbb{E}_\theta\Big[\prod_{i=1}^{\lfloor (k+m-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{-m+ri})\big)^2 \wedge \prod_{i=1}^{\lfloor (t-k-1)/r\rfloor} \big(1-\omega(\mathbf{V}_{t+r-ri})\big)^2\Big]\Bigg]^{\frac12}
\end{aligned}
$$

$$
+ \sum_{k=-m'+1}^{-m} \left[\mathbb{E}_\theta \Big[\prod_{i=1}^{\lfloor (t-k-1)/r \rfloor} \big(1-\omega(\mathbf{V}_{t+r-ri})\big)^2 \Big] \right]^{\frac{1}{2}} \Bigg\}
$$

$$
\leq 8\big(\mathbb{E}_\theta \|\dot{\phi}_0\|_\infty^2\big)^{\frac{1}{2}} \frac{\rho^{\frac{t+m-2}{4r}-1}}{1-\rho^{\frac{1}{2r}}},
$$

which converges to 0. Similarly, we can obtain (A.36) by using (A.31). ■

LEMMA 13. *Assume (A2)–(A5) and (A7)–(A8), for all* $\overline{\mathbf{s}} \in \overline{\mathbb{S}}$ *and* $m \geq 0$*, the function* $\theta \mapsto \Gamma_{0,m,\overline{\mathbf{s}}}$ *is* $\mathbb{P}_{\theta^*} - a.s.$ *continuous on* G*. In addition, for all* $\theta \in G$ *and for all* $\overline{\mathbf{s}} \in \overline{\mathbb{S}}$*,*

$$
\lim_{\delta \mapsto 0} \mathbb{E}_{\theta^*} \Big[\sup_{|\theta'-\theta| \leq \delta} |\Gamma_{0,m,\overline{\mathbf{s}}}(\theta') - \Gamma_{0,m,\overline{\mathbf{s}}}(\theta)| \Big] = 0.
$$

Proof. Note that $|\Gamma_{0,m,\overline{\mathbf{s}}}(\theta)| \leq 2\sum_{k=-m+1}^{0} \|\dot{\phi}_k\|_\infty$. By Assumption (A8), $\Gamma_{0,m,\overline{\mathbf{s}}}(\theta)$ is uniformly bounded with respect to θ by a random variable in $L^1(\mathbb{P}_{\theta^*})$. It hence suffices to show that for $-m < k \leq 0$,

$$
\lim_{\delta \to 0} \sup_{|\theta'-\theta| \leq \delta} \Big| \mathbb{E}_{\theta'}[\dot{\phi}_{\theta'}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^0]
$$

$$
- \mathbb{E}_{\theta}[\dot{\phi}_{\theta}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^0] \Big| = 0, \quad \mathbb{P}_{\theta^*} - a.s.
$$

Write

$$
\mathbb{E}_{\theta}[\dot{\phi}_{\theta}(Z_k, \overline{\mathbf{Z}}_{k-1}, X_k) | \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^0]
$$

$$
= \sum_{s_k, \overline{\mathbf{s}}_{k-1}} \dot{\phi}_\theta\big((Y_k, s_k), (\overline{\mathbf{Y}}_{k-1}, \overline{\mathbf{s}}_{k-1}), X_k\big) \mathbb{P}_\theta(S_k = s_k, \overline{\mathbf{S}}_{k-1} = \overline{\mathbf{s}}_{k-1} | \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^0).
$$

Note that $\dot{\phi}_\theta\big((Y_k, s_k), (\overline{\mathbf{Y}}_{k-1}, \overline{\mathbf{s}}_{k-1}), X_k\big)$ is continuous with respect to θ by Assumption (A7). Similar to the proof of Proposition 1, we can show $\mathbb{P}_\theta(S_k = s_k, \overline{\mathbf{S}}_{k-1} = \overline{\mathbf{s}}_{k-1} | \overline{\mathbf{Y}}_{-m}^0, \overline{\mathbf{S}}_{-m} = \overline{\mathbf{s}}, \mathbf{X}_{-m}^0)$ is continuous with respect to θ from Assumption (A5). The proof is complete. ■

By Lemmas 11 and 12, $\{\Gamma_{0,m,\overline{\mathbf{s}}}(\theta)\}_{m\geq 0}$ is a uniform Cauchy sequence with respect to θ $\mathbb{P}_{\theta^*}$-a.s. and in $L^1(\mathbb{P}_{\theta^*})$. Lemma (13) shows that $\Gamma_{0,m,\overline{\mathbf{s}}}(\theta)$ is continuous with respect to θ on G $\mathbb{P}_{\theta^*} - a.s.$ and in $L^1(\mathbb{P}_{\theta^*})$ for each m. Hence, $\Gamma_{0,\infty}(\theta)$ is continuous with respect to θ on G $\mathbb{P}_{\theta^*}$-a.s. and in $L^1(\mathbb{P}_{\theta^*})$.

Proof of Proposition 3. Write

$$\begin{aligned}
&\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\Gamma_{t,0,\bar{\mathbf{s}}}(\theta')-\mathbb{E}_{\theta^*}[\Gamma_{0,\infty}(\theta)]\right|\\
\le&\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\frac{1}{n}\sum_{t=1}^{n}\left|\Gamma_{t,0,\bar{\mathbf{s}}}(\theta')-\Gamma_{t,\infty}(\theta')\right|\\
&+\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\left|\frac{1}{n}\sum_{t=1}^{n}\Gamma_{t,\infty}(\theta')-\mathbb{E}_{\theta^*}[\Gamma_{0,\infty}(\theta')]\right|\\
&+\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{\theta':|\theta'-\theta|<\delta}\mathbb{E}_{\theta^*}\left|\Gamma_{0,\infty}(\theta')-\Gamma_{0,\infty}(\theta)\right|.
\end{aligned}$$

The first term on the right-hand side is zero $\mathbb{P}_{\theta^*}-a.s.$ by Lemma 11. The second term is zero $\mathbb{P}_{\theta^*}-a.s.$, owing to the ergodic theorem. The third term is zero from the continuity of $\Gamma_{0,\infty}(\theta)$ in $L^1(\mathbb{P}_{\theta^*})$. ■

LEMMA 14. *Assume (A2)–(A4) and (A7)–(A8). Then, a random variable $K\in L^1(\mathbb{P}_{\theta^*})$ exists such that, for all $t\ge 1$ and $0\le m\le m'$,*

$$\mathbb{P}_{\theta^*}\left(\max_{\bar{\mathbf{s}}}\sup_{\theta\in G}\|\Phi_{t,m,\bar{\mathbf{s}}}(\theta)-\Phi_{t,m}(\theta)\|\le K(t\vee m)^3\rho^{(t+m)/8r}\ ev.\right)=1,$$
$$\mathbb{P}_{\theta^*}\left(\max_{\bar{\mathbf{s}}}\sup_{\theta\in G}\|\Phi_{t,m,\bar{\mathbf{s}}}(\theta)-\Phi_{t,m',\bar{\mathbf{s}}}(\theta)\|\le K(t\vee m)^3\rho^{(t+m)/8r}\ ev.\right)=1.$$

Proof. Put $\|\phi_k\|_\infty=\max_{s_k,\bar{\mathbf{s}}_{k-1}}\sup_{\theta\in G}\|\phi_\theta((s_k,Y_k),(\bar{\mathbf{s}}_{k-1},\overline{\mathbf{Y}}_{k-1}),X_k)\|$. For $m'\ge m\ge 0$, all $-m<\ell\le k\le n$, all $\theta\in G$, and all $\bar{\mathbf{s}}_{-m}\in\overline{\mathbb{S}}$,

$$\|\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^n,\mathbf{X}_{-m}^n]\|\le 2\|\phi_k\|_\infty\|\phi_\ell\|_\infty\prod_{i=1}^{\lfloor(k-\ell-1)/r\rfloor}(1-\omega(\mathbf{V}_{\ell+ri})),$$

$$\|\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^n,\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m}^n]\|\le 2\|\phi_k\|_\infty\|\phi_\ell\|_\infty\prod_{i=1}^{\lfloor(k-\ell-1)/r\rfloor}(1-\omega(\mathbf{V}_{\ell+ri})),$$

$$\begin{aligned}
&\|\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^n,\overline{\mathbf{S}}_{-m}=\bar{\mathbf{s}}_{-m},\mathbf{X}_{-m}^n]-\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^n,\mathbf{X}_{-m}^n]\|\\
&\qquad\le 6\|\phi_k\|_\infty\|\phi_\ell\|_\infty\prod_{i=1}^{\lfloor(\ell+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri})),
\end{aligned}$$

$$\begin{aligned}
&\|\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^n,\mathbf{X}_{-m}^n]-\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{X}_{-m}^{n-1}]\|\\
&\qquad\le 6\|\phi_k\|_\infty\|\phi_\ell\|_\infty\prod_{i=1}^{\lfloor(n-k-1)/r\rfloor}(1-\omega(\mathbf{V}_{n+r-ri})),
\end{aligned}$$

$$\|\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}^n_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_{-m},\mathbf{X}^n_{-m}]-\mathrm{Cov}_\theta[\phi_{\theta,k},\phi_{\theta,\ell}|\overline{\mathbf{Y}}^{n-1}_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}}_{-m},\mathbf{X}^{n-1}_{-m}]\| \\ \leq 6\|\phi_k\|_\infty\|\phi_\ell\|_\infty \prod_{i=1}^{\lfloor (n-k-1)/r\rfloor}(1-\omega(\mathbf{V}_{n+r-ri})).$$

We define $\Lambda_a^b=\sum_{i=a}^b \phi_{\theta,i}$. Then, $\Phi_{t,m,\overline{\mathbf{s}}}(\theta)-\Phi_{t,m',\overline{\mathbf{s}}}(\theta)$ may be decomposed as $A+2B+C$, where

$$\begin{aligned}
A =&\mathrm{Var}_\theta[\Lambda^{t-1}_{-m+1}|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}},\mathbf{X}^t_{-m}]-\mathrm{Var}_\theta[\Lambda^{t-1}_{-m+1}|\overline{\mathbf{Y}}^{t-1}_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}},\mathbf{X}^{t-1}_{-m}]\\
&-\mathrm{Var}_\theta[\Lambda^{t-1}_{-m+1}|\overline{\mathbf{Y}}^t_{-m},\mathbf{X}^t_{-m}]+\mathrm{Var}_\theta[\Lambda^{t-1}_{-m+1}|\overline{\mathbf{Y}}^{t-1}_{-m},\mathbf{X}^{t-1}_{-m}],\\
B =&\mathrm{Cov}_\theta[\Lambda^{t-1}_{-m+1},\phi_{\theta,t}|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}},\mathbf{X}^t_{-m}]-\mathrm{Cov}_\theta[\Lambda^{t-1}_{-m+1},\phi_{\theta,t}|\overline{\mathbf{Y}}^t_{-m},\mathbf{X}^t_{-m}],\\
C =&\mathrm{Var}_\theta[\phi_{\theta,t}|\overline{\mathbf{Y}}^t_{-m},\overline{\mathbf{S}}_{-m}=\overline{\mathbf{s}},\mathbf{X}^t_{-m}]-\mathrm{Var}_\theta[\phi_{\theta,t}|\overline{\mathbf{Y}}^t_{-m},\mathbf{X}^t_{-m}].
\end{aligned}$$

We have

$$\begin{aligned}
\|A\| \leq& \max_{-m+1\leq \ell\leq k\leq t-1}\|\phi_k\|_\infty\|\phi_\ell\|_\infty \times 2\sum_{-m+1\leq\ell\leq k\leq t-1}\Bigg\{\Bigg[2\times 6\prod_{i=1}^{\lfloor(\ell+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri}))\Bigg]\\
&\wedge\Bigg[4\times 2\prod_{i=1}^{\lfloor(k-\ell-1)/r\rfloor}(1-\omega(\mathbf{V}_{\ell+ri}))\Bigg]\wedge\Bigg[2\times 6\prod_{i=1}^{\lfloor(t-k-1)/r\rfloor}(1-\omega(\mathbf{V}_{t+r-ri}))\Bigg]\Bigg\},\\
\|B\| \leq& \max_{-m+1\leq k\leq t-1}\|\phi_k\|_\infty\|\phi_t\|_\infty\times\sum_{-m+1\leq k\leq t-1}\Bigg\{\Bigg[6\prod_{i=1}^{\lfloor(k+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri}))\Bigg]\\
&\wedge\Bigg[2\times 2\prod_{i=1}^{\lfloor(t-k-1)/r\rfloor}(1-\omega(\mathbf{V}_{k+ri}))\Bigg]\Bigg\},\\
\|C\| \leq& \|\phi_t\|^2_\infty\times 6\prod_{i=1}^{\lfloor(t+m-1)/r\rfloor}(1-\omega(\mathbf{V}_{-m+ri})).
\end{aligned}$$

Similar to the calculation on p. 2299 of Douc et al. (2004), we derive

$$\max_{-m+1\leq\ell\leq k\leq t-1}\|\phi_k\|_\infty\|\phi_\ell\|_\infty\leq(m^3+t^3)\sum_{k=-\infty}^{\infty}\frac{1}{(|k|\vee 1)^2}\|\phi_k\|^2_\infty.$$

In view of Lemma 10 and

$$\rho^{-2\varepsilon_0(t+m-r+1)}\sum_{-m+1\leq\ell\leq k\leq t-1}\left(\rho^{\lfloor(\ell+m-1)/r\rfloor}\wedge\rho^{\lfloor(k-\ell-1)/r\rfloor}\wedge\rho^{\lfloor(t-k-1)/r\rfloor}\right)$$

$$\leq 4 \frac{1}{\rho(1-\rho)(1-\rho^{\frac{1}{2}})} \rho^{\frac{t+m-4}{8r}},$$

$$\mathbb{P}_{\theta^*}\Big(\|A\| \leq (m^3+t^3) \sum_{k=-\infty}^{\infty} \frac{1}{(|k| \vee 1)^2} \|\phi_k\|_\infty^2 \frac{96\rho^{\frac{t+m}{8r}-1}}{\rho^{\frac{1}{r}}(1-\rho^{\frac{1}{r}})(1-\rho^{\frac{1}{2r}})} \ ev. \Big) = 1,$$

$$\mathbb{P}_{\theta^*}\Big(\|B\| \leq (m^3+t^3) \sum_{k=-\infty}^{\infty} \frac{1}{(|k| \vee 1)^2} \|\phi_k\|_\infty^2 \frac{12\rho^{\frac{t+m}{4r}-1}}{\rho^{\frac{1}{r}}(1-\rho^{\frac{1}{r}})} \ ev. \Big) = 1,$$

$$\mathbb{P}_{\theta^*}\Big(\|C\| \leq (m^3+t^3) \sum_{k=-\infty}^{\infty} \frac{1}{(|k| \vee 1)^2} \|\phi_k\|_\infty^2 6\rho^{\lfloor (t+m-1)/4r \rfloor} \ ev. \Big) = 1.$$

The difference $\Phi_{t,m,\bar{\mathbf{s}}}(\theta) - \Phi_{t,m',\bar{\mathbf{s}}}(\theta)$ can be decomposed as $A + 2B + C - D - 2E - 2F$, where

$$\begin{aligned}
A =& \mathrm{Var}_\theta[\Lambda_{-m+1}^{t-1} | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}_{-m}^t] - \mathrm{Var}_\theta[\Lambda_{-m+1}^{t-1} | \overline{\mathbf{Y}}_{-m}^{t-1}, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}_{-m}^{t-1}] \\
& - \mathrm{Var}_\theta[\Lambda_{-m+1}^{t-1} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t] + \mathrm{Var}_\theta[\Lambda_{-m+1}^{t-1} | \overline{\mathbf{Y}}_{-m'}^{t-1}, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^{t-1}], \\
B =& \mathrm{Cov}_\theta[\Lambda_{-m+1}^{t-1}, \phi_{\theta,t} | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}_{-m}^t] \\
& - \mathrm{Cov}_\theta[\Lambda_{-m+1}^{t-1}, \phi_{\theta,t} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t], \\
C =& \mathrm{Var}_\theta[\phi_{\theta,t} | \overline{\mathbf{Y}}_{-m}^t, \overline{\mathbf{S}}_{-m} = \bar{\mathbf{s}}, \mathbf{X}_{-m}^t] - \mathrm{Var}_\theta[\phi_{\theta,t} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t], \\
D =& \mathrm{Var}_\theta[\Lambda_{-m'+1}^{-m} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t] - \mathrm{Var}_\theta[\Lambda_{-m'+1}^{-m} | \overline{\mathbf{Y}}_{-m'}^{t-1}, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^{t-1}], \\
E =& \mathrm{Cov}_\theta[\Lambda_{-m+1}^{t-1}, \Lambda_{-m'+1}^{-m} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t] \\
& - \mathrm{Cov}_\theta[\Lambda_{-m+1}^{t-1}, \Lambda_{-m'+1}^{-m} | \overline{\mathbf{Y}}_{-m'}^{t-1}, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^{t-1}], \\
F =& \mathrm{Cov}_\theta[\Lambda_{-m'+1}^{-m}, \phi_{\theta,t} | \overline{\mathbf{Y}}_{-m'}^t, \overline{\mathbf{S}}_{-m'} = \bar{\mathbf{s}}, \mathbf{X}_{-m'}^t].
\end{aligned}$$

$\|A\|, \|B\|$, and $\|C\|$ are bounded as above. For the other terms, we have

$$\begin{aligned}
\|D\| \leq 2 \sum_{-m'+1 \leq \ell \leq k \leq -m} \Big[& 6 \prod_{i=1}^{\lfloor (t-k-1)/r \rfloor} (1 - \omega(\mathbf{V}_{t+r-ri})) \\
& \wedge 2 \times 2 \prod_{i=1}^{\lfloor (k-\ell-1)/r \rfloor} (1 - \omega(\mathbf{V}_{\ell+ri})) \Big] \|\phi_k\|_\infty \|\phi_\ell\|_\infty, \\
\|E\| \leq \sum_{k=-m+1}^{t-1} \sum_{\ell=-m'+1}^{-m} \Big[& 6 \prod_{i=1}^{\lfloor (t-k-1)/r \rfloor} (1 - \omega(\mathbf{V}_{t+r-ri}))
\end{aligned}$$

$$\wedge 2 \times 2 \prod_{i=1}^{\lfloor (k-\ell-1)/r \rfloor} (1-\omega(\mathbf{V}_{\ell+ri}))\Big] \|\phi_k\|_\infty \|\phi_\ell\|_\infty,$$

$$\|F\| \leq \sum_{-m'+1 \leq \ell \leq -m} 2 \prod_{i=1}^{\lfloor (t-\ell-1)/r \rfloor} (1-\omega(\mathbf{V}_{\ell+ri})) \|\phi_\ell\|_\infty \|\phi_t\|_\infty.$$

Define ρ, I_t and ε_0 as in (A.4), (A.5) and (A.8), respectvely. Notice that the ρ and ε_0 are the same as the ones in Lemma 2. Using $1 - \mathbf{V}_t \leq \rho^{1-I_t}$,

$$\begin{aligned}
&\sum_{-m'+1 \leq \ell \leq k \leq -m} \Big[6 \prod_{i=1}^{\lfloor (t-k-1)/r \rfloor} (1-\omega(\mathbf{V}_{-m+ri})) \wedge 2 \times 2 \prod_{i=1}^{\lfloor (k-\ell-1)/r \rfloor} (1-\omega(\mathbf{V}_{\ell+ri}))\Big] \|\phi_k\|_\infty \|\phi_\ell\|_\infty \\
&\leq 12 \sum_{-m'+1 \leq \ell \leq k \leq -m} \big(\rho^{\lfloor (t-k-1)/r \rfloor - \sum_{i=1}^{\lfloor (t-k-1)/r \rfloor} I_{-m+ri}} \wedge \rho^{\lfloor (k-\ell-1)/r \rfloor - \sum_{i=1}^{\lfloor (k-\ell-1)/r \rfloor} I_{\ell+ri}}\big) \|\phi_k\|_\infty \|\phi_\ell\|_\infty \\
&\leq 12 \rho^{-\sum_{i=-m'+1}^{t} I_i} \sum_{-m'+1 \leq \ell \leq k \leq -m} \big(\rho^{\lfloor (t-k-1)/r \rfloor} \wedge \rho^{\lfloor (k-\ell-1)/r \rfloor}\big) \|\phi_k\|_\infty \|\phi_\ell\|_\infty
\end{aligned}$$

Using (A.6) and (A.7), $\mathbb{E}_{\theta^*}[\rho^{-\sum_{i=-m'}^{t} I_i}] \leq 2$ for all m' and t. Then $\rho^{-\sum_{i=-m'}^{t} I_i} \leq \rho^{-\sum_{i=-\infty}^{+\infty} I_i}$, which is in $L^1(\mathbb{P}_{\theta^*})$. Following a similar procedure to Douc et al. (2004, pp. 2301), we obtain

$$\begin{aligned}
&\sum_{-m'+1 \leq \ell \leq k \leq -m} \rho^{\lfloor (t-k-1)/r \rfloor} \wedge \rho^{\lfloor (k-\ell-1)/r \rfloor} \|\phi_k\|_\infty \|\phi_\ell\|_\infty \\
&\qquad \leq \rho^{\frac{t+m-2}{8r} - 1} \sum_{\ell=-\infty}^{\infty} \rho^{\frac{|\ell|}{8r}} \|\phi_\ell\|_\infty \sum_{k=-\infty}^{\infty} \rho^{\frac{|k|}{4r}} \|\phi_k\|_\infty
\end{aligned}$$

Thus,

$$\|D\| \leq 12 \rho^{\frac{t+m-2}{8r} - 1} \rho^{-\sum_{i=-\infty}^{+\infty} I_i} \sum_{\ell=-\infty}^{\infty} \rho^{\frac{|\ell|}{8r}} \|\phi_\ell\|_\infty \sum_{k=-\infty}^{\infty} \rho^{\frac{|k|}{4r}} \|\phi_k\|_\infty.$$

Similarly, we can derive

$$\|E\| \leq 6 \rho^{\frac{t+m-2}{8r} - 1} \rho^{-\sum_{i=-\infty}^{+\infty} I_i} \sum_{\ell=-\infty}^{\infty} \rho^{\frac{|\ell|}{8r}} \|\phi_\ell\|_\infty \sum_{k=-\infty}^{\infty} \rho^{\frac{|k|}{4r}} \|\phi_k\|_\infty,$$

$$\|F\| \leq 2 \rho^{\frac{t+m}{4r} - 1} \rho^{-\sum_{i=-\infty}^{+\infty} I_i} \sum_{\ell=-\infty}^{\infty} \rho^{\frac{3|\ell|}{4r}} \|\phi_\ell\|_\infty \sum_{k=-\infty}^{\infty} \rho^{\frac{|k|}{2r}} \|\phi_k\|_\infty.$$

The proof is complete. ■

Using a similar procedure to that in Lemma 12 and the inequalities in Lemma 14, we obtain the following lemma.

LEMMA 15. *Assume (A2)–(A4) and (A7)–(A8). Then, for all $t \geq 1$ and $0 \leq m \leq m'$,*

$$\lim_{m\to\infty} \mathbb{E}_{\theta^*}[\max_{\bar{\mathbf{s}}} \sup_{\theta\in G} \|\Phi_{t,m,\bar{\mathbf{s}}}(\theta) - \Phi_{t,m',\bar{\mathbf{s}}}(\theta)\|] = 0,$$

$$\lim_{m\to\infty} \mathbb{E}_{\theta^*}[\max_{\bar{\mathbf{s}}} \sup_{\theta\in G} \|\Phi_{t,m,\bar{\mathbf{s}}}(\theta) - \Phi_{t,m}(\theta)\|] = 0.$$

By Lemmas 14 and 15, $\{\Phi_{t,m,\bar{\mathbf{s}}}(\theta)\}_{m\geq 0}$ is a uniform Cauchy sequence w.r.t. θ $\mathbb{P}_{\theta^*} - a.s.$ and in $L^1(\mathbb{P}_{\theta^*})$. The rest of the proof of Proposition 4 follows along the same line as the proof of Proposition 3.

A.5 Supplemental proof of Proposition 6

Show that for all $s' \in \mathbb{S}$ and $\bar{\mathbf{s}} \in \overline{\mathbb{S}}$, $\theta \in \Theta$, $\mathbb{E}_{\theta^*}|\log q_\theta(s'|\bar{\mathbf{s}}, \overline{\mathbf{Y}}_0, X_1)|^2 < \infty$.

First, consider the case of $s_1 = 0$ and $s_0 = 0$,

$$q_\theta(s_1 = 0|s_0 = 0, s_{-1}, \ldots, s_{-r+1}, \overline{\mathbf{Y}}_0, X_1) = \frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi\left(\frac{\tau-\rho U_{t-1}}{\sqrt{1-\rho^2}} - \frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx}{\Phi(\tau\sqrt{1-\alpha^2})}.$$

The following result holds for the numerator:

$$\begin{aligned}
&\int_{-\infty}^{\tau\sqrt{1-\alpha^2}} \Phi\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} - \frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx \\
\geq &\int_{-\infty}^{-|\tau\sqrt{1-\alpha^2}|} \Phi\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} + \frac{|\alpha| x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx \\
\geq &\int_{-|\tau\sqrt{1-\alpha^2}|-1}^{-|\tau\sqrt{1-\alpha^2}|} \Phi\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} + \frac{|\alpha| x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx \\
\geq &\Phi\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} - \frac{|\alpha\tau\sqrt{1-\alpha^2}| + |\alpha|}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}| - |\alpha|)\right) \\
\geq &\Phi(-D)\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}| - 1)\right) \\
= &\Phi^c(D)\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}| - 1)\right) \\
\geq &\sqrt{\frac{2}{\pi}} e^{-\frac{D^2}{2}} \frac{1}{D+\sqrt{D^2+1}}\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}| - 1)\right)
\end{aligned}$$

where $D \triangleq \left|\frac{\tau-\rho U_{t-1}}{\sqrt{1-\rho^2}} - \frac{|\alpha\tau\sqrt{1-\alpha^2}|+|\alpha|}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right|$ and $\Phi^c(x) \triangleq 1-\Phi(x)$. The last inequality follows from $\Phi^c(x) \geq \sqrt{\frac{2}{\pi}} e^{-x^2/2}\frac{1}{x+\sqrt{x^2+4}}$.

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\left[|\log q_\theta(s_1=0|s_0=0,s_{-1},\ldots,s_{-r+1},\overline{\mathbf{Y}}_0,X_1)|^2\right]\\
\leq&\mathbb{E}_{\theta^*}\Bigg[\Bigg|\log\left(e^{-\frac{D^2}{2}}\frac{1}{D+\sqrt{D^2+1}}\right)\\
&+\log\sqrt{\frac{2}{\pi}}+\log\left(\Phi(-|\tau\sqrt{1-\alpha^2}|)-\Phi(-|\tau\sqrt{1-\alpha^2}|-1)\right)-\log(\Phi(\tau\sqrt{1-\alpha^2}))\Bigg|^2\Bigg].
\end{aligned}$$

We need show only that $\mathbb{E}_{\theta^*}\left[\left(\log\left(e^{-\frac{D^2}{2}}\frac{1}{D+\sqrt{D^2+1}}\right)\right)^2\right] < \infty$. Note that D is folded normal distributed and has finite moments. We use $\log(x) \leq x-1$ for $x>0$ to obtain

$$\begin{aligned}
&\mathbb{E}_{\theta^*}\left[\left(\log\left(e^{-\frac{D^2}{2}}\frac{1}{D+\sqrt{D^2+1}}\right)\right)^2\right]=\mathbb{E}_{\theta^*}\left[\left(-\frac{D^2}{2}-\log(D+\sqrt{D^2+4})\right)^2\right]\\
=&\mathbb{E}_{\theta^*}\left[\frac{D^4}{4}+\left(\log(D+\sqrt{D^2+4})\right)^2+D^2\log(D+\sqrt{D^2+4})\right]\\
\leq&\mathbb{E}_{\theta^*}\left[\frac{D^4}{4}+(D+\sqrt{D^2+4}-1)^2+D^2(D+\sqrt{D^2+4}-1)\right]\\
\leq&\mathbb{E}_{\theta^*}\left[\frac{D^4}{4}+(2D+1)^2+D^2(2D+1)^2\right]<\infty.
\end{aligned} \tag{A.37}$$

Next, consider the case of $s_1=1$ and $s_0=0$,

$$q_\theta(s_1=1|s_0=0,s_{-1},\ldots,s_{-r+1},\overline{\mathbf{Y}}_0,X_1)=\frac{\int_{-\infty}^{\tau\sqrt{1-\alpha^2}}\Phi^c\left(\frac{\tau-\rho U_{t-1}}{\sqrt{1-\rho^2}}-\frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx}{\Phi(\tau\sqrt{1-\alpha^2})}.$$

The following result holds for the numerator

$$\begin{aligned}
&\int_{-\infty}^{\tau\sqrt{1-\alpha^2}}\Phi^c\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}}-\frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx\\
\geq&\int_{-\infty}^{-|\tau\sqrt{1-\alpha^2}|}\Phi^c\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}}-\frac{|\alpha|x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx\\
\geq&\int_{-|\tau\sqrt{1-\alpha^2}|-1}^{-|\tau\sqrt{1-\alpha^2}|}\Phi^c\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}}-\frac{|\alpha|x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx\\
\geq&\Phi^c\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}}+\frac{|\alpha\tau\sqrt{1-\alpha^2}|+|\alpha|}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\left(\Phi(-|\tau\sqrt{1-\alpha^2}|)-\Phi(-|\tau\sqrt{1-\alpha^2}|-1)\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \Phi^c(D_2)\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}|-1)\right)\\
&\geq \sqrt{\frac{2}{\pi}} e^{-\frac{D_2^2}{2}} \frac{1}{D_2+\sqrt{D_2^2+1}}\left(\Phi(-|\tau\sqrt{1-\alpha^2}|) - \Phi(-|\tau\sqrt{1-\alpha^2}|-1)\right)
\end{aligned}
$$

where $D_2 \triangleq \left|\frac{\tau-\rho U_{t-1}}{\sqrt{1-\rho^2}} + \frac{|\alpha\tau\sqrt{1-\alpha^2}|+|\alpha|}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right|$. Then, we can show $\mathbb{E}_{\theta^*}|\log q_\theta(s_1=1|s_0=0,s_{-1},\ldots,s_{-r+1},\overline{\mathbf{Y}}_0,X_1)|^2 < \infty$ by showing that $\mathbb{E}_{\theta^*}\left[\left(\log\left(e^{-\frac{D_2^2}{2}}\frac{1}{D_2+\sqrt{D_2^2+1}}\right)\right)^2\right] < \infty$. The proof follows a similar procedure to that in (A.37).

For the case of $s_1=0, s_0=1$,

$$
\begin{aligned}
&\int_{\tau\sqrt{1-\alpha^2}}^{-\infty} \Phi\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} - \frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx\\
\geq &\sqrt{\frac{2}{\pi}} e^{-\frac{D^2}{2}} \frac{1}{D+\sqrt{D^2+1}}\left(\Phi(|\tau\sqrt{1-\alpha^2}+1|) - \Phi(|\tau\sqrt{1-\alpha^2}|)\right)
\end{aligned}
$$

and for the case of $s_1=1$, $s_0=1$,

$$
\begin{aligned}
&\int_{\tau\sqrt{1-\alpha^2}}^{-\infty} \Phi^c\left(\frac{\tau-\rho U_0}{\sqrt{1-\rho^2}} - \frac{\alpha x}{\sqrt{1-\alpha^2}\sqrt{1-\rho^2}}\right)\varphi(x)dx\\
\geq &\sqrt{\frac{2}{\pi}} e^{-\frac{D_2^2}{2}} \frac{1}{D_2+\sqrt{D_2^2+1}}\left(\Phi(|\tau\sqrt{1-\alpha^2}+1|) - \Phi(|\tau\sqrt{1-\alpha^2}|)\right)
\end{aligned}
$$

and the result follows.