# Asymptotic Properties of the Maximum Likelihood Estimator in Regime Switching Econometric Models

By

Hiroyuki Kasahara and Katsumi Shimotsu

April 2018

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION
DISCUSSION PAPER NO. 35

CENTER FOR RESEARCH AND EDUCATION FOR POLICY EVALUATION (CREPE)
THE UNIVERSITY OF TOKYO
http://www.crepe.e.u-tokyo.ac.jp/

# Asymptotic Properties of the Maximum Likelihood Estimator in Regime Switching Econometric Models*

Hiroyuki Kasahara
Vancouver School of Economics
University of British Columbia
hkasahar@mail.ubc.ca

Katsumi Shimotsu
Faculty of Economics
University of Tokyo
shimotsu@e.u-tokyo.ac.jp

June 6, 2017

**Abstract**

Markov regime switching models have been widely used in numerous empirical applications in economics and finance. However, the asymptotic distribution of the maximum likelihood estimator (MLE) has not been proven for some empirically popular Markov regime switching models. In particular, the asymptotic distribution of the MLE has been unknown for models in which the regime-specific density depends on both the current and the lagged regimes, which include the seminal model of Hamilton (1989) and the switching ARCH model of Hamilton and Susmel (1994). This paper shows the asymptotic normality of the MLE and the consistency of the asymptotic covariance matrix estimate of these models.

Key words: asymptotic distribution; autoregressive conditional heteroscedasticity; maximum likelihood estimator; Markov regime switching

JEL classification numbers: C12, C13, C22

## 1 Introduction

Since the seminal contribution of Hamilton (1989), Markov regime switching models have become a popular framework for applied empirical work because they can capture the important features of time series, such as structural changes, nonlinearity, high persistence, fat tails, leptokurtosis, and asymmetric dependence (see, e.g., Evans and Wachtel, 1993; Hamilton and Susmel, 1994; Gray, 1996; Sims and Zha, 2006; Inoue and Okimoto, 2008; Ang and Bekaert, 2002; Okimoto, 2008; Dai et al., 2007).

Consider the Markov regime switching model defined by a discrete-time stochastic process $\{Y_k, X_k\}$ written as

$$Y_k = f_\theta(Y_{k-1}, \dots, Y_{k-s}, X_k; \varepsilon_k), \tag{1}$$

*This research was supported by the Natural Science and Engineering Research Council of Canada and JSPS KAKENHI Grant Number JP17K03653.

where $\{\varepsilon_k\}$ is an independent and identically distributed sequence of random variables, $\{Y_k\}$ is an inhomogeneous $s$-order Markov chain on a state space $\mathcal{Y}$ conditional on $X_k$ such that the conditional distribution of $Y_k$ only depends on $X_k$ and the lagged $Y$'s, $X_k$ is a first-order Markov process in a state space $\mathcal{X}$, and $f_\theta$ is a family of functions indexed by a finite-dimensional parameter $\theta \in \Theta$. In (1), the Markov chain $\{X_k\}$ is not observable.

Surprisingly, the asymptotic distribution of the maximum likelihood estimator (MLE) of the Markov regime switching model (1) has not been fully established in the existing literature. Bickel et al. (1998) and Jensen and Petersen (1999) derive the asymptotic normality of the MLE of hidden Markov models, in which the conditional distribution of $Y_k$ depends on $X_k$ but not on the lagged $Y$'s. For hidden Markov models and Markov regime switching models with a finite state space, the consistency of the MLE has been proven by Leroux (1992), Francq and Roussignol (1998), and Krishnamurthy and Rydén (1998).

In an influential paper, Douc et al. (2004) [DMR, hereafter] establish the consistency and asymptotic normality of the MLE in autoregressive Markov regime switching models (1) with a nonfinite hidden state space $\mathcal{X}$ under two assumptions. First, DMR assume that the conditional distribution of $Y_k$ does not depend on the lagged $X_k$'s. Specifically, on page 2259, DMR assume that

> for each $n \geq 1$ and given $\{Y_k\}_{k=n-s}^{n-1}$ and $X_n$, $Y_n$ is conditionally independent of $\{Y_k\}_{k=-s+1}^{n-s-1}$ and $\{X_k\}_{k=0}^{n-1}$.

Second, DMR assume in their Assumption A1(a) that the transition density of $X_k$ is bounded away from 0.

These two assumptions together rule out models in which the conditional density $Y_k$ depends on both the current and the lagged regimes. Suppose that we specify $X_k$ in (1) as

$$X_k = (\widetilde{X}_k, \widetilde{X}_{k-1}, \ldots, \widetilde{X}_{k-p+1}), \tag{2}$$

where $p \geq 2$, and $\widetilde{X}_k$ follows a first-order Markov process and is called the *regime*. Then, the transition density of $X_k$ inevitably has zeros. For example, when $p = 2$ and $X_k = (\tilde{X}_k, \tilde{X}_{k-1})$, we have $\Pr(X_{k+1} = (i', j') | X_k = (i, j)) = 0$ when $j' \neq i$. Consequently, the asymptotic distribution of the MLE has not been proven for some popular Markov regime switching models including the seminal model of Hamilton (1989) and the switching ARCH (SWARCH) model of Hamilton and Susmel (1994).

**Example 1** (Hamilton (1989))**.** *Consider the following model:*

$$Y_k = \mu_{\tilde{X}_k} + u_k \quad \textit{with} \quad u_k = \sum_{\ell=1}^{p-1} \gamma_\ell u_{k-\ell} + \sigma\varepsilon_k \quad \textit{for } p \geq 2, \tag{3}$$

*where* $\varepsilon_k \sim$ *i.i.d.* $N(0,1)$ *and* $\tilde{X}_k$ *follows a Markov chain on* $\tilde{\mathcal{X}} = \{1, 2, \ldots, M\}$ *with* $\Pr(\tilde{X}_k = j | \tilde{X}_{k-1} = i) = p_{ij}$*, where* $M$ *represents the number of regimes. Let*

*$\theta = (\mu_1', \ldots, \mu_M', \gamma', \{p_{ij}\}_{1\le i\le M, 1\le j\le M-1})'$ with $\gamma = (\gamma_1, \ldots, \gamma_{p-1}, \sigma)'$. Then, the conditional density of $Y_k$ given $(Y_{k-1}, \ldots, Y_{k-p})'$ and $X_k = (\tilde{X}_k, \ldots, \tilde{X}_{k-p+1})'$ is*

$$g_\theta(Y_k|Y_{k-1}, \ldots, Y_{k-p+1}, X_k) = \frac{1}{\sigma}\phi\left(\frac{(Y_k - \mu_{\tilde{X}_k}) - \sum_{\ell=1}^{p-1}\gamma_\ell(Y_{k-\ell} - \mu_{\tilde{X}_{k-\ell}})}{\sigma}\right),$$

*where $\phi(z) = (1/\sqrt{2\pi})\exp(-z^2/2)$. Hamilton (1989) estimates model (3) with $M = 2$ and $p = 5$ using data on U.S. real GNP growth.*

**Example 2** (SWARCH model of Hamilton and Susmel (1994))**.** *Consider the following model:*

$$Y_k = \mu + \gamma_y Y_{k-1} + \sigma_{\tilde{X}_k} h_k \varepsilon_k \quad \text{with} \quad h_k^2 = \gamma_0 + \sum_{\ell=1}^{p-1}\gamma_\ell\left(h_{k-\ell}\varepsilon_{k-\ell}\right)^2 \quad \text{for } p \ge 2,$$

*where $\varepsilon_k \sim$ i.i.d. $N(0,1)$ or Student $t$ with $v$ degrees of freedom, and $\tilde{X}_k$ follows a Markov chain on $\tilde{\mathcal{X}} = \{1, 2, \ldots, M\}$ with $\Pr(\tilde{X}_k = j|\tilde{X}_{k-1} = i) = p_{ij}$. For a Gaussian SWARCH model, let $\theta = (\theta_1', \ldots, \theta_M', \gamma', \{p_{ij}\}_{1\le i\le M, 1\le j\le M-1})'$ with $\theta_j = \sigma_j^2$ and $\gamma = (\mu, \gamma_y, \gamma_0, \gamma_1, \ldots, \gamma_{p-1})'$. Then, the conditional density of $Y_k$ given $(Y_{k-1}, \ldots, Y_{k-p})'$ and $X_k = (\tilde{X}_k, \ldots, \tilde{X}_{k-p+1})'$ is*

$$g_\theta(Y_k|Y_{k-1}, \ldots, Y_{k-p}, X_k) = \frac{1}{\sigma_{\tilde{X}_k} h_k}\phi\left(\frac{Y_k - \mu - \gamma_y Y_{k-1}}{\sigma_{\tilde{X}_k} h_k}\right),$$

*where $h_k^2 = \gamma_0 + \sum_{\ell=1}^{p-1}\gamma_\ell\left(\frac{Y_{k-\ell} - \mu - \gamma_y Y_{k-\ell-1}}{\sigma_{\tilde{X}_{k-\ell}}}\right)^2$.*

In the models of Hamilton (1989) and Hamilton and Susmel (1994), the transition probability of $X_k = (\tilde{X}_k, \ldots, \tilde{X}_{k-p+1})'$ has zeros when $p \ge 2$. Therefore, Assumption A1(a) of DMR is violated. As discussed on pages 2257–2258 of DMR, Assumption A1(a) is a crucial assumption for their Corollary 1 (page 2262) that establishes the *deterministic* geometrically decaying bound on the mixing rate of the conditional chain, $X|Y$. As DMR recognize on page 2258, this deterministic nature of the bound is vital to their proof of the asymptotic normality of the MLE.

This paper shows the consistency and asymptotic normality of the MLE of the Markov regime switching model (1)–(2) with $p \ge 2$, including the models of Hamilton (1989) and Hamilton and Susmel (1994). To the best of our knowledge, there exists no rigorous proof in the literature of the asymptotic normality of the MLE of these models, even though these models are popular in applied work and empirical researchers regularly make inferences based on the presumed asymptotic normality (see, e.g., Goodwin, 1993; Garcia and Perron, 1996; Hamilton and Lin, 1996; Fong, 1997; Ramchand and Susmel, 1998; Maheu and McCurdy, 2000; Edwards and Susmel, 2001). This paper therefore provides the theoretical basis for statistical inferences associated with these models.

To derive the asymptotic normality of the MLE, we first establish a bound on the mixing rate of the conditional chain, $X|Y$, in Corollary 1. Our bound is written as a product of random variables, where all but finitely many of them are strictly less than 1. Consequently, the mixing rate of the conditional chain is geometrically decaying almost surely. We then use this mixing rate to show that the sequence of the conditional scores and conditional Hessians given the $m$ past periods converge to the conditional score and conditional Hessian given the "infinite past" as $m \to \infty$. Given these results, we show the asymptotic normality of the MLE under standard regularity assumptions by applying a martingale central limit theorem to the score function (Proposition 2) as well as by proving a uniform law of large numbers for the observed Fisher information (Proposition 3). These results extend those in DMR to an empirically important class of models where the density depends on lagged regimes. Another feature of the present study is that we introduce an additional weakly exogenous regressor, $W_k$.

The remainder of this paper is organized as follows. Section 2 introduces the notation, model, and assumptions. Section 3 derives the bound on the mixing rate of the conditional chain, $X|Y$. Section 4 derives the consistency of the MLE, and the asymptotic normality of the MLE is shown in Section 5. Section 6 reports the simulation results. Section 7 collects the proofs and Section 8 collects the auxiliary results.

# 2 Model and assumptions

Our notation largely follows the notation in DMR. Let $:=$ denote "equals by definition." For a $k \times 1$ vector $x = (x_1, \ldots, x_k)'$ and a matrix $B$, define $|x| := \sqrt{x'x}$ and $|B| := \sqrt{\lambda_{\max}(B'B)}$, where $\lambda_{\max}(B'B)$ denotes the largest eigenvalue of $B'B$. For a $k \times 1$ vector $a = (a_1, \ldots, a_k)'$ and a function $f(a)$, let $\nabla_a^2 f(a) := \nabla_{aa'} f(a)$. For two probability measures $\mu_1$ and $\mu_2$, the total variation distance between $\mu_1$ and $\mu_2$ is defined as $\|\mu_1 - \mu_2\|_{TV} := \sup_A |\mu_1(A) - \mu_2(A)|$. $\|\cdot\|_{TV}$ satisfies $\sup_{f(x):0 \le f(x) \le 1} |\int f(x)\mu_1(dx) - \int f(x)\mu_2(dx)| = \|\mu_1 - \mu_2\|_{TV}$ and $\sup_{f(x):\max_x |f(x)| \le 1} |\int f(x)\mu_1(dx) - \int f(x)\mu_2(dx)| = 2\|\mu_1 - \mu_2\|_{TV}$ for any two probability measures $\mu_1$ and $\mu_2$ (see, e.g., Levin et al. (2009, Proposition 4.5)). Let $\mathbb{I}\{A\}$ denote an indicator function that takes the value of 1 when $A$ is true and 0 otherwise. $\mathcal{C}$ denotes a generic finite positive constant whose value may change from one expression to another. Let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Let $\lfloor x \rfloor$ denote the largest integer less than or equal to $x$, and define $(x)_+ := \max\{x, 0\}$. For any $\{x_i\}$, we define $\sum_{i=a}^{b} x_i := 0$ and $\prod_{i=a}^{b} x_i := 1$ when $b < a$. "i.o." stands for "infinitely often." All limits below are taken as $n \to \infty$ unless stated otherwise.

We consider the Markov regime switching process defined by a discrete-time stochastic process $\{(X_k, Y_k, W_k)\}$, where $(X_k, Y_k, W_k)$ takes the values in a set $\mathcal{X} \times \mathcal{Y} \times \mathcal{W}$ with the associated Borel $\sigma$-field $\mathcal{B}(\mathcal{X} \times \mathcal{Y} \times \mathcal{W})$. We use $p_\theta(\cdot)$ to denote densities with respect to the probability measure on $\mathcal{B}(\mathcal{X} \times \mathcal{Y} \times \mathcal{W})^{\otimes \mathbb{Z}}$. For a stochastic process $\{U_k\}$ and $a < b$, define $\mathbf{U}_a^b := (U_a, U_{a+1}, \ldots, U_b)$. Denote $\overline{\mathbf{Y}}_{k-1} := (Y_{k-1}, \ldots, Y_{k-s})$ for a fixed integer $s$ and $\overline{\mathbf{Y}}_a^b := (\overline{\mathbf{Y}}_a, \overline{\mathbf{Y}}_{a+1}, \ldots, \overline{\mathbf{Y}}_b)$. Define $Z_k := (X_k, \overline{\mathbf{Y}}_k)$. Let $Q_\theta(x, A) := \mathbb{P}_\theta(X_k \in A | X_{k-1} = x)$ denote the transition kernel of $\{X_k\}_{k=0}^{\infty}$.

We now introduce our assumptions, which mainly follow the assumptions in DMR.

**Assumption 1.** *(a) The parameter $\theta$ belongs to $\Theta$, a compact subset of $\mathbb{R}^q$, and the true parameter* *value $\theta^*$ lies in the interior of $\Theta$. (b) $\{X_k\}_{k=0}^{\infty}$ is a Markov chain that lies in a compact set $\mathcal{X} \subset \mathbb{R}^{d_x}$. (c) $Q_\theta(x,\cdot)$ has a density $q_\theta(x,\cdot)$ with respect to a* finite *dominating measure $\mu$ on $\mathcal{B}(\mathcal{X})$ such that $\mu(\mathcal{X}) = 1$, and $\sigma_+^0 := \sup_{\theta\in\Theta}\sup_{x,x'\in\mathcal{X}} q_\theta(x,x') < \infty$. (d) There exists a finite $p \geq 1$ such that $0 < \sigma_- := \inf_{\theta\in\Theta}\inf_{x,x'\in\mathcal{X}} p_\theta(x_k|x_{k-p} = x')$ and $\sigma_+ := \sup_{\theta\in\Theta}\sup_{x,x'\in\mathcal{X}} p_\theta(x_k|x_{k-p} = x') < \infty$. (e) $\{(Y_k, W_k)\}_{k=-s+1}^{\infty}$ takes the values in a set $\mathcal{Y} \times \mathcal{W} \subset \mathbb{R}^{d_y} \times \mathbb{R}^{d_w}$.*

**Assumption 2.** *(a) For each $k \geq 1$, $X_k$ is conditionally independent of $(\mathbf{X}_0^{k-2}, \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{\infty})$ given $X_{k-1}$. (b) For each $k \geq 1$, $Y_k$ is conditionally independent of $(\mathbf{Y}_{-s+1}^{k-s-1}, \mathbf{X}_0^{k-1}, \mathbf{W}_0^{k-1}, \mathbf{W}_{k+1}^{\infty})$ given $(\overline{\mathbf{Y}}_{k-1}, X_k, W_k)$, and the conditional distribution of $Y_k$ has a density $g_\theta(y_k|\overline{\mathbf{Y}}_{k-1}, X_k, W_k)$ with respect to a $\sigma$-finite measure $\nu$ on $\mathcal{B}(\mathcal{Y})$. (c) $\mathbf{W}_1^{\infty}$ is conditionally independent of $(\overline{\mathbf{Y}}_0, X_0)$ given $W_0$. (d) $\{(Z_k, W_k)\}_{k=0}^{\infty}$ is a strictly stationary ergodic process.*

**Assumption 3.** *For all $y' \in \mathcal{Y}$, $\overline{\mathbf{y}} \in \mathcal{Y}^s$, and $w \in \mathcal{W}$, $0 < \inf_{\theta\in\Theta}\inf_{x\in\mathcal{X}} g_\theta(y'|\overline{\mathbf{y}}, x, w)$ and $\sup_{\theta\in\Theta}\sup_{x\in\mathcal{X}} g_\theta(y'|\overline{\mathbf{y}}, x, w) < \infty$.*

Assumption 1(c) is also assumed on page 2258 of DMR. This assumption excludes the case where $\mathcal{X} = \mathbb{R}$ and $\mu$ is the Lebesgue measure but allows for continuously distributed $X_k$ with finite support. Assumption 1(d) implies that the state space $\mathcal{X}$ of the Markov chain $\{X_k\}$ is $\nu_p$-small for some nontrivial measure $\nu_p$ on $\mathcal{B}(\mathcal{X})$. Therefore, for all $\theta \in \Theta$, the chain $\{X_k\}$ has a unique invariant distribution and is uniformly ergodic (Meyn and Tweedie, 2009, Theorem 16.0.2). Assumptions 2(a)(b) imply that $Z_k$ is conditionally independent of $(\mathbf{Z}_0^{k-2}, \mathbf{W}_0^{k-1}, \mathbf{W}_{k+1}^{\infty})$ given $(Z_{k-1}, W_k)$; hence, $\{Z_k\}_{k=0}^{\infty}$ is a Markov chain on $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}^s$ given $\{W_k\}_{k=0}^{\infty}$. Under Assumptions 2(a)–(c), the conditional density of $\mathbf{Z}_0^n$ given $\mathbf{W}_0^n$ is written as $p_\theta(\mathbf{Z}_0^n|\mathbf{W}_0^n) = p_\theta(Z_0|W_0)\prod_{k=1}^n p_\theta(Z_k|Z_{k-1}, W_k)$. Because $\{(Z_k, W_k)\}_{k=0}^{\infty}$ is stationary, we extend $\{(Z_k, W_k)\}_{k=0}^{\infty}$ to a stationary process $\{(Z_k, W_k)\}_{k=-\infty}^{\infty}$ with doubly infinite time. We denote the probability and associated expectation of $\{(Z_k, W_k)\}_{k=\infty}^{\infty}$ under stationarity by $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$, respectively.[1] Assumption 3 is stronger than Assumption A1(b) in DMR, which assumes only $0 < \inf_{\theta\in\Theta}\int_{x\in\mathcal{X}} g_\theta(y'|\overline{\mathbf{y}}, x)\mu(dx)$ and $\sup_{\theta\in\Theta}\int_{x\in\mathcal{X}} g_\theta(y'|\overline{\mathbf{y}}, x)\mu(dx) < \infty$. When $\mathcal{X}$ is finite, Assumption 3 becomes identical to Assumption A3 of Francq and Roussignol (1998), who prove the consistency of the MLE when $\mathcal{X}$ is finite. It appears that assuming a lower bound on $g_\theta$ similar to Assumption 3 is necessary to derive the asymptotics of the MLE when $\inf_\theta \inf_{x,x'} q_\theta(x,x') = 0$. When $p = 1$, we could weaken Assumption 3 to Assumption A1(b) in DMR, but we retain Assumption 3 to simplify the exposition and proof.

Following DMR, we analyze the conditional log-likelihood function given $\overline{\mathbf{Y}}_0$, $\mathbf{W}_0^n$, and $X_0 = x_0$ rather than the stationary log-likelihood function given $\overline{\mathbf{Y}}_0$ and $\mathbf{W}_0^n$ because, as explained in DMR (pages 2263–2264), the conditional initial density $p_\theta(X_0|\overline{\mathbf{Y}}_0^{k-1})$ cannot be easily computed in practice. The conditional density function of $\mathbf{Y}_1^n$ is

$$p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) = \int \prod_{k=1}^n p_\theta(Y_k, x_k|\overline{\mathbf{Y}}_{k-1}, x_{k-1}, W_k)\mu^{\otimes n}(d\mathbf{x}_1^n), \tag{4}$$

[1] DMR use $\overline{\mathbb{P}}_\theta$ and $\overline{\mathbb{E}}_\theta$ to denote the probability and expectation under stationarity because their Section 7 deals with the case when $Z_0$ is drawn from an arbitrary distribution. Because we assume $\{(Z_k, W_k)\}_{k=\infty}^{\infty}$ is stationary throughout this paper, we use notations such as $\mathbb{P}_\theta$ and $\mathbb{E}_\theta$ without an overline for simplicity.

where $p_\theta(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1}, w_k) = q_\theta(x_{k-1}, x_k)g_\theta(y_k|\overline{\mathbf{y}}_{k-1}, x_k, w_k)$. Assumptions 2(a)(b)(c) imply that, for $k \geq 1$, $W_k$ is conditionally independent of $\mathbf{Z}_0^{k-1}$ given $\mathbf{W}_0^{k-1}$ because $p(W_k|\mathbf{Z}_0^{k-1}, \mathbf{W}_0^{k-1}) = p(\mathbf{W}_0^k, \mathbf{Z}_0^{k-1})/p(\mathbf{W}_0^{k-1}, \mathbf{Z}_0^{k-1})$ and, for $j = k, k-1$, $p(\mathbf{W}_0^j, \mathbf{Z}_0^{k-1}) = p(Z_0, \mathbf{W}_0^j)\prod_{t=1}^{k-1} p(Z_t|Z_{t-1}, W_t) = p(\mathbf{W}_1^j|W_0)p(Z_0|W_0)\prod_{t=1}^{k-1} p(Z_t|Z_{t-1}, W_t)$. Therefore, for $1 \leq k \leq n$, we have

$$p_\theta(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) = p_\theta(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \mathbf{W}_0^k, x_0), \tag{5}$$

$$p_\theta(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n) = p_\theta(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0, \mathbf{W}_0^k). \tag{6}$$

In view of (5) and (6), we can write the conditional log-likelihood function and stationary log-likelihood function as

$$\begin{aligned} l_n(\theta, x_0) &:= \log p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) = \sum_{k=1}^n \log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0), \\ l_n(\theta) &:= \log p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n) = \sum_{k=1}^n \log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k). \end{aligned} \tag{7}$$

Many applications use the log-likelihood function in which the conditional density $p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)$ is integrated with respect to $x_0$ over a probability measure $\xi$ on $\mathcal{B}(\mathcal{X})$, where $\xi$ can be fixed or treated as an additional parameter. We also analyze the resulting objective function:

$$l_n(\theta, \xi) := \log\left(\int p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)\xi(dx_0)\right). \tag{8}$$

# 3 Uniform forgetting of the conditional hidden Markov chain

In this section, we establish a mixing rate of the conditional hidden Markov chain, which is the process $\{X_k\}$ given the sequence of $Y$'s and $W$'s. The bounds on this mixing rate are instrumental in deriving the asymptotic properties of the MLE. Corollary 1 of DMR shows that the conditional hidden Markov chain forgets its past at a deterministic exponential rate. As DMR note on page 2258, their deterministic rate holds only when $p = 1$. We derive the convergence rate of the conditional Markov chain when $p \geq 1$.

First, we derive the minorization condition (Rosenthal, 1995) on the conditional hidden Markov chain. Lemma 1 of DMR derives the minorization condition when $p = 1$ and the covariate $W_k$ is absent. The following lemma generalizes Lemma 1 of DMR to accommodate $p \geq 2$ and covariate $W_k$.[2] When $p \geq 2$, the minorization coefficient $\omega(\cdot)$ depends on $(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ because $\overline{\mathbf{Y}}_{k-p}^{k-1}$ provide information on $X_k$ in addition to the information provided by $X_{k-p}$.

**Lemma 1.** *Assume Assumptions 1–3. Let $m, n \in \mathbb{Z}$ with $-m \leq n$ and $\theta \in \Theta$. Then, (a) under $\mathbb{P}_\theta$, conditionally on $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$, $\{X_k\}_{k=-m}^n$ is an inhomogeneous Markov chain, and (b) for all $-m + p \leq k \leq n$, there exists a function $\mu_k(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n, A)$ such that*

[2] We replace the conditioning variable $\overline{\mathbf{Y}}_m^n$ in DMR with $\overline{\mathbf{Y}}_{-m}^n$, because the subsequent analysis uses $\overline{\mathbf{Y}}_{-m}^n$.

*(i) For any $A \in \mathcal{B}(\mathcal{X})$, $(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n) \mapsto \mu_k(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n, A)$ is a Borel function;*

*(ii) For any $(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n)$, $\mu_k(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n, \cdot)$ is a probability measure on $\mathcal{B}(\mathcal{X})$. Furthermore, $\mu_k(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n, \cdot) \ll \mu$ for all $(\overline{\mathbf{y}}_{k-1}^n, \mathbf{w}_k^n)$, and, for all $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$,*

$$\inf_{x \in \mathcal{X}} \mathbb{P}_\theta \left( X_k \in A \middle| X_{k-p} = x, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n \right) \geq \omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mu_k(\overline{\mathbf{Y}}_{k-1}^n, \mathbf{W}_k^n, A),$$

*where $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) := \sigma_- / \sigma_+$ when $p = 1$, and, when $p \geq 2$,*[3]

$$\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) := \left( \frac{\sigma_- \inf_\theta \inf_{\mathbf{x}_{k-p+1}^{k-1}} \prod_{i=k-p+1}^{k-1} g_\theta(Y_i | \overline{\mathbf{Y}}_{i-1}, x_i, W_i)}{\sigma_+ \sup_\theta \sup_{\mathbf{x}_{k-p+1}^{k-1}} \prod_{i=k-p+1}^{k-1} g_\theta(Y_i | \overline{\mathbf{Y}}_{i-1}, x_i, W_i)} \right)^2 . \tag{9}$$

The following corollary provides the rate at which the conditional chain $\{X_k\}_{k=-m}^n$ given $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$ forgets its past. This corollary is an immediate consequence of Lemmas 1 and 8 when $-m + p \leq k \leq n$. When $k < -m + p$, this corollary holds because $\|\mu_1 - \mu_2\|_{TV} \leq 1$ for any probability measures $\mu_1$ and $\mu_2$.

**Corollary 1.** *Assume Assumptions 1–3. Let $m, n \in \mathbb{Z}$ with $-m \leq n$ and $\theta \in \Theta$. Then, for all $-m \leq k \leq n$, all probability measures $\mu_1$ and $\mu_2$ on $\mathcal{B}(\mathcal{X})$, and all $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$,*

$$\begin{aligned} &\left\| \int_{\mathcal{X}} \mathbb{P}_\theta \left( X_k \in \cdot \middle| X_{-m} = x, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n \right) \mu_1(dx) - \int_{\mathcal{X}} \mathbb{P}_\theta \left( X_k \in \cdot \middle| X_{-m} = x, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n \right) \mu_2(dx) \right\|_{TV} \\ &\leq \prod_{i=1}^{\lfloor (k+m)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{Y}}_{-m+pi-p}^{-m+pi-1}, \mathbf{W}_{-m+pi-p}^{-m+pi-1}) \right). \end{aligned}$$

The convergence rate of the conditional hidden Markov chain depends on the minorization coefficient $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$. If this coefficient is bounded away from 0, the chain forgets its past exponentially fast. When $p \geq 2$, this coefficient is not bounded away from 0 because $\inf_{\overline{\mathbf{y}}_{k-p}^{k-1}, \mathbf{w}_{k-p}^{k-1}} \omega(\overline{\mathbf{y}}_{k-p}^{k-1}, \mathbf{w}_{k-p}^{k-1}) = 0$. However, $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ becomes close to zero only when $\mathbf{Y}_{k-p+1}^{k-1}$ takes an unlikely value because the denominator of $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is finite and the numerator of $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is a product of the conditional density $g_\theta(y | \overline{\mathbf{y}}, x, w)$. As a result, $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is bounded away from 0 with a probability close to 1. In the following sections, we use this fact to establish the consistency and asymptotic normality of the MLE.

[3] Strictly speaking, $W_{k-p}$ in $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is superfluous because $\omega(\overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ does not depend on $W_{k-p}$. We retain $W_{k-p}$ for notational simplicity.

## 4 Consistency of the MLE

Define the conditional MLE of $\theta^*$ given $\overline{\mathbf{Y}}_0$, $\mathbf{W}_0^n$, and $X_0 = x_0$ as

$$\hat{\theta}_{x_0} := \operatorname*{argmax}_{\theta \in \Theta} l_n(\theta, x_0),$$

with $l_n(\theta, x_0)$ defined in (7). In this section, we prove the consistency of the conditional MLE. We introduce additional assumptions required for proving consistency.

**Assumption 4.** *(a)* $b_+ := \sup_{\theta\in\Theta} \sup_{y_1, \overline{\mathbf{y}}_0, x, w} g_\theta(y_1|\overline{\mathbf{y}}_0, x, w) < \infty$. *(b)* $\mathbb{E}_{\theta^*}|\log b_-(\overline{\mathbf{Y}}_0^1, W_1)| < \infty$, *where* $b_-(\overline{\mathbf{Y}}_{k-1}^k, W_k) := \inf_{\theta\in\Theta} \inf_{x_k \in \mathcal{X}} g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k)$.

**Assumption 5.** *There exist constants* $\alpha > 0$, $C_1, C_2 \in (0, \infty)$, *and* $\beta > 1$ *such that, for any* $r > 0$,

$$\mathbb{P}_{\theta^*}\left(b_-(\overline{\mathbf{Y}}_0^1, W_1) \leq C_1 e^{-\alpha r}\right) \leq C_2 r^{-\beta}.$$

Assumption 4(a) is also assumed in Assumption (A3) of DMR. Assumption 4(b) is stronger than Assumption (A3) of DMR, who assume $\mathbb{E}_{\theta^*}|\log(\inf_{\theta\in\Theta}\int g_\theta(Y_1|\overline{\mathbf{Y}}_0, x)\mu(dx)| < \infty$. Assumption 4 implies that $\mathbb{E}_{\theta^*}\sup_{\theta\in\Theta}\sup_{x\in\mathcal{X}}|\log(g_\theta(Y_1|\overline{\mathbf{Y}}_0, x, W_1))| < \infty$, which is similar to the moment condition used in the standard maximum likelihood estimation, but the infimum is taken over $x$ in addition to $\theta$. Assumption 5 restricts the probability that $\inf_{\theta\in\Theta}\inf_{x_k\in\mathcal{X}} g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k)$ takes an extremely small value. Assumption 5 is not restrictive because the right hand side of the inequality inside $\mathbb{P}_{\theta^*}(\cdot)$ is exponential in $r$ and the bound $C_2 r^{-\beta}$ is a polynomial in $r$. An easily verifiable sufficient condition for Assumption 5 is $\mathbb{E}_{\theta^*}|\log b_-(\overline{\mathbf{Y}}_0^1, W_1)|^{1+\delta} < \infty$ for some $\delta > 0$. This is because $\mathbb{P}_{\theta^*}(b_-(\overline{\mathbf{Y}}_0^1, W_1) \leq b_+ e^{-\alpha r}) = \mathbb{P}_{\theta^*}(|\log(b_-(\overline{\mathbf{Y}}_0^1, W_1)/b_+)| \geq \alpha r) \leq (\mathbb{E}_{\theta^*}|\log(b_-(\overline{\mathbf{Y}}_0^1, W_1)/b_+)|^{1+\delta})/(\alpha r)^{1+\delta} \leq C_2 r^{-(1+\delta)}$, where the first equality follows from $b_-(\overline{\mathbf{Y}}_0^1, W_1) \leq b_+$, and the second and third inequalities follow from Markov's inequality and Minkowski's inequality. Examples 1 and 2 satisfy Assumptions 4 and 5.

In the following lemma, we show that the difference between the conditional log-likelihood function $l_n(\theta, x_0)$ and the stationary log-likelihood function $l_n(\theta)$ is $o(n)$ $\mathbb{P}_{\theta^*}$-a.s.

**Lemma 2.** *Assume Assumptions 1–5. Then,*

$$n^{-1}\sup_{x_0\in\mathcal{X}}\sup_{\theta\in\Theta}|l_n(\theta, x_0) - l_n(\theta)| \to 0 \quad \mathbb{P}_{\theta^*}\text{-}a.s.$$

When $p = 1$, Lemma 2 of DMR shows that $\sup_{\theta\in\Theta}|l_n(\theta, x_0) - l_n(\theta)|$ is bounded by a deterministic constant. When $p \geq 2$, Lemma 2 of DMR is no longer applicable because $|l_n(\theta, x_0) - l_n(\theta)|$ depends on the products of $1 - \omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1}, \mathbf{W}_{pi-p}^{pi-1})$'s for $i = 1, \ldots, \lfloor n/p \rfloor$. A key observation is that $\{\omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1}, \mathbf{W}_{pi-p}^{pi-1})\}_{i\geq 1}$ is stationary and ergodic and that $\epsilon := \mathbb{P}_{\theta^*}(\omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1}, \mathbf{W}_{pi-p}^{pi-1}) \leq \delta)$ is small when $\delta > 0$ is sufficiently small. Because the strong law of large numbers implies that $(\lfloor n/p \rfloor)^{-1}\sum_{i=1}^{\lfloor n/p \rfloor}\mathbb{I}\{\omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1}, \mathbf{W}_{pi-p}^{pi-1}) > \delta\}$ converges to $1 - \epsilon$ $\mathbb{P}_{\theta^*}$-a.s. from the strong law of large numbers, $1-\omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1},\mathbf{W}_{pi-p}^{pi-1})>1-\delta$ holds for a large fraction of the $\omega(\overline{\mathbf{Y}}_{pi-p}^{pi-1},\mathbf{W}_{pi-p}^{pi-1})$'s. Consequently, we can establish a $\mathbb{P}_{\theta^*}$-a.s. bound on $n^{-1}|l_n(\theta,x_0)-l_n(\theta)|$.

We proceed to show that, for all $\theta\in\Theta$, $p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1},\mathbf{W}_{-m}^{k})$ converges to $p_\theta(Y_k|\overline{\mathbf{Y}}_{-\infty}^{k-1},\mathbf{W}_{-\infty}^{k})$ $\mathbb{P}_{\theta^*}$-a.s. as $m\to\infty$ and that we can approximate $n^{-1}l_n(\theta)$ by $n^{-1}\sum_{k=1}^n\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-\infty}^{k-1},\mathbf{W}_{-\infty}^{k})$, which is the sample average of the stationary ergodic random variables. For $x\in\mathcal{X}$ and $m\geq 0$, define

$$\begin{aligned}
\Delta_{k,m,x}(\theta) &:= \log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1},\mathbf{W}_{-m}^{k},X_{-m}=x),\\
\Delta_{k,m}(\theta) &:= \log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1},\mathbf{W}_{-m}^{k})\\
&= \log\int p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1},\mathbf{W}_{-m}^{k},X_{-m}=x_{-m})\mathbb{P}_\theta(dx_{-m}|\overline{\mathbf{Y}}_{-m}^{k-1},\mathbf{W}_{-m}^{k}),
\end{aligned}$$

so that $l_n(\theta)=\sum_{k=1}^n\Delta_{k,0}(\theta)$. The following proposition corresponds to Lemma 3 of DMR. This proposition shows that, for any $k\geq 0$, the sequences $\{\Delta_{k,m}(\theta)\}_{m\geq 0}$ and $\{\Delta_{k,m,x}(\theta)\}_{m\geq 0}$ are uniformly Cauchy sequences and hence converge uniformly in $\theta\in\Theta$ with probability one.

**Lemma 3.** *Assume Assumptions 1–5. Then, there exist a constant $\rho\in(0,1)$ and random sequences $\{A_{k,m}\}_{k\geq 1,m\geq 0}$ and $\{B_k\}_{k\geq 1}$ such that, for all $1\leq k\leq n$ and $m'\geq m\geq 0$,*

$$\begin{aligned}
&(a)\quad \sup_{x,x'\in\mathcal{X}}\sup_{\theta\in\Theta}\left|\Delta_{k,m,x}(\theta)-\Delta_{k,m',x'}(\theta)\right|\leq A_{k,m}\rho^{\lfloor (k+m)/3p\rfloor},\\
&(b)\quad \sup_{x\in\mathcal{X}}\sup_{\theta\in\Theta}|\Delta_{k,m,x}(\theta)-\Delta_{k,m}(\theta)|\leq A_{k,m}\rho^{\lfloor (k+m)/3p\rfloor},\\
&(c)\quad \sup_{m\geq 0}\sup_{x\in\mathcal{X}}\sup_{\theta\in\Theta}|\Delta_{k,m,x}(\theta)|+\sup_{m\geq 0}\sup_{\theta\in\Theta}|\Delta_{k,m}(\theta)|\leq B_k,
\end{aligned}$$

*where $\mathbb{P}_{\theta^*}(A_{k,m}\geq M \text{ i.o.})=0$ for a constant $M<\infty$ and $B_k\in L^1(\mathbb{P}_{\theta^*})$.*

Lemma 3(a) implies that $\{\Delta_{k,m,x}(\theta)\}_{m\geq 0}$ is a uniform Cauchy sequence in $\theta\in\Theta$ with probability one and that $\lim_{m\to\infty}\Delta_{k,m,x}(\theta)$ does not depend on $x$. Let $\Delta_{k,\infty}(\theta)$ denote this limit. Because $\{\Delta_{k,m,x}(\theta)\}_{m\geq 0}$ is uniformly bounded in $L^1(\mathbb{P}_{\theta^*})$ from Lemma 3(c), $\{\Delta_{k,m,x}(\theta)\}_{m\geq 0}$ converges to $\Delta_{k,\infty}(\theta)$ in $L^1(\mathbb{P}_{\theta^*})$ and $\Delta_{k,\infty}(\theta)\in L^1(\mathbb{P}_{\theta^*})$ by the dominated convergence theorem. Define $l(\theta):=\mathbb{E}_{\theta^*}[\Delta_{0,\infty}(\theta)]$. Lemma 3 also implies that $n^{-1}l_n(\theta)$ converge to $n^{-1}\sum_{k=1}^n\Delta_{k,\infty}(\theta)$, which converges to $l(\theta)$ by the ergodic theorem. Therefore, the consistency of $\hat{\theta}_{x_0}$ is proven if this convergence of $n^{-1}l_n(\theta)-l(\theta)$ is strengthened to uniform convergence in $\theta\in\Theta$ and the additional regularity conditions are confirmed.

We introduce additional assumptions on the continuity of $q_\theta$ and $g_\theta$ and identification of $\theta^*$.

**Assumption 6.** *(a) For all $(\overline{\mathbf{y}},y',w)\in\mathcal{Y}^s\times\mathcal{Y}\times\mathcal{W}$ and uniformly in $x,x'\in\mathcal{X}$, $q_\theta(x,x')$ and $g_\theta(y'|\overline{\mathbf{y}},x,w)$ are continuous in $\theta$. (b) $\mathbb{P}_{\theta^*}[p_{\theta^*}(Y_1|\overline{\mathbf{Y}}_{-m}^{0},\mathbf{W}_{-m}^{1})\neq p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^{0},\mathbf{W}_{-m}^{1})]>0$ for all $m\geq 0$ and all $\theta\in\Theta$ such that $\theta\neq\theta^*$.*

Assumption 6(b) is a high-level assumption because it is imposed on $p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^{0},\mathbf{W}_{-m}^{1})$. When the covariate $W_k$ is absent, DMR prove consistency under a lower-level assumption (their (A5′)), which is stated in terms of $p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0)$. We use Assumption 6(b) for brevity.

The following proposition shows the strong consistency of the (conditional) MLE.[4]

**Proposition 1.** *Assume Assumptions 1–6. Then,* $\sup_{x_0\in\mathcal{X}}|\hat{\theta}_{x_0}-\theta^*|\to 0$ $\mathbb{P}_{\theta^*}$*-a.s.*

Francq and Roussignol (1998, Theorem 3) prove the consistency of the MLE when the state space of $X_k$ is finite. Proposition 1 generalizes Theorem 3 of Francq and Roussignol (1998) in the following three aspects. First, we allow $X_k$ to be continuously distributed. Second, we analyze the log-likelihood function conditional on $X_0=x_0$, whereas Francq and Roussignol (1998) set the initial distribution of $X_1$ to any probability vector with strictly positive elements. In other words, we allow for zeros in the postulated initial distribution of $\{X_k\}$. Third, we allow for an exogenous covariate $\{W_k\}_{k=0}^n$.

Define the MLE with a probability measure $\xi$ on $\mathcal{B}(\mathcal{X})$ for $x_0$ as $\hat{\theta}_\xi := \operatorname{argmax}_{\theta\in\Theta} l_n(\theta,\xi)$ with $l_n(\theta,\xi)$ defined in (8). Proposition 1 implies the following corollary.

**Corollary 2.** *Assume Assumptions 1–6. Then, for any* $\xi$, $\hat{\theta}_\xi\to\theta^*$ $\mathbb{P}_{\theta^*}$*-a.s.*

# 5 Asymptotic distribution of the MLE

In this section, we derive the asymptotic distribution of the MLE and consistency of the asymptotic covariance matrix estimate. Because $\hat{\theta}_{x_0}$ is consistent, expanding the first-order condition $\nabla_\theta l_n(\hat{\theta}_{x_0},x_0)=0$ around $\theta^*$ gives

$$0=\nabla_\theta l_n(\hat{\theta}_{x_0},x_0)=\nabla_\theta l_n(\theta^*,x_0)+\nabla_\theta^2 l_n(\overline{\theta},x_0)(\hat{\theta}_{x_0}-\theta^*), \tag{10}$$

where $\overline{\theta}\in[\theta^*,\hat{\theta}_{x_0}]$ and $\overline{\theta}$ may take different values across different rows of $\nabla_\theta^2 l_n(\overline{\theta},x_0)$. In the following, we approximate $\nabla_\theta^j l_n(\theta,x_0)=\sum_{k=1}^n\nabla_\theta^j\log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1},\mathbf{W}_0^k,X_0=x_0)$ for $j=1,2$ by $\sum_{k=1}^n\nabla_\theta^j\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-\infty}^{k-1},\mathbf{W}_{-\infty}^k)$, which is a sum of a stationary process. We then apply the central limit theorem and law of large numbers to $n^{-j/2}\sum_{k=1}^n\nabla_\theta^j\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-\infty}^{k-1},\mathbf{W}_{-\infty}^k)$. A similar expansion gives the asymptotic distribution of $n^{1/2}(\hat{\theta}_\xi-\theta^*)$.

We introduce additional assumptions. Define $\mathcal{X}_\theta^+ := \{(x,x')\in\mathcal{X}^2 : q_\theta(x,x')>0\}$.

**Assumption 7.** *There exists a constant $\delta>0$ such that the following conditions hold on $G:=\{\theta\in\Theta : |\theta-\theta^*|<\delta\}$: (a) For all $(\overline{\mathbf{y}},y',w,x,x')\in\mathcal{Y}^s\times\mathcal{Y}\times\mathcal{W}\times\mathcal{X}\times\mathcal{X}$, the functions $g_\theta(y'|\overline{\mathbf{y}},w,x)$ and $q_\theta(x,x')$ are twice continuously differentiable in $\theta\in G$. (b) $\sup_{\theta\in G}\sup_{x,x'\in\mathcal{X}_\theta^+}|\nabla_\theta\log q_\theta(x,x')|<\infty$ and $\sup_{\theta\in G}\sup_{x,x'\in\mathcal{X}_\theta^+}|\nabla_\theta^2\log q_\theta(x,x')|<\infty$. (c) $\mathbb{E}_{\theta^*}[\sup_{\theta\in G}\sup_{x\in\mathcal{X}}|\nabla_\theta\log g_\theta(Y_1|\overline{\mathbf{Y}}_0,x,W_1)|^2]<\infty$ and $\mathbb{E}_{\theta^*}[\sup_{\theta\in G}\sup_{x\in\mathcal{X}}|\nabla_\theta^2\log g_\theta(Y_1|\overline{\mathbf{Y}}_0,x,W_1)|]<\infty$. (d) For almost all $(\overline{\mathbf{y}},y',w)\in\mathcal{Y}^s\times\mathcal{Y}\times$*

[4] A Gaussian regime switching model with regime-specific mean $\mu_j$ and variance $\sigma_j^2$ is subject to the unbounded likelihood problem (Hartigan, 1985) in that the likelihood diverges to infinity if we set $\mu_j=Y_k$ for some $k$ and let $\sigma_j\to 0$. In this paper, the compactness assumption (Assumption 1(a)) in effect imposes a lower bound on $\sigma_j$ and hence rules out the unbounded likelihood problem.

$\mathcal{W}$, *there exists a function* $f_{\overline{\mathbf{y}},y',w} : \mathcal{X} \to \mathbb{R}^+$ *in* $L^1(\mu)$ *such that* $\sup_{\theta \in G} g_\theta(y'|\overline{\mathbf{y}}, x, w) \leq f_{\overline{\mathbf{y}},y',w}(x)$.
*(e) For almost all* $(x, \overline{\mathbf{y}}, w) \in \mathcal{X} \times \mathcal{Y}^s \times \mathcal{W}$ *and* $j = 1, 2$, *there exist functions* $f^j_{x,\overline{\mathbf{y}},w} : \mathcal{Y} \to \mathbb{R}^+$ *in* $L^1(\nu)$ *such that* $|\nabla^j_\theta g_\theta(y'|\overline{\mathbf{y}}, x, w)| \leq f^j_{x,\overline{\mathbf{y}},w}(y')$ *for all* $\theta \in G$.

**Assumption 8.** $\mathbb{E}_{\theta^*}[\sup_{m \geq 0} \sup_{\theta \in G} |\nabla_\theta \log p_\theta(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{W}^1_{-m})|^2] < \infty$,
$\mathbb{E}_{\theta^*}[\sup_{m \geq 0} \sup_{\theta \in G} |\nabla^2_\theta \log p_\theta(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{W}^1_{-m})|] < \infty$,
$\mathbb{E}_{\theta^*}[\sup_{m \geq 0} \sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\nabla_\theta \log p_\theta(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{W}^1_{-m}, X_{-m} = x)|^2] < \infty$, *and*
$\mathbb{E}_{\theta^*}[\sup_{m \geq 0} \sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\nabla^2_\theta \log p_\theta(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{W}^1_{-m}, X_{-m} = x)|] < \infty$.

Assumption 7 is the same as Assumptions (A6)–(A8) of DMR except for accommodating the case $\inf_{(x,x') \in \mathcal{X}^2} q_\theta(x, x') = 0$ and the covariate $W$. Examples 1 and 2 satisfy Assumption 7. Assumption 8 is a high-level assumption that bounds the moments of $\nabla^j_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m})$ and $\nabla^j_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x)$ uniformly in $m$. When $p = 1$, DMR could derive Assumption 8 by using the $L^{3-j}(\mathbb{P}_{\theta^*})$ convergence of $\nabla^j_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m})$ and $\nabla^j_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x)$ to $\nabla^j_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-\infty}, \mathbf{W}^k_{-\infty})$ as $m \to \infty$. When $p \geq 2$, we need to assume Assumption 8 because our Lemma 6 only shows that these sequences converge to $\nabla^j_\theta \log p_\theta(\mathbf{Y}_k|\overline{\mathbf{Y}}^{k-1}_{-\infty}, \mathbf{W}^k_{-\infty})$ in probability.

### 5.1 Asymptotic distribution of the score function

This section derives the asymptotic distribution of $n^{-1/2}\nabla_\theta l_n(\theta^*, x_0)$ and $n^{-1/2}\nabla_\theta l_n(\theta^*, \xi)$. We introduce a result known as the Louis missing information principle (Louis, 1982), which expresses the derivatives of the log-likelihood function of a latent variable model in terms of the conditional expectation of the derivatives of the *complete data* log-likelihood function. Let $(X, Y, W)$ be random variables with $p_\theta(y, x|w)$ denoting the joint density of $(Y, X)$ given $W$, and let $p_\theta(y|w)$ be the marginal density of $Y$ given $W$. Then, a straightforward differentiation that is valid under Assumption 7 gives

$$\begin{aligned}
\nabla_\theta \log p_\theta(Y|W) &= \mathbb{E}_\theta \left[\nabla_\theta \log p_\theta(Y, X|W)|Y, W\right], \\
\nabla^2_\theta \log p_\theta(Y|W) &= \mathbb{E}_\theta \left[\nabla^2_\theta \log p_\theta(Y, X|W)\big|Y, W\right] + \mathrm{var}_\theta \left[\nabla_\theta \log p_\theta(Y, X|W)|Y, W\right].
\end{aligned} \tag{11}$$

Define $\overline{\mathbf{Z}}^k_{k-1} := (Y_k, X_k, \overline{\mathbf{Y}}_{k-1}, X_{k-1})$. For $j = 1, 2$, denote the derivatives of the complete data log-density of $(Y_k, X_k)$ given $(\overline{\mathbf{Y}}_{k-1}, X_{k-1}, W_k)$ by

$$\begin{aligned}
\phi^j(\theta, \overline{\mathbf{Z}}^k_{k-1}, W_k) &:= \nabla^j_\theta \log p_\theta(Y_k, X_k|\overline{\mathbf{Y}}_{k-1}, X_{k-1}, W_k) \\
&= \nabla^j_\theta \log q_\theta(X_{k-1}, X_k) + \nabla^j_\theta \log g_\theta(Y_k, |\overline{\mathbf{Y}}_{k-1}, X_k, W_k).
\end{aligned}$$

We use a short-handed notation $\phi^j_{\theta k} := \phi^j(\theta, \overline{\mathbf{Z}}^k_{k-1}, W_k)$. We also suppress the superscript 1 from $\phi^1_{\theta k}$, so that $\phi_{\theta k} = \phi^1_{\theta k}$. Let $|\phi^j_k|_\infty := \sup_{\theta \in G} \sup_{x,x' \in \mathcal{X}^+_\theta} |\nabla^j_\theta \log q_\theta(x, x')|$
$+ \sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\nabla^j_\theta \log g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x, W_k)|$.

Define, for $x \in \mathcal{X}$, $k \geq 1$, $m \geq 0$, and $j = 1, 2$,[5]

$$\Psi^j_{k,m,x}(\theta) := \mathbb{E}_\theta\left[\sum_{t=-m+1}^{k} \phi^j_{\theta t}\middle|\overline{\mathbf{Y}}^k_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x\right] - \mathbb{E}_\theta\left[\sum_{t=-m+1}^{k-1} \phi^j_{\theta t}\middle|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^{k-1}_{-m}, X_{-m} = x\right]. \tag{12}$$

Because $\Psi^1_{k,m,x}(\theta) = \nabla_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x)$ from (11), we can express $\nabla_\theta l_n(\theta, x_0)$ as

$$\nabla_\theta l_n(\theta, x_0) = \sum_{k=1}^{n} \nabla_\theta \log p_\theta(Y_k|\overline{\mathbf{Y}}^{k-1}_0, \mathbf{W}^k_0, X_0 = x_0) = \sum_{k=1}^{n} \Psi^1_{k,0,x_0}(\theta).$$

We show that $\{\Psi^j_{k,m,x}(\theta)\}_{m\geq 0}$ is a Cauchy sequence that converges to a limit at an exponential rate. Note that $\Psi^j_{k,m,x}(\theta)$ is a function of $\mathbb{E}_\theta[\phi^j_{\theta t}|\cdot]$ for $t = -m+1, \ldots, k$. When $t$ is large, the difference between $\mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{Y}}^k_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x]$ and $\mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{Y}}^k_{-m'}, \mathbf{W}^k_{-m'}, X_{-m'} = x']$ with $m' > m$ is small by virtue of Corollary 1. When $t$ is small, we bound $\mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{Y}}^k_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x] - \mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^{k-1}_{-m}, X_{-m} = x]$ in $\Psi^j_{k,m,x}(\theta)$ by using the following lemma. This lemma corresponds to Lemma 9 of DMR and derives the minorization constant for the time-reversed process $\{X_{n-k}\}_{0\leq k\leq n+m}$ conditional on $(\overline{\mathbf{Y}}^n_{-m}, \mathbf{W}^n_{-m})$.

**Lemma 4.** *Assume Assumptions 1 and 2. Let $m, n \in \mathbb{Z}$ with $-m \leq n$ and $\theta \in \Theta$. Then, (a) under $\mathbb{P}_\theta$, conditionally on $(\overline{\mathbf{Y}}^n_{-m}, \mathbf{W}^n_{-m})$, the time-reversed process $\{X_{n-k}\}_{0\leq k\leq n+m}$ is an inhomogeneous Markov chain, and (b) for all $p \leq k \leq n+m$, there exists a function $\tilde{\mu}_k(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}, A)$ such that*

*(i) For any $A \in \mathcal{B}(\mathcal{X})$, $(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}) \mapsto \tilde{\mu}_k(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}, A)$ is a Borel function;*

*(ii) For any $(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}, A)$, $\tilde{\mu}_k(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}, \cdot)$ is a probability measure on $\mathcal{B}(\mathcal{X})$. Furthermore, $\tilde{\mu}_k(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m}, \cdot) \ll \mu$ for all $(\overline{\mathbf{y}}^{n-k+p-1}_{-m}, \mathbf{w}^{n-k+p-1}_{-m})$, and, for all $(\overline{\mathbf{Y}}^{n-k+p-1}_{-m}, \mathbf{W}^{n-k+p-1}_{-m})$,*

$$\begin{aligned}
&\mathbb{P}_\theta\left(X_{n-k} \in A \middle| X_{n-k+p}, \overline{\mathbf{Y}}^n_{-m}, \mathbf{W}^n_{-m}\right) \\
&= \mathbb{P}_\theta\left(X_{n-k} \in A \middle| X_{n-k+p}, \overline{\mathbf{Y}}^{n-k+p-1}_{-m}, \mathbf{W}^{n-k+p-1}_{-m}\right) \\
&\geq \omega(\overline{\mathbf{Y}}^{n-k+p-1}_{n-k}, \mathbf{W}^{n-k+p-1}_{n-k})\tilde{\mu}_k(\overline{\mathbf{Y}}^{n-k+p-1}_{-m}, \mathbf{W}^{n-k+p-1}_{-m}, A),
\end{aligned}$$

*where $\omega(\overline{\mathbf{Y}}^{n-k+p-1}_{n-k}, \mathbf{W}^{n-k+p-1}_{n-k}) := \sigma_-/\sigma_+$ when $p = 1$, and, when $p \geq 2$, $\omega(\overline{\mathbf{Y}}^{n-k+p-1}_{n-k}, \mathbf{W}^{n-k+p-1}_{n-k})$ is defined as in (9) but replacing $k-1$ and $k-p$ in (9) with $n-k+p-1$ and $n-k$.*

The following lemma shows that the time-reversed process $\{X_{n-k}\}_{0\leq k\leq n+m}$ conditional on

[5] DMR (page 2272) use the symbol $\Delta_{k,m,x}(\theta)$ to denote our $\Psi^1_{k,m,x}(\theta)$, but we use $\Psi_{k,m,x}(\theta)$ to avoid confusion with $\Delta_{k,m,x}(\theta)$ used in Lemma 3.

$(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$ forgets its initial conditioning variable (i.e., $Y_n$ and $W_n$) exponentially fast. Part (b) corresponds to equation (39) on page 2294 of DMR.

**Lemma 5.** *Assume Assumptions 1 and 2. Let $m, n \in \mathbb{Z}$ with $m, n \geq 0$ and $\theta \in \Theta$. Then,*
*(a) for all $-m \leq k \leq n$ and all $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$,*

$$\begin{aligned} &\left\| \mathbb{P}_\theta \left( X_k \in \cdot \middle| \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n \right) - \mathbb{P}_\theta \left( X_k \in \cdot \middle| \overline{\mathbf{Y}}_{-m}^{n-1}, \mathbf{W}_{-m}^{n-1} \right) \right\|_{TV} \\ &\leq \prod_{i=1}^{\lfloor (n-1-k)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{Y}}_{n-2-pi+1}^{n-2-pi+p}, \mathbf{W}_{n-2-pi+1}^{n-2-pi+p}) \right). \end{aligned}$$

*(b) for all $-m+1 \leq k \leq n$ and all $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n, x)$,*

$$\begin{aligned} &\left\| \mathbb{P}_\theta \left( X_k \in \cdot \middle| \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n, X_{-m} = x \right) - \mathbb{P}_\theta \left( X_k \in \cdot \middle| \overline{\mathbf{Y}}_{-m}^{n-1}, \mathbf{W}_{-m}^{n-1}, X_{-m} = x \right) \right\|_{TV} \\ &\leq \prod_{i=1}^{\lfloor (n-1-k)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{Y}}_{n-2-pi+1}^{n-2-pi+p}, \mathbf{W}_{n-2-pi+1}^{n-2-pi+p}) \right). \end{aligned}$$

Define, for $k \geq 0$, $m \geq 0$, and $j = 1, 2$,

$$\Psi_{k,m}^j(\theta) := \mathbb{E}_\theta \left[ \sum_{t=-m+1}^{k} \phi_{\theta t}^j \middle| \overline{\mathbf{Y}}_{-m}^k, \mathbf{W}_{-m}^k \right] - \mathbb{E}_\theta \left[ \sum_{t=-m+1}^{k-1} \phi_{\theta t}^j \middle| \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1} \right].$$

Note that $\Psi_{k,m}^1(\theta) = \nabla_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k)$. From Corollary 1 and Lemma 5, we obtain the following bound on $\Psi_{k,m,x}^j(\theta) - \Psi_{k,m}^j(\theta)$ and $\Psi_{k,m,x}^j(\theta) - \Psi_{k,m',x'}^j(\theta)$.

**Lemma 6.** *Assume Assumptions 1–8. Then, for $j = 1, 2$, there exist a constant $\rho \in (0, 1)$, random sequences $\{A_{k,m}\}_{k \geq 1, m \geq 0}$ and $\{B_m\}_{m \geq 0}$, and a random variable $K_j \in L^{3-j}(\mathbb{P}_{\theta^*})$ such that, for all $1 \leq k \leq n$ and $m' \geq m \geq 0$,*

$$\begin{aligned} (a) \quad & \sup_{\theta \in G} \sup_{x \in \mathcal{X}} \left| \Psi_{k,m,x}^j(\theta) - \Psi_{k,m}^j(\theta) \right| \leq K_j (k+m)^2 \rho^{\lfloor (k+m)/4(p+1) \rfloor} A_{k,m}, \\ (b) \quad & \sup_{\theta \in G} \sup_{x, x' \in \mathcal{X}} \left| \Psi_{k,m,x}^j(\theta) - \Psi_{k,m',x'}^j(\theta) \right| \leq [K_j (k+m)^2 + B_m] \rho^{\lfloor (k+m)/4(p+1) \rfloor} A_{k,m}, \end{aligned}$$

*where $\mathbb{P}_{\theta^*}(A_{k,m} \geq 1 \text{ i.o.}) = 0$, $B_m < \infty$ a.s., and the distribution function of $B_m$ does not depend on $m$.*

Because $B_m \rho^{\lfloor (k+m)/4(p+1) \rfloor / 2} \to_p 0$ as $m \to \infty$, Lemma 6 implies that $\{\Psi_{k,m,x}^1(\theta)\}_{m \geq 0}$ converges to $\Psi_{k,\infty}^1(\theta) = \nabla_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}_{-\infty}^{k-1}, \mathbf{W}_{-\infty}^k)$ in probability uniformly in $\theta \in G$ and $x \in \mathcal{X}$. Define the filtration $\mathcal{F}$ by $\mathcal{F}_k := \sigma((\overline{\mathbf{Y}}_i, W_{i+1}) : -\infty < i \leq k)$. It follows from $\mathbb{E}_{\theta^*}[\Psi_{k,m}^1(\theta^*) | \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k] = 0$, Assumption 8, and combining Exercise 2.3.7 and Theorem 5.5.9 of Durrett (2010) that $\mathbb{E}_{\theta^*}[\Psi_{k,\infty}^1(\theta^*) | \overline{\mathbf{Y}}_{-\infty}^{k-1}, \mathbf{W}_{-\infty}^k] = 0$ and $I(\theta^*) := \mathbb{E}_{\theta^*}[\Psi_{0,\infty}^1(\theta^*)(\Psi_{0,\infty}^1(\theta^*))'] < \infty$. Therefore,

$\{\Psi^1_{k,\infty}(\theta^*)\}_{k=-\infty}^{\infty}$ is an $(\mathcal{F}, \mathbb{P}_{\theta^*})$-adapted stationary, ergodic, and square integrable martingale difference sequence, to which a martingale central limit theorem is applicable. The following proposition establishes the asymptotic distribution of the score function.

**Proposition 2.** *Assume Assumptions 1–8. Then, (a) for any $x_0 \in \mathcal{X}$, $n^{-1/2}\nabla_\theta l_n(\theta^*, x_0) \to_d N(0, I(\theta^*))$; (b) for any probability measure $\xi$ on $\mathcal{B}(\mathcal{X})$ for $x_0$, $n^{-1/2}\nabla_\theta l_n(\theta^*, \xi) \to_d N(0, I(\theta^*))$.*

## 5.2 Convergence of the Hessian

This section derives the probability limit of $n^{-1}\nabla^2_\theta l_n(\theta, x_0)$ and $n^{-1}\nabla^2_\theta l_n(\theta, \xi)$ when $\theta$ is in a neighborhood of $\theta^*$. Define

$$\Gamma_{k,m,x}(\theta) := \mathrm{var}_\theta \left[ \sum_{t=-m+1}^{k} \phi_{\theta t} \middle| \overline{\mathbf{Y}}^k_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x \right] - \mathrm{var}_\theta \left[ \sum_{t=-m+1}^{k-1} \phi_{\theta t} \middle| \overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^{k-1}_{-m}, X_{-m} = x \right], \tag{13}$$

$$\Gamma_{k,m}(\theta) := \mathrm{var}_\theta \left[ \sum_{t=-m+1}^{k} \phi_{\theta t} \middle| \overline{\mathbf{Y}}^k_{-m}, \mathbf{W}^k_{-m} \right] - \mathrm{var}_\theta \left[ \sum_{t=-m+1}^{k-1} \phi_{\theta t} \middle| \overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^{k-1}_{-m} \right]. \tag{14}$$

From the Louis missing information principle (11), we can write $\nabla^2_\theta l_n(\theta, x_0)$ in terms of $\{\Psi^2_{k,m,x}(\theta)\}$ and $\{\Gamma_{k,m,x}(\theta)\}$ as

$$\nabla^2_\theta l_n(\theta, x_0) = \sum_{k=1}^{n} \nabla^2_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}^{k-1}_0, \mathbf{W}^k_0, X_0 = x_0) = \sum_{k=1}^{n} [\Psi^2_{k,0,x_0}(\theta) + \Gamma_{k,0,x_0}(\theta)].$$

The following lemma provides the bounds on $\Gamma_{k,m,x}(\theta)$ that are analogous to Lemma 6.

**Lemma 7.** *Assume Assumptions 1–8. Then, there exist a constant $\rho \in (0, 1)$, random sequences $\{C_{k,m}\}_{k\geq 1, m\geq 0}$ and $\{D_m\}_{m\geq 0}$, and a random variable $K \in L^1(\mathbb{P}_{\theta^*})$ such that, for all $1 \leq k \leq n$ and $m' \geq m \geq 0$,*

$$(a) \quad \sup_{\theta\in G} \sup_{x\in\mathcal{X}} |\Gamma_{k,m,x}(\theta) - \Gamma_{k,m}(\theta)| \leq K(k+m)^3 \rho^{\lfloor (k+m)/8(p+1) \rfloor} C_{k,m},$$

$$(b) \quad \sup_{\theta\in G} \sup_{x,x'\in\mathcal{X}} \left|\Gamma_{k,m,x}(\theta) - \Gamma_{k,m',x'}(\theta)\right| \leq K[(k+m)^3 + D_m] \rho^{\lfloor (k+m)/16(p+1) \rfloor} C_{k,m},$$

*where $\mathbb{P}_{\theta^*}(C_{k,m} \geq 1 \text{ i.o.}) = 0$, $D_m < \infty$ $\mathbb{P}_{\theta^*}$-a.s. and the distribution function of $D_m$ does not depend on $m$.*

Lemma 7 implies that $\{\Gamma_{k,m,x}(\theta)\}_{m\geq 0}$ converges to $\Gamma_{k,\infty}(\theta)$ in probability uniformly in $x \in \mathcal{X}$ and $\theta \in G$. The following proposition is a local uniform law of large numbers for the observed Hessian.

**Proposition 3.** *Assume Assumptions 1–8. Then,*

$$\sup_{x \in \mathcal{X}} \left| n^{-1} \nabla_\theta^2 l_n(\theta, x) - \mathbb{E}_{\theta^*}[\Psi_{0,\infty}^2(\theta) + \Gamma_{0,\infty}(\theta)] \right| \to_p 0.$$

The following proposition shows the asymptotic normality of the MLE.

**Proposition 4.** *Assume Assumptions 1–8. Then, (a) for any $x_0 \in \mathcal{X}$, $n^{-1/2}(\hat{\theta}_{x_0} - \theta^*) \to_d N(0, I(\theta^*)^{-1})$; (b) for any probability measure $\xi$ on $\mathcal{B}(\mathcal{X})$ for $x_0$, $n^{-1/2}(\hat{\theta}_\xi - \theta^*) \to_d N(0, I(\theta^*)^{-1})$.*

### 5.3 Convergence of the covariance matrix estimate

When conducting statistical inferences with the MLE, the researcher needs to estimate the asymptotic covariance matrix of the MLE. Proposition 3 already derived the consistency of the observed Hessian. We derive the consistency of the outer-product-of-gradients estimates:

$$\hat{I}_{x_0}(\theta) := n^{-1} \sum_{k=1}^{n} \nabla_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0)(\nabla_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0))', \tag{15}$$

$$\hat{I}_\xi(\theta) := n^{-1} \sum_{k=1}^{n} \nabla_\theta \log p_{\theta\xi}(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)(\nabla_\theta \log p_{\theta\xi}(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k))', \tag{16}$$

where $\nabla_\theta \log p_{\theta\xi}(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k) := \nabla_\theta \log \int p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0)\xi(dx_0)$. In applications, $\nabla_\theta \log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0)$ can be computed by numerically differentiating $\log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0)$, which in turn can be computed by using the recursive algorithm of Hamilton (1996).

The following proposition shows the consistency of the outer-product-of-gradients estimate. Its proof is similar to that of Proposition 3 and hence omitted.

**Proposition 5.** *Assume Assumptions 1–8. Then,* $\sup_{x_0 \in \mathcal{X}} |\hat{I}_{x_0}(\hat{\theta}) - I(\theta^*)| \to_p 0$ *and* $\hat{I}_\xi(\hat{\theta}) \to_p I(\theta^*)$ *for any $\hat{\theta}$ such that $\hat{\theta} \to_p \theta^*$ and any $\xi$.*

## 6 Simulation

As an illustration, we provide a small simulation study based on the Markov regime switching model (3). The simulation was conducted with an R package we developed for Markov regime switching models.[6] We generate 1000 data sets of sample sizes $n = 200, 400$, and 800 from model (3) with $p = 5$, using the parameter value estimated by Hamilton (1989) for U.S. real GNP growth from 1952Q2 to 1984Q4. Specifically, the true parameter value of our simulated data is taken from Table I of Hamilton (1989) with $\theta = (\mu_1, \mu_2, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \sigma, p_{11}, p_{22})' = (1.522, -0.3577, 0.014, -0.058, -0.247, -0.213, 0.7690, 0.9049, 0.7550)'$.[7] For each of the 1000 data

[6] The R package is available at `https://github.com/chiyahn/rMSWITCH`.

[7] We simulate $(800 + n)$ periods and use the last $n$ observations as our sample, so that the initial value for our data set is approximately drawn from the stationary distribution.

sets, we estimate the parameter $\theta$ by the MLE, where we also treat the initial distribution of $X_0$, $\xi = (\Pr(X_0 = 1), \Pr(X_0 = 2))'$, as a parameter to be estimated together with $\theta$. Given the estimate, we construct the asymptotic 95 percent confidence interval by using the outer-product-of-gradients estimator (16), while fixing $\xi$ at its estimated value, and then examine the frequency at which the 95 percent confidence interval contains the true parameter value. Table 1 presents the results. At $n = 200$, the asymptotic 95 percent confidence intervals slightly undercover the true parameter with the coverage probability ranging from 0.875 to 0.944. As the sample size increases from $n = 200$ to 400, and then to 800, the actual coverage probability approaches 95 percent, indicating that the asymptotic approximation improves as the sample size increases. Table 2 presents the coverage probabilities of the asymptotic 95 percent confidence intervals when we use the outer-product-of-gradients estimator (15) by setting $x_0 = 2$ rather than (16). Consistent with our theoretical derivation, the results in Table 2 are similar to those in Table 1, suggesting that the choice of the initial value of $x_0$ in constructing the covariance matrix estimate does not affect the performance of the asymptotic confidence intervals.

Table 1: Coverage probability of the asymptotic 95 percent confidence intervals with $\hat{\xi}$

| | $p_{11}$ | $p_{21}$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\mu_1$ | $\mu_2$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|---|
| $n = 200$ | 0.916 | 0.911 | 0.938 | 0.926 | 0.944 | 0.925 | 0.916 | 0.896 | 0.875 |
| $n = 400$ | 0.938 | 0.933 | 0.930 | 0.944 | 0.943 | 0.937 | 0.946 | 0.929 | 0.922 |
| $n = 800$ | 0.942 | 0.942 | 0.945 | 0.941 | 0.950 | 0.956 | 0.939 | 0.941 | 0.930 |

Notes: Based on 1000 replications. Each entry reports the frequency at which the asymptotic 95 percent confidence interval constructed from (16) contains the true parameter value.

Table 2: Coverage probability of the asymptotic 95 percent confidence intervals with $x_0 = 2$

| | $p_{11}$ | $p_{21}$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\mu_1$ | $\mu_2$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|---|
| $n = 200$ | 0.915 | 0.920 | 0.938 | 0.927 | 0.941 | 0.934 | 0.922 | 0.901 | 0.884 |
| $n = 400$ | 0.932 | 0.932 | 0.938 | 0.949 | 0.942 | 0.939 | 0.945 | 0.929 | 0.923 |
| $n = 800$ | 0.943 | 0.945 | 0.945 | 0.939 | 0.949 | 0.956 | 0.936 | 0.937 | 0.929 |

Notes: Based on 1000 replications. Each entry reports the frequency at which the asymptotic 95 percent confidence interval constructed from (15) with $x_0 = 2$ contains the true parameter value.

# 7 Proofs

Throughout these proofs, define $\overline{\mathbf{V}}_b^a := (\overline{\mathbf{Y}}_b^a, \mathbf{W}_b^a)$.

*Proof of Lemma 1.* The proof uses a similar argument to the proof of Lemma 1 in DMR. Because

$\{Z_k\}_{k=-m}^n$ is a Markov chain given $\{W_k\}_{k=-m}^n$, we have, for $-m < k \leq n$,

$$\mathbb{P}_\theta(X_k \in A | \mathbf{X}_{-m}^{k-1}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n) = \mathbb{P}_\theta(X_k \in A | X_{k-1}, \overline{\mathbf{Y}}_{k-1}^n, \mathbf{W}_k^n).$$

Therefore, $\{X_k\}_{k=-m}^n$ conditional on $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$ is an inhomogeneous Markov chain, and part (a) follows.

We proceed to prove part (b). Observe that if $-m+p \leq k \leq n$,

$$\mathbb{P}_\theta(X_k \in A | X_{k-p}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n) = \mathbb{P}_\theta(X_k \in A | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^n, \mathbf{W}_{k-p}^n), \tag{17}$$

because the left hand side of (17) can be written as

$$\frac{\mathbb{P}_\theta(X_k \in A, \mathbf{Y}_{k-p+1}^n | X_{k-p}, \overline{\mathbf{Y}}_{-m}^{k-p}, \mathbf{W}_{-m}^n)}{\mathbb{P}_\theta(\mathbf{Y}_{k-p+1}^n | X_{k-p}, \overline{\mathbf{Y}}_{-m}^{k-p}, \mathbf{W}_{-m}^n)} = \frac{\mathbb{P}_\theta(X_k \in A, \mathbf{Y}_{k-p+1}^n | X_{k-p}, \overline{\mathbf{Y}}_{k-p}, \mathbf{W}_{k-p}^n)}{\mathbb{P}_\theta(\mathbf{Y}_{k-p+1}^n | X_{k-p}, \overline{\mathbf{Y}}_{k-p}, \mathbf{W}_{k-p}^n)}.$$

The equality (17) holds even when the conditioning variable $\mathbf{W}_{k-p}^n$ on the right hand side is replaced with $\mathbf{W}_{k-p+1}^n$, but we use $\mathbf{W}_{k-p}^n$ for notational simplicity. Write the right hand side of (17) as

$$\begin{aligned}
&\mathbb{P}_\theta(X_k \in A | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^n, \mathbf{W}_{k-p}^n) \\
&= \frac{\int_A p_\theta(X_k = x, \mathbf{Y}_k^n | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^n)\mu(dx)}{p_\theta(\mathbf{Y}_k^n | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^n)} \\
&= \int_A p_\theta(X_k = x | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) p_\theta(\mathbf{Y}_k^n | X_k = x, \overline{\mathbf{Y}}_{k-1}, \mathbf{W}_k^n)\mu(dx) \\
&\times \left( \int_{\mathcal{X}} p_\theta(X_k = x | X_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) p_\theta(\mathbf{Y}_k^n | X_k = x, \overline{\mathbf{Y}}_{k-1}, \mathbf{W}_k^n)\mu(dx) \right)^{-1}.
\end{aligned}$$

When $p = 1$, we have $p_\theta(x_k | x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) = p_\theta(x_k | x_{k-1}) \in [\sigma_-, \sigma_+]$. Therefore, the stated result follows with $\mu_k(\overline{\mathbf{Y}}_{k-1}^n, \mathbf{W}_k^n, A)$ defined as

$$\mu_k(\overline{\mathbf{Y}}_{k-1}^n, \mathbf{W}_k^n, A) := \int_A p_\theta(\mathbf{Y}_k^n | X_k = x, \overline{\mathbf{Y}}_{k-1}, \mathbf{W}_k^n)\mu(dx) \Big/ \int_{\mathcal{X}} p_\theta(\mathbf{Y}_k^n | X_k = x, \overline{\mathbf{Y}}_{k-1}, \mathbf{W}_k^n)\mu(dx). \tag{18}$$

Note that $\int_{\mathcal{X}} p_\theta(\mathbf{Y}_k^n | X_k = x, \overline{\mathbf{Y}}_{k-1}, \mathbf{W}_k^n)\mu(dx) > 0$ from Assumption 3.

When $p \geq 2$, a lower bound on $p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is obtained as

$$
\begin{aligned}
&p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \\
&= \frac{p_\theta(x_k, \mathbf{Y}_{k-p+1}^{k-1}|x_{k-p}, \overline{\mathbf{Y}}_{k-p}, \mathbf{W}_{k-p}^{k-1})}{p_\theta(\mathbf{Y}_{k-p+1}^{k-1}|x_{k-p}, \overline{\mathbf{Y}}_{k-p}, \mathbf{W}_{k-p}^{k-1})} \\
&= \frac{\int \prod_{i=k-p+1}^{k} q_\theta(x_{i-1}, x_i) \prod_{i=k-p+1}^{k-1} g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1}, x_i, W_i) \mu^{\otimes(p-1)}(d\mathbf{x}_{k-p+1}^{k-1})}{\int \prod_{i=k-p+1}^{k} q_\theta(x_{i-1}, x_i) \prod_{i=k-p+1}^{k-1} g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1}, x_i, W_i) \mu^{\otimes p}(d\mathbf{x}_{k-p+1}^{k})} \\
&\geq \frac{\inf_\theta \inf_{x_k, x_{k-p}} p_\theta(x_k|x_{k-p}) \inf_\theta \inf_{\mathbf{x}_{k-p+1}^{k-1}} \prod_{i=k-p+1}^{k-1} g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1}, x_i, W_i)}{\sup_\theta \sup_{x_k, x_{k-p}} p_\theta(x_k|x_{k-p}) \sup_\theta \sup_{\mathbf{x}_{k-p+1}^{k-1}} \prod_{i=k-p+1}^{k-1} g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1}, x_i, W_i)}. 
\end{aligned}
\tag{19}
$$

Similarly, an upper bound on $p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1})$ is given by the reciprocal of (19). Therefore, the stated result holds with $\mu_k(\overline{\mathbf{Y}}_{k-1}^{n}, \mathbf{W}_k^n, A)$ defined in (18). $\square$

*Proof of Lemma 2.* In view of (7), the stated result holds if there exist constants $\rho \in (0,1)$ and $M < \infty$ and a random sequence $\{b_k\}$ with $\mathbb{P}_{\theta^*}(b_k \geq M \text{ i.o.}) = 0$ such that, for $k = 1, \ldots, n$,

$$
\sup_{x_0 \in \mathcal{X}} \sup_{\theta \in \Theta} \left| \log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0) - \log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k) \right| \leq \min\left\{ \frac{b_+}{b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k)}, \rho^{\lfloor k/3p \rfloor} b_k \right\},
\tag{20}
$$

because $b_+/b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k) < \infty$ $\mathbb{P}_{\theta^*}$-a.s. from Assumption 3.

First, it follows from $p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0) = \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) \mathbb{P}_\theta(dx_k|x_0, \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)$, $p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k) = \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) \mathbb{P}_\theta(dx_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)$, and Assumption 4(a) that $p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0), p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k) \in [b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k), b_+]$ uniformly in $\theta \in \Theta$ and $x_0 \in \mathcal{X}$. Hence, from the inequality $|\log x - \log y| \leq |x - y|/(x \wedge y)$, we have, for $k = 1, \ldots, n$,

$$
\sup_{x_0 \in \mathcal{X}} \sup_{\theta \in \Theta} |\log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0) - \log p_\theta(Y_k|\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)| \leq b_+/b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k).
\tag{21}
$$

This gives the first bound in (20).

We proceed to derive the second bound in (20). Using a derivation similar to (17) and noting that $X_k$ is independent of $W_k$ given $X_{k-1}$ gives, for any $-m + p \leq k \leq n$,

$$
\mathbb{P}_\theta(X_k \in \cdot|X_{k-p}, \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k}) = \mathbb{P}_\theta(X_k \in \cdot|X_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}).
\tag{22}
$$

Consequently, for any $-m+p \le k \le n$,

$$
\begin{aligned}
&p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k}, x_{-m}) \\
&= \int\int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mathbb{P}_\theta(dx_{k-p}|x_{-m}, \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1}) \mu(dx_k), \quad (23)\\
&p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k}) \\
&= \int\int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mathbb{P}_\theta(dx_{k-p}|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1}) \mu(dx_k). \quad (24)
\end{aligned}
$$

Furthermore,

$$
\mathbb{P}_\theta(X_{k-p} \in \cdot|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1}) = \int \mathbb{P}_\theta(X_{k-p} \in \cdot|x_{-m}, \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1}) \mathbb{P}_\theta(dx_{-m}|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^{k-1}). \quad (25)
$$

Combining (23), (24), and (25) for $m = 0$ and applying Corollary 1 and the property of the total variation distance gives that, for any $p \le k \le n$ and uniformly in $x_0 \in \mathcal{X}$,

$$
\begin{aligned}
&\left| p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}, x_0) - p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}) \right| \\
&\le \left| \int\int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mu(dx_k) \right. \\
&\quad \left. \times \left( \mathbb{P}_\theta(dx_{k-p}|x_0, \overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k-1}) - \mathbb{P}_\theta(dx_{k-p}|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k-1}) \right) \right| \\
&\le \prod_{i=1}^{\lfloor (k-p)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}) \right) \sup_{x_{k-p}} \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) p_\theta(x_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mu(dx_k) \\
&\le \prod_{i=1}^{\lfloor (k-p)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}) \right) \sup_{x'_k, x_{k-p} \in \mathcal{X}} p_\theta(x'_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k) \mu(dx_k).
\end{aligned}
\quad (26)
$$

Furthermore, (23) and (24) imply that, for any $k \ge p$, $(p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}, x_0) \wedge p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}))$ $\ge \inf_{x'_k, x_{k-p} \in \mathcal{X}} p_\theta(x'_k|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k)\mu(dx_k)$. Therefore, it follows from $|\log x - \log y| \le |x-y|/(x \wedge y)$, (26), and (19) and the subsequent argument that, for $p \le k \le n$,

$$
\sup_{x_0 \in \mathcal{X}} \sup_{\theta \in \Theta} \left| \log p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}, x_0) - \log p_\theta(Y_k|\overline{\mathbf{Y}}_{0}^{k-1}, \mathbf{W}_{0}^{k}) \right| \le \frac{\prod_{i=1}^{\lfloor (k-p)/p \rfloor} \left( 1 - \omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}) \right)}{\omega(\overline{\mathbf{V}}_{k-p}^{k-1})}. \quad (27)
$$

We first bound $\prod_{i=1}^{\lfloor (k-p)/p \rfloor}(1-\omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}))$ on the right hand side of (27). Fix $\epsilon \in (0, 1/8]$. Because $\omega(\overline{\mathbf{V}}_{t-p}^{t-1}) > 0$ for all $\overline{\mathbf{V}}_{t-p}^{t-1} \in \mathcal{Y}^{p+s-1} \times \mathcal{W}^p$ from Assumption 3 (note that $\omega(\overline{\mathbf{V}}_{t-p}^{t-1}) = \sigma_-/\sigma_+ > 0$ when $p = 1$), there exists $\rho \in (0,1)$ such that $\mathbb{P}_{\theta^*}(1-\omega(\overline{\mathbf{V}}_{t-p}^{t-1}) \ge \rho) \le \epsilon$. Define $I_i := \mathbb{I}\{1-\omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}) \ge \rho\}$;

then, we have $\mathbb{E}_{\theta^*}[I_i] \leq \epsilon$ and $1 - \omega(\overline{\mathbf{V}}_{pi-p}^{pi-1}) \leq \rho^{1-I_i}$. Consequently, with $a_k := \rho^{-\sum_{i=1}^{\lfloor (k-p)/p \rfloor} I_i}$,

$$\prod_{i=1}^{\lfloor (k-p)/p \rfloor} \left(1 - \omega(\overline{\mathbf{V}}_{pi-p}^{pi-1})\right) \leq \rho^{\lfloor (k-p)/p \rfloor - \sum_{i=1}^{\lfloor (k-p)/p \rfloor} I_i} = \rho^{\lfloor (k-p)/p \rfloor} a_k. \tag{28}$$

Because $\overline{\mathbf{V}}_{t-p}^{t-1}$ is stationary and ergodic, it follows from the strong law of large numbers that $(\lfloor (k-p)/p \rfloor)^{-1} \sum_{i=1}^{\lfloor (k-p)/p \rfloor} I_i \to \mathbb{E}_{\theta^*}[I_i] \leq \epsilon$ $\mathbb{P}_{\theta^*}$-a.s. as $k \to \infty$. Therefore, $a_k$ is bounded as

$$\mathbb{P}_{\theta^*}(a_k \geq \rho^{-2\epsilon \lfloor (k-p)/p \rfloor} \text{ i.o.}) = 0. \tag{29}$$

We then bound $1/\omega(\overline{\mathbf{V}}_{k-p}^{k-1})$ on the right hand side of (27). Let $C_3 := (\sigma_-/\sigma_+)^2 (C_1/b_+)^{2(p-1)} > 0$; then, we have $\mathbb{P}_{\theta^*}(\omega(\overline{\mathbf{V}}_{k-p}^{k-1}) \leq C_3 e^{-2\alpha(p-1)r}) \leq (p-1)\mathbb{P}_{\theta^*}(b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k) \leq C_1 e^{-\alpha r})$ for any $r > 0$. In view of $\rho \in (0,1)$, there exists a finite and positive constant $C_4$ such that $\rho^\epsilon = e^{-2\alpha(p-1)C_4}$. For $k \geq 2p$, set $r = C_4 \lfloor (k-p)/p \rfloor > 0$ so that $\rho^{\epsilon \lfloor (k-p)/p \rfloor} = e^{-2\alpha(p-1)r}$. Then, $\mathbb{P}_{\theta^*}(\omega(\overline{\mathbf{V}}_{k-p}^{k-1}) \leq C_3 \rho^{\epsilon \lfloor (k-p)/p \rfloor}) \leq (p-1)\mathbb{P}_{\theta^*}(b_-(\overline{\mathbf{Y}}_{k-1}^{k}, W_k) \leq C_1 e^{-\alpha C_4 \lfloor (k-p)/p \rfloor})$ for $k \geq 2p$, and it follows from Assumption 5 that $\sum_{k=p}^{\infty} \mathbb{P}_{\theta^*}(\omega(\overline{\mathbf{V}}_{k-p}^{k-1}) \leq C_3 \rho^{\epsilon \lfloor (k-p)/p \rfloor}) < \infty$. Therefore, $\mathbb{P}_{\theta^*}(\omega(\overline{\mathbf{V}}_{k-p}^{k-1}) \leq C_3 \rho^{\epsilon \lfloor (k-p)/p \rfloor} \text{ i.o.}) = 0$ from the Borel-Cantelli lemma. Substituting this bound and (28) and (29) into (27) gives, for $p \leq k \leq n$,

$$\sup_{x_0 \in \mathcal{X}} \sup_{\theta \in \Theta} \left| \log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0) - \log p_\theta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k) \right| \leq \rho^{(1-3\epsilon)\lfloor (k-p)/p \rfloor} b_k, \tag{30}$$

where $\mathbb{P}_{\theta^*}(b_k \geq M \text{ i.o.}) = 0$ for a constant $M < \infty$.

The right hand side of (30) gives the second bound in (20) because $(1-3\epsilon)\lfloor (k-p)/p \rfloor \geq \lfloor (k-p)/p \rfloor / 2 \geq \lfloor (k-p)/2p \rfloor \geq \lfloor k/3p \rfloor$, where the last inequality holds because, for any numbers $a, b > 0$ and $k \geq 0$,

$$\lfloor (k-a)_+/b \rfloor \geq \lfloor k/(a+b) \rfloor. \tag{31}$$

Therefore, (20) holds, and the stated result is proven.

□

*Proof of Lemma 3.* The proof uses a similar argument to the proof of Lemma 3 in DMR and the proof of Lemma 2. We first show part (a) for $-m + p \leq k \leq n$. Using a similar argument to (23) and (26) in conjunction with Corollary 1 gives

$$\begin{aligned} & p_\theta(Y_k | \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m} = x) - p_\theta(Y_k | \overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k, X_{-m'} = x') \\ &= \int \int \int g_\theta(Y_k | \overline{\mathbf{Y}}_{k-1}, x_k, W_k) p_\theta(x_k | x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \mu(dx_k) \\ & \quad \times \mathbb{P}_\theta(dx_{k-p} | X_{-m} = x_{-m}, \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k) \left[ \delta_x(dx_{-m}) - \mathbb{P}_\theta(dx_{-m} | X_{-m'} = x', \overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k) \right] \quad (32) \\ &\leq \prod_{i=1}^{\lfloor (k-p+m)/p \rfloor} \left(1 - \omega(\overline{\mathbf{V}}_{-m+pi-p}^{-m+pi-1})\right) \sup_{x'_k, x_{k-p} \in \mathcal{X}} p_\theta(x'_k | x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \int g_\theta(Y_k | \overline{\mathbf{Y}}_{k-1}, x_k, W_k) \mu(dx_k), \end{aligned}$$

where the first equality uses the fact $\mathbb{P}_\theta(X_{k-p} \in \cdot|X_{-m}, \overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k)$ $= \mathbb{P}_\theta(X_{k-p} \in \cdot|X_{-m}, \overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k)$, which is proven as (22).

Furthermore, (23) and (24) imply that, for any $k \geq -m+p$, $(p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, x_{-m}) \wedge p_\theta(Y_k|\overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k, x_{-m'})) \geq \inf_{x_k', x_{k-p} \in \mathcal{X}} p_\theta(x_k'|x_{k-p}, \overline{\mathbf{Y}}_{k-p}^{k-1}, \mathbf{W}_{k-p}^{k-1}) \int g_\theta(Y_k|\overline{\mathbf{Y}}_{k-1}, x_k, W_k)\mu(dx_k)$. Therefore, it follows from the inequality $|\log x - \log y| \leq |x-y|/(x \wedge y)$ that

$$\begin{aligned} &\left|\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m} = x) - \log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k, X_{-m'} = x')\right| \\ &\leq \frac{\prod_{i=1}^{\lfloor (k-p+m)/p \rfloor} \left(1 - \omega(\overline{\mathbf{V}}_{-m+pi-p}^{-m+pi-1})\right)}{\omega(\overline{\mathbf{V}}_{k-p}^{k-1})}. \end{aligned} \tag{33}$$

Proceeding as in (28)–(30) in the proof of Lemma 2, we find that there exist $\rho \in (0,1)$ and $\epsilon \in (0, 1/8]$ such that the right hand side of (33) is bounded by $\rho^{(1-2\epsilon)\lfloor (k-p+m)/p \rfloor}\rho^{-\epsilon\lfloor (k-p)/p \rfloor}B_{k,m}$, where $\mathbb{P}_{\theta^*}(B_{k,m} \geq M \text{ i.o.}) = 0$ for a constant $M < \infty$. Therefore, part (a) is proven for $-m+p \leq k \leq n$ by noting that $\rho^{-\epsilon\lfloor (k-p)/p \rfloor} \leq \rho^{-\epsilon\lfloor (k-p+m)/p \rfloor}$ and using the argument following (30). Part (a) holds for $1 \leq k \leq -m+p-1$ because $|\log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m} = x) - \log p_\theta(Y_k|\overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k, X_{-m'} = x')|$ is bounded by $b_+/b_-(\overline{\mathbf{Y}}_{k-1}^k, W_k)$, which is finite $\mathbb{P}_{\theta^*}$-a.s. Part (b) follows from replacing $\mathbb{P}_\theta(dx_{-m}|X_{-m'} = x', \overline{\mathbf{Y}}_{-m'}^{k-1}, \mathbf{W}_{-m'}^k)$ in (32) with $\mathbb{P}_\theta(dx_{-m}|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k)$. Part (c) follows from $b_-(\overline{\mathbf{Y}}_{k-1}^k, W_k) \leq p_\theta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m} = x) \leq b_+$ and Assumption 4. $\square$

*Proof of Proposition 1.* The proof follows the argument of the proof of Proposition 2 and Theorem 1 in DMR. From Property 24.2 of Gourieroux and Monfort (1995, page 385), the stated result holds if (i) $\Theta$ is compact, (ii) $l_n(\theta, x_0)$ is continuous uniformly in $x_0 \in \mathcal{X}$, (iii) $\sup_{x_0 \in \mathcal{X}} \sup_{\theta \in \Theta} |n^{-1}l_n(\theta, x_0) - l(\theta)| \to 0$ $\mathbb{P}_{\theta^*}$-a.s., and (iv) $l(\theta)$ is uniquely maximized at $\theta^*$.

(i) follows from Assumption 1(a). (ii) follows from Assumption 6(a). In view of Lemma 2 and the compactness of $\Theta$, (iii) holds if, for all $\theta \in \Theta$,

$$\limsup_{\delta \to 0} \limsup_{n \to \infty} \sup_{|\theta' - \theta| \leq \delta} |n^{-1}l_n(\theta') - l(\theta)| = 0 \quad \mathbb{P}_{\theta^*}\text{-a.s.} \tag{34}$$

Noting that $l_n(\theta) = \sum_{k=1}^n \Delta_{k,0}(\theta)$, the left hand side of (34) is bounded by $A + B + C$, where

$$\begin{aligned} A &:= \limsup_{n \to \infty} \sup_{\theta' \in \Theta} \left| n^{-1} \sum_{k=1}^n (\Delta_{k,0}(\theta') - \Delta_{k,\infty}(\theta')) \right|, \\ B &:= \limsup_{\delta \to 0} \limsup_{n \to \infty} \sup_{|\theta' - \theta| \leq \delta} \left| n^{-1} \sum_{k=1}^n (\Delta_{k,\infty}(\theta') - \Delta_{k,\infty}(\theta)) \right|, \\ C &:= \limsup_{n \to \infty} \left| n^{-1} \sum_{k=1}^n (\Delta_{k,\infty}(\theta) - \mathbb{E}_{\theta^*}\Delta_{k,\infty}(\theta)) \right|. \end{aligned}$$

Fix $x \in \mathcal{X}$. Setting $m = 0$ and letting $m' \to \infty$ in Lemma 3(a)(b) show that $\sup_{\theta \in \Theta} |\Delta_{k,0}(\theta) - \Delta_{k,\infty}(\theta)| \leq \sup_{\theta \in \Theta} |\Delta_{k,0}(\theta) - \Delta_{k,0,x}(\theta)| + \sup_{\theta \in \Theta} |\Delta_{k,0,x}(\theta) - \Delta_{k,\infty}(\theta)| \leq 2A_{k,0}\rho^{\lfloor k/3p \rfloor}$ while $\sup_{\theta \in \Theta} |\Delta_{k,0}(\theta) -$

$\Delta_{k,0,x}(\theta)| + \sup_{\theta\in\Theta} |\Delta_{k,0,x}(\theta) - \Delta_{k,\infty}(\theta)| \leq 4B_k$ follows from Lemma 3(c). Consequently, $A = 0$ $\mathbb{P}_{\theta^*}$-a.s. $B$ is bounded by, from the ergodic theorem and Lemma 9,

$$\begin{aligned}
&\lim_{\delta\to 0}\limsup_{n\to\infty} n^{-1}\sum_{k=1}^{n}\sup_{|\theta'-\theta|\leq\delta}|\Delta_{k,\infty}(\theta') - \Delta_{k,\infty}(\theta)| \\
&= \lim_{\delta\to 0}\mathbb{E}_{\theta^*}\left[\sup_{|\theta'-\theta|\leq\delta}|\Delta_{0,\infty}(\theta') - \Delta_{0,\infty}(\theta)|\right] = 0 \quad \mathbb{P}_{\theta^*}\text{-a.s.}
\end{aligned}$$

$C = 0$ $\mathbb{P}_{\theta^*}$-a.s. by the ergodic theorem, and hence (iii) holds. For (iv), observe that $\mathbb{E}_{\theta^*}|\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)| < \infty$ from Lemma 3(c). Therefore, for any $m$, $\mathbb{E}_{\theta^*}[\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)]$ is uniquely maximized at $\theta^*$ from Lemma 2.2 of Newey and McFadden (1994) and Assumption 6(b). Then, (iv) follows because $\mathbb{E}_{\theta^*}[\log p_\theta(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)]$ converges to $l(\theta)$ uniformly in $\theta$ as $m \to \infty$ from Lemma 3 and the dominated convergence theorem. Therefore, (iv) holds, and the stated result is proven. □

*Proof of Corollary 2.* Observe that $|n^{-1}l_n(\theta,\xi) - l(\theta)| \leq \sup_{x_0\in\mathcal{X}} |n^{-1}l_n(\theta,x_0) - l(\theta)|$ because $\inf_{x_0\in\mathcal{X}} l_n(\theta,x_0) \leq l_n(\theta,\xi) \leq \sup_{x_0\in\mathcal{X}} l_n(\theta,x_0)$. Furthermore, $l_n(\theta,\xi)$ is continuous in $\theta$ from the continuity of $l_n(\theta,x_0)$. Therefore, the stated result follows from the proof of Proposition 1. □

*Proof of Lemma 4.* The proof is similar to the proof of Lemma 1. Because the time-reversed process $\{Z_{n-k}\}_{0\leq k\leq n+m}$ is Markov conditional on $\mathbf{W}_{-m}^n$, we have, for $1 \leq k \leq n+m$,

$$\mathbb{P}_\theta(X_{n-k} \in A|\mathbf{X}_{n-k+1}^n, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n) = \mathbb{P}_\theta(X_{n-k} \in A|X_{n-k+1}, \overline{\mathbf{Y}}_{-m}^{n-k+1}, \mathbf{W}_{-m}^n).$$

Therefore, $\{X_{n-k}\}_{0\leq k\leq n+m}$ is an inhomogeneous Markov chain given $(\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$, and part (a) follows.

For part (b), because (i) the time-reversed process $\{Z_{n-k}\}_{0\leq k\leq n+m}$ is Markov conditional on $\mathbf{W}_{-m}^n$, (ii) $Y_{n-k+p}$ is independent of $\mathbf{X}_{-m}^{n-k+p-1}$ given $(X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^n)$, (iii) $X_{n-k+p}$ is independent of the other random variables given $X_{n-k+p-1}$, and (iv) $W_{n-k+p}$ is independent of $\mathbf{Z}_{-m}^{n-k+p-1}$ given $\mathbf{W}_{-m}^{n-k+p-1}$, we have, for $1 \leq k \leq n+m$,

$$\mathbb{P}_\theta\left(X_{n-k} \in A\middle|X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n\right) = \mathbb{P}_\theta\left(X_{n-k} \in A\middle|X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^{n-k+p-1}\right). \tag{35}$$

Observe that in view of $n-k \geq -m$,

$$\begin{aligned}
&\mathbb{P}_\theta\left(X_{n-k} \in A, X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^{n-k+p-1}\right) \\
&= \mathbb{P}_\theta\left(X_{n-k+p}\middle|X_{n-k} \in A, \overline{\mathbf{Y}}_{n-k}^{n-k+p-1}, \mathbf{W}_{n-k}^{n-k+p-1}\right) \\
&\quad\times \mathbb{P}_\theta\left(X_{n-k} \in A, \mathbf{Y}_{-m+1}^{n-k+p-1}, \mathbf{W}_{-m+1}^{n-k+p-1}\middle|\overline{\mathbf{Y}}_{-m}, W_{-m}\right)\mathbb{P}_\theta\left(\overline{\mathbf{Y}}_{-m}, W_{-m}\right).
\end{aligned}$$

It follows that

$$\mathbb{P}_\theta\left(X_{n-k}\in A\middle|X_{n-k+p},\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1}\right)=\frac{\int_A G_\theta(x,X_{n-k+p},\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1})\mu(dx)}{\int_{\mathcal{X}} G_\theta(x,X_{n-k+p},\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1})\mu(dx)},$$

where $G_\theta(x,X_{n-k+p},\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1}) := p_\theta(X_{n-k+p}|X_{n-k}=x,\overline{\mathbf{Y}}_{n-k}^{n-k+p-1},\mathbf{W}_{n-k}^{n-k+p-1})\times p_\theta(X_{n-k}=x,\mathbf{Y}_{-m+1}^{n-k+p-1},\mathbf{W}_{-m+1}^{n-k+p-1}|\overline{\mathbf{Y}}_{-m},W_{-m})$.

When $p=1$, we have $p_\theta(X_{n-k+p}|X_{n-k}=x,\overline{\mathbf{Y}}_{n-k}^{n-k+p-1},\mathbf{W}_{n-k}^{n-k+p-1})=p_\theta(X_{n-k+1}|X_{n-k}=x)\in[\sigma_-,\sigma_+]$. Therefore, the stated result follows with $\tilde{\mu}_k(\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1},A)$ defined as

$$\tilde{\mu}_k(\overline{\mathbf{Y}}_{-m}^{n-k+p-1},\mathbf{W}_{-m}^{n-k+p-1},A):=\frac{\int_A p_\theta(X_{n-k}=x,\mathbf{Y}_{-m+1}^{n-k+p-1},\mathbf{W}_{-m+1}^{n-k+p-1}|\overline{\mathbf{Y}}_{-m},W_{-m})\mu(dx)}{\int_{\mathcal{X}} p_\theta(X_{n-k}=x,\mathbf{Y}_{-m+1}^{n-k+p-1},\mathbf{W}_{-m+1}^{n-k+p-1}|\overline{\mathbf{Y}}_{-m},W_{-m})\mu(dx)}. \quad (36)$$

Note that $\int_{\mathcal{X}} p_\theta(X_{n-k}=x,\mathbf{Y}_{-m+1}^{n-k+p-1},\mathbf{W}_{-m+1}^{n-k+p-1}|\overline{\mathbf{Y}}_{-m},W_{-m})\mu(dx)>0$ from Assumption 3.

When $p\geq 2$, it follows from a derivation similar to (19) that $p_\theta(x_{n-k+p}|x_{n-k},\overline{\mathbf{Y}}_{n-k}^{n-k+p-1},\mathbf{W}_{n-k}^{n-k+p-1})$ is bounded from below by

$$\frac{\inf_\theta\inf_{x_{n-k+p},x_{n-k}}p_\theta(x_{n-k+p}|x_{n-k})\inf_\theta\inf_{\mathbf{x}_{n-k+1}^{n-k+p-1}}\prod_{i=n-k+1}^{n-k+p-1}g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1},x_i,W_i)}{\sup_\theta\sup_{x_{n-k+p},x_{n-k}}p_\theta(x_{n-k+p}|x_{n-k})\sup_\theta\sup_{\mathbf{x}_{n-k+1}^{n-k+p-1}}\prod_{i=n-k+1}^{n-k+p-1}g_\theta(Y_i|\overline{\mathbf{Y}}_{i-1},x_i,W_i)}, \quad (37)$$

and an upper bound on $p_\theta(x_{n-k+p}|x_{n-k},\overline{\mathbf{Y}}_{n-k}^{n-k+p-1},\mathbf{W}_{n-k}^{n-k+p-1})$ is given by the inverse of (19). Therefore, the stated result holds with $\tilde{\mu}_k$ defined in (36). □

*Proof of Lemma 5.* When $k\geq n-1$, the stated result holds trivially because $\prod_{i=1}^j a_i=1$ when $j<i$. We first show part (a) for $k\leq n-2$. Because the time-reversed process $\{Z_{n-k}\}_{0\leq k\leq n+m}$ is Markov conditional on $\mathbf{W}_{-m}^n$ and $W_n$ is independent of $Z_{n-1}$ given $W_{n-1}$, we have $\mathbb{P}_\theta(X_k\in\cdot|\overline{\mathbf{Y}}_{-m}^n,\mathbf{W}_{-m}^n)=\int\mathbb{P}_\theta(X_k\in\cdot|x_{n-1},\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1})\mathbb{P}_\theta(dx_{n-1}|\overline{\mathbf{Y}}_{-m}^n,\mathbf{W}_{-m}^n)$. Similarly, we obtain $\mathbb{P}_\theta(X_k\in\cdot|\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1})=\int\mathbb{P}_\theta(X_k\in\cdot|x_{n-1},\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1})\mathbb{P}_\theta(dx_{n-1}|\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1})$. It follows that

$$\begin{aligned}&\left|\mathbb{P}_\theta\left(X_k\in\cdot|\overline{\mathbf{Y}}_{-m}^n,\mathbf{W}_{-m}^n\right)-\mathbb{P}_\theta\left(X_k\in\cdot\middle|\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1}\right)\right|\\&\leq\int\mathbb{P}_\theta\left(X_k\in\cdot\middle|x_{n-1},\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1}\right)\left|\mathbb{P}_\theta\left(dx_{n-1}|\overline{\mathbf{Y}}_{-m}^n,\mathbf{W}_{-m}^n\right)-\mathbb{P}_\theta\left(dx_{n-1}\middle|\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1}\right)\right|.\end{aligned}$$

Therefore, the stated result follows from applying Lemmas 4 and 8 to the time-reversed process $\{X_{n-i}\}_{i=1}^{n-k}$ conditional on $(\overline{\mathbf{Y}}_{-m}^{n-1},\mathbf{W}_{-m}^{n-1})$.

For part (b) for $k\leq n-2$, by using a similar argument to the proof of Lemma 4, we can show that (i) conditionally on $(\overline{\mathbf{Y}}_{-m}^n,\mathbf{W}_{-m}^n,X_{-m})$, the time-reversed process $\{X_{n-k}\}_{0\leq k\leq n+m-1}$ is an inhomogeneous Markov chain, and (ii) for all $p\leq k\leq n+m-1$, there exists a probability measure

$\breve{\mu}_k(\overline{\mathbf{y}}_{-m}^{n-k+p-1}, \mathbf{w}_{-m}^{n-k+p-1}, x, A)$ such that, for all $(\overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^{n-k+p-1}, x)$,

$$\begin{aligned}
&\mathbb{P}_\theta \left(X_{n-k} \in A \middle| X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^{n}, \mathbf{W}_{-m}^{n}, X_{-m} = x\right) \\
&= \mathbb{P}_\theta \left(X_{n-k} \in A \middle| X_{n-k+p}, \overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^{n-k+p-1}, X_{-m} = x\right) \\
&\geq \omega(\overline{\mathbf{Y}}_{n-k}^{n-k+p-1}, \mathbf{W}_{n-k}^{n-k+p-1}) \breve{\mu}_k(\overline{\mathbf{Y}}_{-m}^{n-k+p-1}, \mathbf{W}_{-m}^{n-k+p-1}, X_{-m} = x, A),
\end{aligned}$$

with the same $\omega(\overline{\mathbf{Y}}_{n-k}^{n-k+p-1}, \mathbf{W}_{n-k}^{n-k+p-1})$ as in Lemma 4. Therefore, the stated result follows from a similar argument to the proof of part (a). $\square$

*Proof of Lemma 6.* The proof follows the argument of the proof of Lemma 13 in DMR. When $(k, m) = (1, 0)$, the stated result follows from $\Psi_{1,0,x}^j(\theta) = \mathbb{E}_{\theta^*}[\phi_{\theta 1}^j | \overline{\mathbf{V}}_0, X_0 = x]$, $\Psi_{1,0}^j(\theta) = \mathbb{E}_{\theta^*}[\phi_{\theta 1}^j | \overline{\mathbf{V}}_0]$, $\sup_{\theta \in G} |\phi_{\theta k}^j| \leq |\phi_k^j|_\infty$, and Assumption 7. Henceforth, assume $(k, m) \neq (1, 0)$ so that $k + m \geq 2$.

For part (a), it follows from Lemma 10(a)–(e) that

$$\begin{aligned}
\left| \Psi_{k,m,x}^j(\theta) - \Psi_{k,m}^j(\theta) \right| &\leq 4 \sum_{t=-m+1}^{k} |\phi_t^j|_\infty \left( \Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1} \right) \\
&\leq 4 \max_{-m \leq t' \leq k} |\phi_{t'}^j|_\infty \sum_{t=-m+1}^{k} \left( \Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1} \right), \qquad (38)
\end{aligned}$$

where $\Omega_{t-1,-m} := \prod_{i=1}^{\lfloor (t-1+m)/p \rfloor} (1 - \omega(\overline{\mathbf{V}}_{-m+pi-p}^{-m+pi-1}))$ and $\tilde{\Omega}_{t,k-1} := \prod_{i=1}^{\lfloor (k-1-t)/p \rfloor} (1 - \omega(\overline{\mathbf{V}}_{k-2-pi+1}^{k-2-pi+p}))$ as defined in the paragraph preceding Lemma 10. As shown on page 2294 of DMR, we have $\max_{-m \leq t' \leq k} |\phi_{t'}^j|_\infty \leq \sum_{t=-m}^{k} (|t| \vee 1)^2 |\phi_t^j|_\infty / (|t| \vee 1)^2 \leq 2(k \vee m)^2 [\sum_{t=-\infty}^{\infty} |\phi_t^j|_\infty / (|t| \vee 1)^2] \leq (k + m)^2 K_j$ with $K_j \in L^{3-j}(\mathbb{P}_{\theta^*})$.

We proceed to bound $\sum_{t=-m+1}^{k} (\Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1})$ on the right hand side of (38). Similar to the proof of Lemma 2, fix $\epsilon \in (0, 1/8p(p+1)]$; then, there exists $\rho \in (0, 1)$ such that $\mathbb{P}_{\theta^*}(1 - \omega(\overline{\mathbf{V}}_{k-p}^{k-1}) \geq \rho) \leq \epsilon$. Define $I_{p,i} := \sum_{t=0}^{(p-2)_+} \mathbb{I}\{1 - \omega(\overline{\mathbf{V}}_{t+i}^{t+i+p-1}) \geq \rho\}$ and $\nu_b^a := \sum_{i=b}^{a} I_{p,i}$. Observe that (recall we define $\prod_{i=c}^{d} x_i = 1$ when $c > d$)

$$\begin{aligned}
\prod_{i=\lfloor (b-s)/p \rfloor + 1}^{\lfloor (a-s)/p \rfloor} (1 - \omega(\overline{\mathbf{V}}_{s+pi-p}^{s+pi-1})) &\leq \rho^{(\lfloor (a-s)/p \rfloor - \lfloor (b-s)/p \rfloor)_+ - \sum_{i=\lfloor (b-s)/p \rfloor + 1}^{\lfloor (a-s)/p \rfloor} \mathbb{I}\{1 - \omega(\overline{\mathbf{V}}_{s+pi-p}^{s+pi-1}) \geq \rho\}} \\
&\leq \rho^{\lfloor (a-b)_+/p \rfloor - \nu_{b-p}^{a-p}},
\end{aligned} \qquad (39)$$

where the second inequality follows from $\lfloor x \rfloor - \lfloor y \rfloor \geq \lfloor x - y \rfloor$, $(\lfloor x/p \rfloor)_+ = \lfloor x_+/p \rfloor$, $s + p(\lfloor (b - s)/p \rfloor + 1) - p \geq b - p$, and $s + p\lfloor (a - s)/p \rfloor - 1 \leq a - 1$. Similarly, we obtain

$$\prod_{i=\lfloor (k-1-a)/p \rfloor + 1}^{\lfloor (k-1-b)/p \rfloor} (1 - \omega(\overline{\mathbf{V}}_{k-2-pi+1}^{k-2-pi+p})) \leq \rho^{\lfloor (a-b)_+/p \rfloor - \nu_b^a}, \qquad (40)$$

because $k - 2 - p\lfloor (k - 1 - b)/p \rfloor + 1 \geq b$ and $k - 2 - p(\lfloor (k - 1 - a)/p \rfloor + 1) + p \leq a + p - 1$. By applying (39) to $\Omega_{t-1,-m}$ with $a = t-1, b = s = -m$, applying (40) to $\tilde{\Omega}_{t,k-1}$ with $a = k-1$ and $b = t$, and using (31) and $-m+1 \le t \le k$, we obtain

$$\begin{aligned} \Omega_{t-1,-m} &\le \rho^{\lfloor (t-1+m)_+/p \rfloor - \nu_{-m-p}^{t-1-p}} \le \rho^{\lfloor (t+m)/(p+1) \rfloor - \nu_{-m-p}^{k}}, \\ \tilde{\Omega}_{t,k-1} &\le \rho^{\lfloor (k-1-t)_+/p \rfloor - \nu_t^{k-1}} \le \rho^{\lfloor (k-t)/(p+1) \rfloor - \nu_{-m-p}^{k}}. \end{aligned} \tag{41}$$

Observe that, for any $\rho \in (0,1)$, $c > 0$ and any integers $a < b$,

$$\begin{aligned} \sum_{t=-\infty}^{\infty} \left( \rho^{\lfloor (t+a)/c \rfloor} \wedge \rho^{\lfloor (b-t)/c \rfloor} \right) &\le \sum_{t=-\infty}^{\lfloor (b-a)/2 \rfloor} \rho^{\lfloor (b-t)/c \rfloor} + \sum_{t=\lfloor (b-a)/2 \rfloor + 1}^{\infty} \rho^{\lfloor (t+a)/c \rfloor} \\ &\le \frac{c}{1-\rho} \left( \rho^{\lfloor (b - \lfloor (b-a)/2 \rfloor)/c \rfloor} + \rho^{\lfloor (\lfloor (b-a)/2 \rfloor + 1 + a)/c \rfloor} \right) \\ &\le \frac{2c}{1-\rho} \rho^{\lfloor (a+b)/2c \rfloor}. \end{aligned} \tag{42}$$

From (42), $\sum_{t=-m+1}^{k} (\Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1})$ is bounded by $2(p+1)\rho^{\lfloor (k+m)/2(p+1) \rfloor - \nu_{-m-p}^{k}}/(1-\rho)$. Because $\overline{\mathbf{V}}_i^{i+p-1}$ is stationary and ergodic, it follows from the strong law of large numbers that $(\lfloor (k+m)/2(p+1) \rfloor)^{-1} \nu_{-m-p}^{k} \to 2(p+1)\mathbb{E}_{\theta^*}[I_{p,i}] \le 2p(p+1)\epsilon$ $\mathbb{P}_{\theta^*}$-a.s. as $k + m \to \infty$. In view of $\epsilon < 1/8p(p+1)$, we have $\mathbb{P}_{\theta^*}(\rho^{\lfloor (k+m)/2(p+1) \rfloor - \nu_{-m-p}^{k}} \ge \rho^{\lfloor (k+m)/2(p+1) \rfloor /2}$ i.o.$) = 0$. Henceforth, let $\{b_{k,m}\}_{k \ge 1, m \le 0}$ denote a generic nonnegative random sequence such that $\mathbb{P}_{\theta^*}(b_{k,m} \ge M$ i.o.$) = 0$ for a finite constant $M$. With this notation and the fact that $\lfloor (k+m)/2(p+1) \rfloor /2 \ge \lfloor (k+m)/4(p+1) \rfloor$, $\sum_{t=-m+1}^{k} (\Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1})$ is bounded by

$$\rho^{\lfloor (k+m)/4(p+1) \rfloor} b_{k,m}, \tag{43}$$

and part (a) is proven.

For part (b), it follows from (12) and Lemma 10(a)–(e) that

$$\left| \Psi_{k,m,x}^{j}(\theta) - \Psi_{k,m',x'}^{j}(\theta) \right| \le 4 \sum_{t=-m+1}^{k} \left( \Omega_{t-1,-m} \wedge \tilde{\Omega}_{t,k-1} \right) |\phi_t^j|_\infty + 2 \sum_{t=-m'+1}^{-m} \tilde{\Omega}_{t,k-1} |\phi_t^j|_\infty.$$

The first term on the right hand side is bounded by $(k+m)^2 K_j \rho^{\lfloor (k+m)/4(p+1) \rfloor} b_{k,m}$ with $K_j \in L^{3-j}(\mathbb{P}_{\theta^*})$ from the same argument as the proof of part (a). For the second term on the right hand side, write $\tilde{\Omega}_{t,k-1}$ as $\tilde{\Omega}_{t,k-1} = \tilde{\Omega}_{-m,k-1} \tilde{\Omega}_{t,k-1}^{-m}$, where $\tilde{\Omega}_{t,k-1}^{-m} := \prod_{i=\lfloor (k-1+m)/p \rfloor + 1}^{\lfloor (k-1-t)/p \rfloor} (1 - \omega(\overline{\mathbf{V}}_{k-2-pi+1}^{k-2-pi+p}))$. By applying (40) to $\tilde{\Omega}_{t,k-1}^{-m}$ with $a = -m$ and $b = t$, we obtain $\tilde{\Omega}_{t,k-1}^{-m} \le \rho^{\lfloor (-m-t)/p \rfloor - \nu_t^{-m}}$. In conjunction with $\tilde{\Omega}_{t,k-1}^{-m} \le 1$, the second term on the right hand side is bounded by $2\tilde{\Omega}_{-m,k-1} R_{m,m'}$, where

$$R_{m,m'} := \sum_{t=-m'+1}^{-m} d_{t,m} |\phi_t^j|_\infty, \quad d_{t,m} := \min\{1, \rho^{\lfloor (-m-t)/p \rfloor - \nu_t^{-m}}\}. \tag{44}$$

From a similar argument to (41)–(43), we can bound $\tilde{\Omega}_{-m,k-1}$ as $\tilde{\Omega}_{-m,k-1} \le \rho^{\lfloor (k+m)/4(p+1) \rfloor} b_{k,m}$.

It follows from $(-m-t)^{-1}\nu_t^{-m} \to \mathbb{E}_{\theta^*}[I_{p,i}] \leq p\epsilon$ $\mathbb{P}_{\theta^*}$-a.s. as $t+m \to -\infty$ that $\mathbb{P}_{\theta^*}(d_{t,m} \geq \rho^{\lfloor(-m-t)/p\rfloor/2} \text{ i.o.}) = 0$. Furthermore, $|\phi_t^j|_\infty$ satisfies $\mathbb{P}_{\theta^*}(|\phi_t^j|_\infty \geq \rho^{-\lfloor(-m-t)/p\rfloor/4} \text{ i.o.}) = 0$ from Markov's inequality and the Borel-Cantelli lemma. Therefore, $\mathbb{P}_{\theta^*}(d_{t,m}|\phi_t^j|_\infty \geq \rho^{\lfloor(-m-t)/p\rfloor/4} \text{ i.o.}) = 0$. In conjunction with $0 \leq d_{t,m}|\phi_t^j|_\infty < \infty$ $\mathbb{P}_{\theta^*}$-a.s., we obtain $\overline{R}_m := \sup_{m' \geq m} R_{m,m'} < \infty$ $\mathbb{P}_{\theta^*}$-a.s., and the distribution of $\overline{R}_m$ does not depend on $m$ because $\overline{\mathbf{V}}_t$ is stationary. Therefore, part (b) is proven by setting $B_m = \overline{R}_m$. □

*Proof of Proposition 2.* By setting $m = 0$ and letting $m' \to \infty$ in Lemma 6, we obtain $\sup_{\theta\in G}\sup_{x\in\mathcal{X}} |\Psi^1_{k,0,x}(\theta) - \Psi^1_{k,\infty}(\theta)| \leq (K_1 + B_0)k^2\rho^{\lfloor k/4(p+1)\rfloor}A_{k,0}$. Furthermore, the sum over finitely many $\sup_{\theta\in G}\sup_{x\in\mathcal{X}} |\Psi^1_{k,0,x}(\theta) - \Psi^1_{k,\infty}(\theta)|$ is $o(n^{1/2})$ $\mathbb{P}_{\theta^*}$-a.s. because $\mathbb{E}_{\theta^*}[\sup_{\theta\in G}\sup_{x\in\mathcal{X}} |\Psi^1_{k,0,x}(\theta)|] < \infty$ and $\mathbb{E}_{\theta^*}[\sup_{\theta\in G} |\Psi^1_{k,\infty}(\theta)|] < \infty$ from Assumption 8. Therefore, we have $n^{-1/2}\nabla_\theta l_n(\theta^*, x_0) = n^{-1/2}\sum_{k=1}^n \Psi^1_{k,0,x_0}(\theta^*) = n^{-1/2}\sum_{k=1}^n \Psi^1_{k,\infty}(\theta^*) + o_p(1)$.

Because $\{\Psi^1_{k,\infty}(\theta^*)\}_{k=-\infty}^\infty$ is a stationary, ergodic, and square integrable martingale difference sequence, it follows from a martingale difference central limit theorem (McLeish, 1974, Theorem 2.3) that $n^{-1/2}\sum_{k=1}^n \Psi^1_{k,\infty}(\theta^*) \to_d N(0, I(\theta^*))$, and part (a) follows. For part (b), let $p_{n\theta}(x_0)$ denote $p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)$, and observe that

$$\nabla_\theta l_n(\theta, \xi) = \frac{\nabla_\theta \int p_{n\theta}(x_0)\xi(dx_0)}{\int p_{n\theta}(x_0)\xi(dx_0)} = \frac{\int \nabla_\theta \log p_{n\theta}(x_0) p_{n\theta}(x_0)\xi(dx_0)}{\int p_{n\theta}(x_0)\xi(dx_0)}.$$

Therefore, $\min_{x_0}\nabla_\theta l_n(\theta^*, x_0) \leq \nabla_\theta l_n(\theta^*, \xi) \leq \max_{x_0}\nabla_\theta l_n(\theta^*, x_0)$ holds, and part (b) follows. □

*Proof of Lemma 7.* The proof follows the argument of the proof of Lemma 17 in DMR and the proof of Lemma 6. Fix $\epsilon \in (0, 1/32p(p+1)]$ and choose $\rho \in (0,1)$ as in the proof of Lemma 6. When $(k,m) = (1,0)$, the stated result follows from $\sup_{\theta\in G}|\phi_{\theta k}| \leq |\phi_k|_\infty$. Henceforth, assume $(k,m) \neq (1,0)$ so that $k + m \geq 2$. For $a \leq b$, define $S_a^b := \sum_{t=a}^b \phi_{\theta t}$. Let $\{b_{k,m}\}_{k\geq 1, m\leq 0}$ denote a generic nonnegative random sequence such that $\mathbb{P}_{\theta^*}(b_{k,m} \geq M \text{ i.o.}) = 0$ for a finite constant $M$. We prove part (a) first. Write $\Gamma_{k,m,x}(\theta) - \Gamma_{k,m}(\theta) = A + 2B + C$, where

$$\begin{aligned}
A &:= \mathrm{var}_\theta[S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^k, X_{-m} = x] - \mathrm{var}_\theta[S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^{k-1}, X_{-m} = x] \\
&\quad - \mathrm{var}_\theta[S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^k] + \mathrm{var}_\theta[S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^{k-1}], \\
B &:= \mathrm{cov}_\theta[\phi_{\theta k}, S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^k, X_{-m} = x] - \mathrm{cov}_\theta[\phi_{\theta k}, S_{-m+1}^{k-1}|\overline{\mathbf{V}}_{-m}^k], \\
C &:= \mathrm{var}_\theta[\phi_{\theta k}|\overline{\mathbf{V}}_{-m}^k, X_{-m} = x] - \mathrm{var}_\theta[\phi_{\theta k}|\overline{\mathbf{V}}_{-m}^k].
\end{aligned}$$

From Lemma 10(f)–(k), $A$ is bounded as

$$|A| \leq 24 \sum_{-m+1\leq s\leq t\leq k-1} \left(\Omega_{s-1,-m} \wedge \Omega_{t-1,s} \wedge \tilde{\Omega}_{t,k-1}\right) \max_{-m+1\leq s\leq t\leq k-1} |\phi_t|_\infty |\phi_s|_\infty.$$

From equation (46) of DMR on page 2299, we have $\max_{-m+1\leq s\leq t\leq k-1}|\phi_t|_\infty|\phi_s|_\infty \leq (m^3 + k^3)\sum_{t=-\infty}^\infty |\phi_t|_\infty^2/(|t| \vee 1)^2 \leq (k+m)^3 K$ for $K \in L^1(\mathbb{P}_{\theta^*})$.

We proceed to bound $\Omega_{s-1,-m} \wedge \Omega_{t-1,s} \wedge \tilde{\Omega}_{t,k-1}$. By using the argument in (39)-(41), we obtain

$$\Omega_{s-1,-m} \wedge \Omega_{t-1,s} \wedge \tilde{\Omega}_{t,k-1} \leq \left(\rho^{\lfloor (s+m)/(p+1) \rfloor} \wedge \rho^{\lfloor (t-s)/(p+1) \rfloor} \wedge \rho^{\lfloor (k-t)/(p+1) \rfloor}\right) \rho^{-\nu^k_{-m-p}}.$$

Furthermore, a derivation similar to DMR (page 2299) gives, for $n \geq 2$,

$$\sum_{0 \leq s \leq t \leq n} \left(\rho^{\lfloor s/(p+1) \rfloor} \wedge \rho^{\lfloor (t-s)/(p+1) \rfloor} \wedge \rho^{\lfloor (n-t)/(p+1) \rfloor}\right) \leq 2 \sum_{s=0}^{n/2} \sum_{t=s}^{n-s} \left(\rho^{\lfloor (t-s)/(p+1) \rfloor} \wedge \rho^{\lfloor (n-t)/(p+1) \rfloor}\right).$$

From (42), the right hand side is bounded by

$$\mathcal{C} \sum_{s=0}^{n/2} \rho^{\lfloor (n-s)/2(p+1) \rfloor} \leq \mathcal{C} \rho^{\lfloor n/4(p+1) \rfloor}, \tag{45}$$

where the inequality holds because $\sum_{t=a}^{\infty} \rho^{\lfloor t/b \rfloor} \leq b\rho^{\lfloor a/b \rfloor}/(1-\rho)$ for any integers $a \geq 0$ and $b > 0$. Hence, $A$ is bounded by $K(k+m)^3 \rho^{\lfloor (k+m)/4(p+1) \rfloor} b_{k,m}$ by setting $n = k+m$ in (45) and noting that $(\lfloor (k+m)/4(p+1) \rfloor)^{-1} \nu^k_{-m-p} \to 4(p+1)\mathbb{E}_{\theta^*}[I_{p,i}] \leq 4p(p+1)\epsilon < 1/2$ $\mathbb{P}_{\theta^*}$-a.s. as $k+m \to \infty$. For $B$, from Lemma 10(f)–(i), (39), (41), $t \geq -m$, and (42), $B$ is bounded as, with $M_k := \max_{-m+1 \leq t \leq k-1} |\phi_k|_\infty |\phi_t|_\infty$,

$$\begin{aligned}
|B| &\leq 12 \sum_{-m+1 \leq t \leq k-1} \left(\Omega_{t-1,-m} \wedge \Omega_{k-1,t}\right) M_k \\
&\leq 12 \sum_{-m+1 \leq t \leq k-1} \left(\rho^{(t+m)/(p+1)} \wedge \rho^{(k-t)/(p+1)}\right) \rho^{-\nu^k_{-m-p}} M_k \\
&\leq \mathcal{C} \rho^{\lfloor (k+m)/2(p+1) \rfloor - \nu^k_{-m-p}} M_k,
\end{aligned}$$

which is written as $K(k+m)^3 \rho^{\lfloor (k+m)/4(p+1) \rfloor} b_{k,m}$ for $K \in L^1(\mathbb{P}_{\theta^*})$. $C$ is bounded by $6\Omega_{k-1,-m}|\phi_k|^2_\infty$ from Lemma 10(h), and part (a) is proven.

We proceed to prove part (b). Write $\Gamma_{k,m',x'}(\theta) = A + 2B + 2C + D$, where

$$\begin{aligned}
A &:= \mathrm{var}_\theta[S^k_{-m+1} | \overline{\mathbf{V}}^k_{-m'}, X_{-m'} = x'] - \mathrm{var}_\theta[S^{k-1}_{-m+1} | \overline{\mathbf{V}}^{k-1}_{-m'}, X_{-m'} = x'], \\
B &:= \mathrm{cov}_\theta[\phi_{\theta k}, S^{-m}_{-m'+1} | \overline{\mathbf{V}}^k_{-m'}, X_{-m'} = x'], \\
C &:= \mathrm{cov}_\theta[S^{k-1}_{-m+1}, S^{-m}_{-m'+1} | \overline{\mathbf{V}}^k_{-m'}, X_{-m'} = x'] - \mathrm{cov}_\theta[S^{k-1}_{-m+1}, S^{-m}_{-m'+1} | \overline{\mathbf{V}}^{k-1}_{-m'}, X_{-m'} = x'], \\
D &:= \mathrm{var}_\theta[S^{-m}_{-m'+1} | \overline{\mathbf{V}}^k_{-m'}, X_{-m'} = x'] - \mathrm{var}_\theta[S^{-m}_{-m'+1} | \overline{\mathbf{V}}^{k-1}_{-m'}, X_{-m'} = x'].
\end{aligned}$$

$|\Gamma_{k,m,x}(\theta) - A|$ is bounded similarly to $|\Gamma_{k,m,x}(\theta) - \Gamma_{k,m}(\theta)|$ in part (a) by using Lemma 10. From

Lemma 10(g), $B$ is bounded by $2\sum_{t=-m'+1}^{-m}\Omega_{k-1,t}|\phi_k|_\infty|\phi_t|_\infty = B_1\times B_2$, where

$$B_1 := 2|\phi_k|_\infty \prod_{i=\lfloor(-m-t)/p\rfloor+1}^{\lfloor(k-1-t)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{t+pi-p}^{t+pi-1})), \quad B_2 := \sum_{t=-m'+1}^{-m}\prod_{i=1}^{\lfloor(-m-t)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{t+pi-p}^{t+pi-1}))|\phi_t|_\infty.$$

$B_1$ is bounded by $|\phi_k|_\infty\rho^{\lfloor(k+m)/2(p+1)\rfloor}b_{k,m}$ from the same argument as part (a). Because $\mathbb{P}_{\theta^*}(|\phi_k|_\infty \geq \rho^{-\lfloor(k+m)/2(p+1)\rfloor/2}$ i.o.$) = 0$, $B_1$ is bounded by $\rho^{\lfloor(k+m)/4(p+1)\rfloor}b_{k,m}$. For $B_2$, because $\prod_{i=1}^{\lfloor(-m-t)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{t+pi-p}^{t+pi-1}))$ is bounded by $\rho^{\lfloor(-m-t)/p\rfloor-\nu_{t-p}^{-m}}$ from (39), we can use the same argument as the one for $R_{m,m'}$ defined in (44) to show that $\overline{B}_{2m} := \sup_{m'\geq m}B_2 < \infty$ $\mathbb{P}_{\theta^*}$-a.s. and $\overline{B}_{2m}$ is stationary. Therefore, $B$ is bounded by $\rho^{\lfloor(k+m)/4(p+1)\rfloor}b_{k,m}\overline{B}_{2m}$.

$|C|+|D|$ is bounded by, with $\Delta_{t,s} := |\mathrm{cov}_\theta[\phi_{\theta t},\phi_{\theta s}|\overline{\mathbf{V}}_{-m'}^{k}, X_{-m'}=x'] - \mathrm{cov}_\theta[\phi_{\theta t},\phi_{\theta s}|\overline{\mathbf{V}}_{-m'}^{k-1}, X_{-m'}=x']|$,

$$\sum_{t=-m'+1}^{k-1}\sum_{s=-m'+1}^{-m}\Delta_{t,s} \leq 2\sum_{s=-m'+1}^{-m}\sum_{t=s}^{k-1}\Delta_{t,s} \leq 2\sum_{s=-m'+1}^{-m}\sum_{t=s}^{k-1}\left(\Omega_{t-1,s}\wedge\tilde{\Omega}_{t,k-1}\right)|\phi_t|_\infty|\phi_s|_\infty. \tag{46}$$

Similar to (41), we obtain

$$\Omega_{t-1,s}\wedge\tilde{\Omega}_{t,k-1} \leq \left(\rho^{\lfloor(t-s)/(p+1)\rfloor-\nu_{s-p}^{t-1-p}}\wedge\rho^{\lfloor(k-t)/(p+1)\rfloor-\nu_t^{k-1}}\right) \leq \left(\rho^{\lfloor(t-s)/(p+1)\rfloor}\wedge\rho^{\lfloor(k-t)/(p+1)\rfloor}\right)\rho^{-\nu_{s-p}^{k-1}}.$$

Therefore, the right hand side of (46) is bounded by

$$2\sum_{s=-m'+1}^{-m}\sum_{t=s}^{k-1}\left(\rho^{\lfloor(t-s)/(p+1)\rfloor}\wedge\rho^{\lfloor(k-t)/(p+1)\rfloor}\right)\rho^{-\nu_{-m}^{k-1}}\rho^{-\nu_{s-p}^{-m}}|\phi_t|_\infty|\phi_s|_\infty. \tag{47}$$

DMR (page 2300) show that the following holds for $k\geq 1$, $m\geq 0$ and $t,s\leq 0$:

$$\begin{aligned}&\text{if } t\leq(k+s-1)/2, \text{ then } (|t|-1)/2\leq(3k+s-3)/4-t,\\ &\text{if } (k+s-1)/2\leq t\leq k-1, \text{ then } (|t|-1)/4\leq t+(-k-3s+1)/4.\end{aligned}$$

Consequently, (47) is bounded by

$$\begin{aligned}&2\rho^{\lfloor(k+m-2)/8(p+1)\rfloor-\nu_{-m}^{k-1}}\sum_{s=-m'+1}^{-m}\rho^{\lfloor(k-2s-m)/8(p+1)\rfloor}\rho^{-\nu_{s-p}^{-m}}|\phi_s|_\infty\\ &\quad\times\left(\sum_{t=s}^{(k+s)/2}\rho^{\lfloor((3k+s-3)/4-t)/(p+1)\rfloor}|\phi_t|_\infty + \sum_{t=(k+s)/2}^{k-1}\rho^{\lfloor(t+(-k-3s+1)/4)/(p+1)\rfloor}|\phi_t|_\infty\right)\\ &\leq\mathcal{C}\rho^{\lfloor(k+m-2)/8(p+1)\rfloor-\nu_{-m}^{k-1}}\sum_{s=-m'+1}^{-m}\rho^{\lfloor(-m-s)/8(p+1)\rfloor-\nu_{s-p}^{-m}}|\phi_s|_\infty\sum_{t=s}^{k-1}\rho^{\lfloor(|t|-1)/4(p+1)\rfloor}|\phi_t|_\infty\\ &\leq\rho^{\lfloor(k+m)/16(p+1)\rfloor}b_{k,m}\times E\times F_{m,m'},\end{aligned}$$

where $E := \sum_{t=-\infty}^{\infty} \rho^{\lfloor (|t|-1)/4(p+1) \rfloor} |\phi_t|_\infty$, and $F_{m,m'} := \sum_{s=-m'+1}^{-m} \rho^{\lfloor (-m-s)/8(p+1) \rfloor - \nu_{s-p}^{-m}} |\phi_s|_\infty$. Because $E \in L^1(\mathbb{P}_{\theta^*})$, $\overline{F}_m := \sup_{m' \geq m} F_{m,m'} < \infty$ $\mathbb{P}_{\theta^*}$-a.s., and $\overline{F}_m$ is stationary, (47) is bounded by $\rho^{\lfloor (k+m)/16(p+1) \rfloor} b_{k,m} E \overline{F}_m$, and part (b) is proven. □

*Proof of Proposition 3.* Define $\Upsilon_{k,m,x}(\theta) := \Psi^2_{k,m,x}(\theta) + \Gamma_{k,m,x}(\theta)$ and $\Upsilon_{k,\infty}(\theta) := \Psi^2_{k,\infty}(\theta) + \Gamma_{k,\infty}(\theta)$, so that $\nabla^2_\theta l_n(\theta, x) = \sum_{k=1}^n \Upsilon_{k,0,x}(\theta)$. By setting $m = 0$ and letting $m' \to \infty$ in Lemmas 6 and 7, we obtain $\sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\Upsilon_{k,0,x}(\theta) - \Upsilon_{k,\infty}(\theta)| \leq (K_2 + B_0) k^2 \rho^{\lfloor k/4(p+1) \rfloor} A_{k,0} + K(k^3 + D_0) \rho^{\lfloor k/16(p+1) \rfloor} C_{k,0}$. Furthermore, the sum over finitely many $\sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\Upsilon_{k,0,x}(\theta) - \Upsilon_{k,\infty}(\theta)|$ is $o(n)$ $\mathbb{P}_{\theta^*}$-a.s. because $\mathbb{E}_{\theta^*} \sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\Upsilon_{k,0,x}(\theta)| < \infty$ and $\mathbb{E}_{\theta^*} \sup_{\theta \in G} |\Upsilon_{k,\infty}(\theta)| < \infty$ from Assumption 8. Therefore, we have $\sup_{\theta \in G} \sup_{x \in \mathcal{X}} |n^{-1} \nabla^2_\theta l_n(\theta, x) - n^{-1} \sum_{k=1}^n \Upsilon_{k,\infty}(\theta)| = o_p(1)$.

Consequently, it suffices to show that

$$\sup_{\theta \in G} \left| n^{-1} \sum_{k=1}^n \Upsilon_{k,\infty}(\theta) - \mathbb{E}_{\theta^*}[\Upsilon_{0,\infty}(\theta)] \right| \to_p 0. \tag{48}$$

Because $G$ is compact, (48) holds if, for all $\theta \in G$,

$$n^{-1} \sum_{k=1}^n \Upsilon_{k,\infty}(\theta) - \mathbb{E}_{\theta^*}[\Upsilon_{0,\infty}(\theta)] \to_p 0, \tag{49}$$

$$\lim_{\delta \to 0} \lim_{n \to \infty} \sup_{|\theta' - \theta| \leq \delta} \left| n^{-1} \sum_{k=1}^n \Upsilon_{k,\infty}(\theta') - n^{-1} \sum_{k=1}^n \Upsilon_{k,\infty}(\theta) \right| = 0 \quad \mathbb{P}_{\theta^*}\text{-}a.s. \tag{50}$$

(49) holds by ergodic theorem. Note that the left hand side of (50) is bounded by $\lim_{\delta \to 0} \lim_{n \to \infty} n^{-1} \sum_{k=1}^n \sup_{|\theta' - \theta| \leq \delta} |\Upsilon_{k,\infty}(\theta') - \Upsilon_{k,\infty}(\theta)|$, which equals $\lim_{\delta \to 0} \mathbb{E}_{\theta^*} \sup_{|\theta' - \theta| \leq \delta} |\Upsilon_{0,\infty}(\theta') - \Upsilon_{0,\infty}(\theta)|$ $\mathbb{P}_{\theta^*}$-a.s. from ergodic theorem. Therefore, (50) holds if

$$\lim_{\delta \to 0} \mathbb{E}_{\theta^*} \sup_{|\theta' - \theta| \leq \delta} \left| \Upsilon_{0,\infty}(\theta') - \Upsilon_{0,\infty}(\theta) \right| = 0. \tag{51}$$

Fix a point $x_0 \in \mathcal{X}$. The left hand side of (51) is bounded by $2A_m + C_m$, where

$$A_m := \mathbb{E}_{\theta^*} \sup_{\theta \in G} |\Upsilon_{0,m,x_0}(\theta) - \Upsilon_{0,\infty}(\theta)|, \quad C_m := \lim_{\delta \to 0} \mathbb{E}_{\theta^*} \sup_{|\theta' - \theta| \leq \delta} \left| \Upsilon_{0,m,x_0}(\theta') - \Upsilon_{0,m,x_0}(\theta) \right|.$$

From Lemmas 6 and 7, $\sup_{\theta \in G} |\Upsilon_{0,m,x_0}(\theta) - \Upsilon_{0,\infty}(\theta)| \to_p 0$ as $m \to \infty$. Furthermore, we have $\mathbb{E}_{\theta^*} \sup_{m \geq 1} \sup_{\theta \in G} |\Upsilon_{0,m,x_0}(\theta)| < \infty$ and $\mathbb{E}_{\theta^*} \sup_{\theta \in G} |\Upsilon_{0,\infty}(\theta)| < \infty$ from Assumption 8. Therefore, $A_m \to 0$ as $m \to \infty$ by the dominated convergence theorem (Durrett, 2010, Exercise 2.3.7). $C_m = 0$ from Lemma 11 if $m \geq p$. Therefore, (51) holds, and the stated result is proven. □

*Proof of Proposition 4.* In view of (10) and Propositions 1, 2, and 3, part (a) holds if (i) $\mathbb{E}_{\theta^*}[\Psi^2_{0,\infty}(\theta) + \Gamma_{0,\infty}(\theta)]$ is continuous in $\theta \in G$ and (ii) $\mathbb{E}_{\theta^*}[\Psi^2_{0,\infty}(\theta^*) + \Gamma_{0,\infty}(\theta^*)] = -I(\theta^*)$. (i) follows from (51). For (ii), it follows from the Louis information principle and information matrix equality that, for all $m \geq 1$, $\mathbb{E}_{\theta^*}[\Psi^1_{0,m}(\theta^*)(\Psi^1_{0,m}(\theta^*))'] = -\mathbb{E}_{\theta^*}[\Psi^2_{0,m}(\theta^*) + \Gamma_{0,m}(\theta^*)]$. From Lemmas 6 and 7, Assumption 8, and the dominated convergence theorem, the left hand side converges to $\mathbb{E}_{\theta^*}[\Psi^1_{0,\infty}(\theta^*)(\Psi^1_{0,\infty}(\theta^*))'] = I(\theta^*)$, and the right hand side converges to $-\mathbb{E}_{\theta^*}[\Psi^2_{0,\infty}(\theta^*) + \Gamma_{0,\infty}(\theta^*)]$. Therefore, (ii) holds, and part (a) is proven.

For part (b), an elementary calculation gives, with $p_{n\theta}(x)$ denoting $p_\theta(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x)$,

$$n^{-1}\nabla^2_\theta l_n(\theta,\xi) = \frac{\int n^{-1}\nabla^2_\theta \log p_{n\theta}(x_0) p_{n\theta}(x_0)\xi(dx_0)}{\int p_{n\theta}(x_0)\xi(dx_0)} + \frac{\int (n^{-1/2}\nabla_\theta \log p_{n\theta}(x_0))^2 p_{n\theta}(x_0)\xi(dx_0)}{\int p_{n\theta}(x_0)\xi(dx_0)}$$
$$- \left( \frac{\int n^{-1/2}\nabla_\theta \log p_{n\theta}(x_0) p_{n\theta}(x_0)\xi(dx_0)}{\int p_{n\theta}(x_0)\xi(dx_0)} \right)^2.$$

The sum of the last two terms is $o_p(1)$ because $\sup_{\theta\in G}\sup_{x\in\mathcal{X}} |n^{-1/2}\nabla_\theta \log p_{n\theta}(x) - n^{-1/2}\sum_{k=1}^n \Psi^1_{k,\infty}(\theta)| = o_p(1)$. Therefore, $\min_{x_0} n^{-1}\nabla^2_\theta l_n(\theta,x_0) + o_p(1) \le n^{-1}\nabla^2_\theta l_n(\theta,\xi) \le \max_{x_0} n^{-1}\nabla^2_\theta l_n(\theta,x_0) + o_p(1)$ holds, and part (b) follows. □

# 8 Auxiliary results

The following lemma provides the convergence rate of a Markov chain $X_t$. When $X_t$ is time-homogeneous, this result has been proven by Theorem 1 of Rosenthal (1995). This lemma extends Rosenthal (1995) to time-inhomogeneous $X_t$.

**Lemma 8.** *Let $\{X_t\}_{t\ge1}$ be a Markov process that lies in $\mathcal{X}$, and let $P_t(x,A) := \mathbb{P}(X_t \in A|X_{t-1} = x)$. Suppose there is a probability measure $Q_t(\cdot)$ on $\mathcal{X}$, a positive integer $p$, and $\varepsilon_t \ge 0$ such that*

$$P_t^p(x,A) := \mathbb{P}(X_t \in A|X_{t-p} = x) \ge \varepsilon_t Q_t(A),$$

*for all $x \in \mathcal{X}$ and all measurable subsets $A \subset \mathcal{X}$. Let $X_0$ and $Y_0$ be chosen from the initial distributions $\pi_1$ and $\pi_2$, respectively, and update them according to $P_t(x,A)$. Then,*

$$\|\mathbb{P}(X_k \in \cdot) - \mathbb{P}(Y_k \in \cdot)\|_{TV} \le \prod_{i=1}^{\lfloor k/p \rfloor} (1 - \varepsilon_{ip}).$$

*Proof.* The proof follows the line of argument in the proof of Theorem 1 of Rosenthal (1995). Starting from $(X_0, Y_0)$, we let $X_t$ and $Y_t$ for $t \ge 1$ progress as follows. Given the value of $X_t$ and $Y_t$, flip a coin with the probability of heads equal to $\varepsilon_{t+p}$. If the coin comes up heads, then choose a point $x \in \mathcal{X}$ according to $Q_{t+p}(\cdot)$ and set $X_{t+p} = Y_{t+p} = x$, choose $(X_{t+1},\ldots,X_{t+p-1})$ and $(Y_{t+1},\ldots,Y_{t+p-1})$ independently according to the transition kernel $P_{t+1}(x_{t+1}|x_t),\ldots,P_{t+p-1}(x_{t+p-1}|x_{t+p-2})$ conditional on $X_{t+p} = x$ and $Y_{t+p} = x$, and update the processes after $t+p$ so that they remain equal for all future time. If the coin comes up tails, then choose $X_{t+p}$ and $Y_{t+p}$ independently according to the distributions $(P^p_{t+p}(X_t,\cdot) - \varepsilon_{t+p}Q_{t+p}(\cdot))/(1-\varepsilon_{t+p})$ and $(P^p_{t+p}(Y_t,\cdot) - \varepsilon_{t+p}Q_{t+p}(\cdot))/(1-\varepsilon_{t+p})$, respectively, and choose $(X_{t+1},\ldots,X_{t+p-1})$ and $(Y_{t+1},\ldots,Y_{t+p-1})$ independently according to the transition kernel $P_{t+1}(x_{t+1}|x_t),\ldots,P_{t+p-1}(x_{t+p-1}|x_{t+p-2})$ conditional on the value of $X_{t+p}$ and

$Y_{t+p}$. It is easily checked that $X_t$ and $Y_t$ are each marginally updated according to the transition kernel $P_t(x, A)$.

Furthermore, $X_t$ and $Y_t$ are coupled the first time (call it $T$) when we choose $X_{t+p}$ and $Y_{t+p}$ both from $Q_{t+p}(\cdot)$ as earlier. It now follows from the coupling inequality that

$$\|\mathbb{P}(X_k \in \cdot) - \mathbb{P}(Y_k \in \cdot)\|_{TV} \leq \mathbb{P}(X_k \neq Y_k) \leq \mathbb{P}(T > k).$$

By construction, when $t$ is a multiple of $p$, $X_t$ and $Y_t$ will couple with probability $\varepsilon_t$. Hence,

$$\mathbb{P}(T > k) \leq (1 - \varepsilon_p) \cdots (1 - \varepsilon_{\lfloor k/p \rfloor \cdot p}),$$

and the stated result follows. $\square$

The following lemma corresponds to Lemma 4 of DMR and implies that $\mathbb{E}_{\theta^*}[\Delta_{0,\infty}(\theta)]$ is continuous in $\theta$. This lemma is used in the proof of the consistency of the MLE.

**Lemma 9.** *Assume Assumptions 1–6. Then, for all $\theta \in \Theta$,*

$$\lim_{\delta \to 0} \mathbb{E}_{\theta^*} \left[ \sup_{|\theta - \theta'| \leq \delta} |\Delta_{0,\infty}(\theta') - \Delta_{0,\infty}(\theta)| \right] = 0.$$

*Proof.* The proof is similar to the proof of Lemma 4 in DMR but requires a small adjustment when $p \geq 2$. We first show that $\Delta_{0,m,x}(\theta)$ is continuous in $\theta$ for any fixed $x \in \mathcal{X}$ and any $m \geq p+1$. Recall that $\Delta_{0,m,x}(\theta) = \log p_\theta(Y_0 | \overline{\mathbf{Y}}_{-m}^{-1}, \mathbf{W}_{-m}^0, X_{-m} = x)$ and

$$p_\theta(Y_0 | \overline{\mathbf{Y}}_{-m}^{-1}, \mathbf{W}_{-m}^0, X_{-m} = x) = \frac{p_\theta(\mathbf{Y}_{-m+1}^0 | \overline{\mathbf{Y}}_{-m}, \mathbf{W}_{-m}^0, X_{-m} = x)}{p_\theta(\mathbf{Y}_{-m+1}^{-1} | \overline{\mathbf{Y}}_{-m}, \mathbf{W}_{-m}^{-1}, X_{-m} = x)}.$$

For $j \in \{-1, 0\}$, we have

$$\begin{aligned} & p_\theta(\mathbf{Y}_{-m+1}^j | \overline{\mathbf{Y}}_{-m}, \mathbf{W}_{-m}^j, X_{-m} = x) \\ &= \int q_\theta(x, x_{-m+1}) \prod_{i=-m+2}^{j} q_\theta(x_{i-1}, x_i) \prod_{i=-m+1}^{j} g_\theta(Y_i | \overline{\mathbf{Y}}_{i-1}, x_i, W_i) \mu^{\otimes(m+j)}(d\mathbf{x}_{-m+1}^j). \end{aligned} \tag{52}$$

Because the integrand is bounded by $(\sigma_+^0 b_+)^{m+j}$, $p_\theta(\mathbf{Y}_{-m+1}^j | \overline{\mathbf{Y}}_{-m}, X_{-m} = x, \mathbf{W}_{-m}^j)$ is continuous in $\theta$ by the continuity of $q_\theta$ and $g_\theta$ and the bounded convergence theorem. Furthermore, when $j \geq -m + p$, the infimum of the right hand side of (52) in $\theta$ is strictly positive $\mathbb{P}_{\theta^*}$-a.s. from Assumptions 1(d) and 3. Therefore, $\Delta_{0,m,x}(\theta)$ is continuous in $\theta$ $\mathbb{P}_{\theta^*}$-a.s. Because $\{\Delta_{0,m,x}(\theta)\}$ is continuous in $\theta$ and converges uniformly in $\theta \in \Theta$ $\mathbb{P}_{\theta^*}$-a.s., $\Delta_{0,\infty}(\theta)$ is continuous in $\theta \in \Theta$ $\mathbb{P}_{\theta^*}$-a.s. The stated result then follows from $\mathbb{E}_{\theta^*} \sup_{\theta \in \Theta} |\Delta_{0,\infty}(\theta)| < \infty$ by Lemma 3(c) and the dominated convergence theorem. $\square$

The following lemma is used in the proof of Lemmas 6 and 7. This lemma provides the bounds on the difference in the conditional expectations of $\phi_{\theta t}^j = \phi^j(\theta, \overline{\mathbf{Z}}_{t-1}^t, W_t)$ when the conditioning sets are different. Define $\Omega_{\ell,k} := \prod_{i=1}^{\lfloor(\ell-k)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{k+pi-p}^{k+pi-1}))$ and $\tilde{\Omega}_{\ell,k} := \prod_{i=1}^{\lfloor(k-\ell)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{k-1-pi+1}^{k-1-pi+p}))$ with defining $\prod_{i=a}^{b} x_i := 1$ if $b < a$, where $\omega(\cdot)$ is defined in Lemma 1 and $\overline{\mathbf{V}}_a^b := (\overline{\mathbf{Y}}_a^b, \mathbf{W}_a^b)$.

**Lemma 10.** *Assume Assumptions 1–7. Then, for all $m' \geq m \geq 0$, all $-m < s \leq t \leq n$, all $\theta \in G$, and all $x, x' \in \mathcal{X}$ and $j = 1, 2$,*

$$
\begin{aligned}
&(a) \quad |\mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n]| \leq 2\Omega_{t-1,-m}|\phi_t^j|_\infty,\\
&(b) \quad |\mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n] - \mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m'}^n]| \leq 2\Omega_{t-1,-m}|\phi_t^j|_\infty,\\
&(c) \quad |\mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m'}^n, X_{-m'} = x']| \leq 2\Omega_{t-1,-m}|\phi_t^j|_\infty,\\
&(d) \quad |\mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^{n-1}, X_{-m} = x]| \leq 2\tilde{\Omega}_{t,n-1}|\phi_t^j|_\infty,\\
&(e) \quad |\mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^n] - \mathbb{E}_\theta[\phi_{\theta t}^j|\overline{\mathbf{V}}_{-m}^{n-1}]| \leq 2\tilde{\Omega}_{t,n-1}|\phi_t^j|_\infty,
\end{aligned}
$$

*and*

$$
\begin{aligned}
&(f) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n]| \leq 2\Omega_{t-1,s}|\phi_t|_\infty|\phi_s|_\infty,\\
&(g) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x]| \leq 2\Omega_{t-1,s}|\phi_t|_\infty|\phi_s|_\infty\\
&(h) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n]| \leq 6\Omega_{s-1,-m}|\phi_t|_\infty|\phi_s|_\infty,\\
&(i) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m'}^n, X_{-m'} = x']| \leq 6\Omega_{s-1,-m}|\phi_t|_\infty|\phi_s|_\infty,\\
&(j) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n] - \overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^{n-1}]| \leq 6\tilde{\Omega}_{t,n-1}|\phi_t|_\infty|\phi_s|_\infty,\\
&(k) \quad |\overline{\mathrm{cov}}_\theta[\phi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^n, X_{-m} = x] - \overline{\mathrm{cov}}_\theta[\varphi_{\theta t}, \phi_{\theta s}|\overline{\mathbf{V}}_{-m}^{n-1}, X_{-m} = x]| \leq 6\tilde{\Omega}_{t,n-1}|\phi_t|_\infty|\phi_s|_\infty.
\end{aligned}
$$

*Proof of Lemma 10.* To prove parts (a)–(c), we first show that, for all $-m \leq k \leq t-1$, all probability measures $\mu_1$ and $\mu_2$ on $\mathcal{B}(\mathcal{X})$, and all $\overline{\mathbf{V}}_{-m}^n$,

$$
\begin{aligned}
&\sup_A \left| \int \mathbb{P}_\theta(\mathbf{X}_{t-1}^t \in A|X_k = x, \overline{\mathbf{V}}_{-m}^n)\mu_1(dx) - \int \mathbb{P}_\theta(\mathbf{X}_{t-1}^t \in A|X_k = x, \overline{\mathbf{V}}_{-m}^n)\mu_2(dx) \right| \\
&\leq \prod_{i=1}^{\lfloor(t-1-k)/p\rfloor} \left(1-\omega(\overline{\mathbf{V}}_{k+pi-p}^{k+pi-1})\right).
\end{aligned} \tag{53}
$$

When $k = t-1$, (53) holds trivially. When $-m \leq k < t-1$, equation (17) and the Markov property of $Z_t$ imply that $\mathbb{P}_\theta(\mathbf{X}_{t-1}^t \in A|X_k, \overline{\mathbf{V}}_{-m}^n) = \mathbb{P}_\theta(\mathbf{X}_{t-1}^t \in A|X_k, \overline{\mathbf{V}}_k^n) = \int \mathbb{P}_\theta(\mathbf{X}_{t-1}^t \in A|X_{t-1} = x_{t-1}, \overline{\mathbf{V}}_k^n)p_\theta(x_{t-1}|X_k, \overline{\mathbf{V}}_k^n)\mu(dx_{t-1})$. Consequently, from the property of the total variation distance, the left hand side of (53) is bounded by

$$
\left\| \int \mathbb{P}_\theta(X_{t-1} \in \cdot|X_k = x, \overline{\mathbf{V}}_k^n)\mu_1(dx) - \int \mathbb{P}_\theta(X_{t-1} \in \cdot|X_k = x, \overline{\mathbf{V}}_k^n)\mu_2(dx) \right\|_{TV}.
$$

This is bounded by $\prod_{i=1}^{\lfloor(t-1-k)/p\rfloor}(1-\omega(\overline{\mathbf{V}}_{k+pi-p}^{k+pi-1}))$ from Corollary 1, and (53) is proven.

We proceed to show parts (a)–(c). For part (a), observe that

$$\mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{V}}^n_{-m}, X_{-m} = x_{-m}] = \int \nabla^j_\theta \log p_\theta(Y_t, x_t|\overline{\mathbf{Y}}_{t-1}, x_{t-1}, W_t) p_\theta(\mathbf{x}^t_{t-1}|\overline{\mathbf{V}}^n_{-m}, x_{-m})\mu^{\otimes 2}(d\mathbf{x}^t_{t-1}), \tag{54}$$

$$\mathbb{E}_\theta[\phi^j_{\theta t}|\overline{\mathbf{V}}^n_{-m}] = \int \nabla^j_\theta \log p_\theta(Y_t, x_t|\overline{\mathbf{Y}}_{t-1}, x_{t-1}, W_t) p_\theta(\mathbf{x}^t_{t-1}|\overline{\mathbf{V}}^n_{-m})\mu^{\otimes 2}(d\mathbf{x}^t_{t-1}), \tag{55}$$

and $p_\theta(\mathbf{x}^t_{t-1}|\overline{\mathbf{V}}^n_{-m}) = \int p_\theta(\mathbf{x}^t_{t-1}|\overline{\mathbf{V}}^n_{-m}, x_{-m})p_\theta(x_{-m}|\overline{\mathbf{V}}^n_{-m})\mu(dx_{-m})$. Note that, for any conditioning set $\mathcal{G}$, we have $\mathbb{P}_\theta(\mathbf{X}^t_{t-1}|\mathcal{G}) = 0$ if $q_\theta(X_{t-1}, X_t) = 0$. Therefore, the right hand side of (54) and (55) are written as

$$\int \nabla^j_\theta \log g_\theta(Y_t|\overline{\mathbf{Y}}_{t-1}, x_t, W_t) p_\theta(\mathbf{x}^t_{t-1}|\mathcal{F})\mu^{\otimes 2}(d\mathbf{x}^t_{t-1}) + \int_{\mathcal{X}^+_\theta} \nabla^j_\theta \log q_\theta(x_{t-1}, x_t) p_\theta(\mathbf{x}^t_{t-1}|\mathcal{F})\mu^{\otimes 2}(d\mathbf{x}^t_{t-1}),$$

with $\mathcal{F} = \{\overline{\mathbf{V}}^n_{-m}, x_{-m}\}, \{\overline{\mathbf{V}}^n_{-m}\}$. Therefore, part (a) follows from the property of the total variation distance and setting $k = -m$ in (53). Parts (b) and (c) are proven similarly.

Part (d) holds if we show that, for all $-m+1 \le t \le n$ and $\overline{\mathbf{V}}^n_{-m}$,

$$\begin{aligned} &\sup_A \left|\mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|X_{-m} = x, \overline{\mathbf{V}}^n_{-m}) - \mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|X_{-m} = x, \overline{\mathbf{V}}^{n-1}_{-m})\right| \\ &\le \prod_{i=1}^{\lfloor (n-1-t)/p \rfloor} \left(1 - \omega(\overline{\mathbf{V}}^{n-2-pi+p}_{n-2-pi+1})\right). \end{aligned} \tag{56}$$

When $t \ge n-1$, (56) holds trivially. When $t \le n-2$, observe that the time-reversed process $\{Z_{n-k}\}_{0\le k\le n+m}$ is Markov. Hence, for any $-m+1 \le t \le k$, we have $\mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|X_{-m}, \overline{\mathbf{V}}^k_{-m}) = \int \mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|X_t = x_t, \overline{\mathbf{V}}^t_{-m})p_\theta(x_t|X_{-m}, \overline{\mathbf{V}}^k_{-m})\mu(dx_t)$. Therefore, (56) is proven similarly to (53) by using Lemma 5(b). Part (e) is proven similarly by using Lemma 5(a).

We proceed to show parts (f)–(k). In view of (55), part (f) holds if we show that, for all $-m < s \le t \le n$,

$$\begin{aligned} &\sup_{A,B\in\mathcal{B}(\mathcal{X}^2)} \left|\mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A, \mathbf{X}^s_{s-1} \in B|\overline{\mathbf{V}}^n_{-m}) - \mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|\overline{\mathbf{V}}^n_{-m})\mathbb{P}_\theta(\mathbf{X}^s_{s-1} \in B|\overline{\mathbf{V}}^n_{-m})\right| \\ &\le \prod_{i=1}^{\lfloor (t-1-s)/p \rfloor} \left(1 - \omega(\overline{\mathbf{V}}^{s+pi-1}_{s+pi-p})\right). \end{aligned} \tag{57}$$

When $s \ge t-1$, (57) holds trivially because $\prod_{i=1}^j a_i = 1$ when $j < i$. When $s \le t-2$, observe that $\mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A, \mathbf{X}^s_{s-1} \in B|\overline{\mathbf{V}}^n_{-m}) = \int_B \mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|\mathbf{X}^s_{s-1} = \mathbf{x}^s_{s-1}, \overline{\mathbf{V}}^n_{-m})p_\theta(\mathbf{x}^s_{s-1}|\overline{\mathbf{V}}^n_{-m})\mu^{\otimes 2}(d\mathbf{x}^s_{s-1})$ and $\mathbb{P}_\theta(\mathbf{X}^s_{s-1} \in B|\overline{\mathbf{V}}^n_{-m}) = \int_B p_\theta(\mathbf{x}^s_{s-1}|\overline{\mathbf{V}}^n_{-m})\mu^{\otimes 2}(d\mathbf{x}^s_{s-1})$. Hence, in view of the Markov property of $\{X_k\}$ given $\overline{\mathbf{V}}^n_{-m}$, the left hand side of (57) is bounded by $\sup_A \sup_{x_s\in\mathcal{X}} |\mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|X_s = x_s, \overline{\mathbf{V}}^n_{-m}) - \mathbb{P}_\theta(\mathbf{X}^t_{t-1} \in A|\overline{\mathbf{V}}^n_{-m})|$. From (53), this is bounded by $\prod_{i=1}^{\lfloor (t-1-s)/p \rfloor}(1 - \omega(\overline{\mathbf{V}}^{s+pi-1}_{s+pi-p}))$, and (57) follows. Part (g) is proven similarly by replacing the conditioning variable $\overline{\mathbf{V}}^n_{-m}$ with

$(X_{-m}, \overline{\mathbf{V}}^n_{-m})$. Parts (h)–(k) follow from (53), (56), and the relation $|\mathrm{cov}(X,Y|\mathcal{F}_1) - \mathrm{cov}(X,Y|\mathcal{F}_2)| \leq |E(XY|\mathcal{F}_1) - E(XY|\mathcal{F}_2)| + |E(X|\mathcal{F}_1) - E(X|\mathcal{F}_2)|E(Y|\mathcal{F}_1) + E(X|\mathcal{F}_2)|E(Y|\mathcal{F}_1) - E(Y|F_2)|$. □

The following lemma corresponds to Lemma 14 of DMR and shows that $\mathbb{E}_{\theta^*}[\Psi^1_{0,m,x}(\theta)]$, $\mathbb{E}_{\theta^*}[\Psi^2_{0,m,x}(\theta)]$, and $\mathbb{E}_{\theta^*}[\Gamma_{0,m,x}(\theta)]$ are continuous in $\theta$.

**Lemma 11.** *Assume Assumptions 1–8. Then, for $j = 1, 2$, all $x \in \mathcal{X}$ and $m \geq p$, the functions $\Psi^j_{0,m,x}(\theta)$ and $\Gamma_{0,m,x}(\theta)$ are continuous in $\theta \in G$ $\mathbb{P}_{\theta^*}$-a.s. In addition,*

$$
\begin{aligned}
(a) \quad & \lim_{\delta \to 0} \mathbb{E}_{\theta^*}\left[\sup_{|\theta' - \theta| \leq \delta} |\Psi^j_{0,m,x}(\theta) - \Psi^j_{0,m,x}(\theta')|^{3-j}\right] = 0, \\
(b) \quad & \lim_{\delta \to 0} \mathbb{E}_{\theta^*}\left[\sup_{|\theta' - \theta| \leq \delta} |\Gamma_{0,m,x}(\theta) - \Gamma_{0,m,x}(\theta')|\right] = 0.
\end{aligned}
$$

*Proof.* The proof is similar to the proof of Lemma 14 in DMR. For brevity, we suppress $W_t$ and $\mathbf{W}^0_{-m}$ from $\phi^j(\theta, \overline{\mathbf{Z}}^t_{t-1}, W_t)$ and the conditioning set. We prove part (a) first. Note that $\sup_{\theta \in G} \sup_{x \in \mathcal{X}} |\Psi^j_{0,m,x}(\theta)|^{3-j} \leq (2 \sum_{t=-m+1}^0 |\phi^j_t|_\infty)^{3-j} \in L^1(\mathbb{P}_{\theta^*})$. Hence, the stated result holds if, for $m \geq p$ and $-m+1 \leq t \leq 0$,

$$
\lim_{\delta \to 0} \sup_{|\theta' - \theta| \leq \delta} \left| \mathbb{E}_{\theta'}[\phi^j(\theta', \overline{\mathbf{Z}}^t_{t-1}) | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x] - \mathbb{E}_{\theta}[\phi^j(\theta, \overline{\mathbf{Z}}^t_{t-1}) | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x] \right| = 0 \quad \mathbb{P}_{\theta^*}\text{-a.s.}
$$

Write

$$
\mathbb{E}_{\theta}[\phi^j(\theta, \overline{\mathbf{Z}}^t_{t-1}) | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x] = \int \phi^j(\theta, \overline{\mathbf{Z}}^t_{t-1}) p_\theta(\mathbf{X}^t_{t-1} = \mathbf{x}^t_{t-1} | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x) \mu^{\otimes 2}(d\mathbf{x}^t_{t-1}). \quad (58)
$$

For all $\mathbf{x}^t_{t-1}$ such that $p_\theta(\mathbf{x}^t_{t-1} | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x) > 0$, $\phi^j(\theta, \mathbf{x}^t_{t-1}, \overline{\mathbf{Y}}^t_{t-1})$ is continuous in $\theta$ and bounded by $|\phi^j_t|_\infty < \infty$. Furthermore, $p_\theta(\mathbf{X}^t_{t-1} = \mathbf{x}^t_{t-1} | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x)$ is continuous in $\theta$ and bounded from above uniformly in $\theta \in G$ because $p_\theta(\mathbf{X}^t_{t-1} = \mathbf{x}^t_{t-1}, \mathbf{Y}^0_{-m+1} | \overline{\mathbf{Y}}_{-m}, X_{-m} = x)$ is continuous in $\theta$ and bounded from above by $(\sigma^0_+ b_+)^m$ and $p_\theta(\mathbf{Y}^0_{-m+1} | \overline{\mathbf{Y}}_{-m}, X_{-m} = x)$ is continuous in $\theta$ and bounded from below by $\sigma_-^{\lfloor m/p \rfloor} \prod_{t=-m+1}^0 \int \inf_{\theta \in G} g_\theta(Y_t | \overline{\mathbf{Y}}_{t-1}, x_t) \mu(dx_t) > 0$. Consequently, the integrand on the right hand side of (58) is continuous in $\theta$ and bounded from above uniformly in $\theta \in G$. From the dominated convergence theorem, the left hand side of (58) is continuous in $\theta$, and part (a) is proven.

Part (b) holds if, for $-m+1 \leq s \leq t \leq 0$,

$$
\begin{aligned}
\lim_{\delta \to 0} \sup_{|\theta' - \theta| \leq \delta} \Big| & \mathrm{cov}_{\theta'}[\phi(\theta', \overline{\mathbf{Z}}^s_{s-1}), \phi(\theta', \overline{\mathbf{Z}}^t_{t-1}) | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x] \\
& - \mathrm{cov}_{\theta}[\phi(\theta', \overline{\mathbf{Z}}^s_{s-1}), \phi(\theta, \overline{\mathbf{Z}}^t_{t-1}) | \overline{\mathbf{Y}}^0_{-m}, X_{-m} = x] \Big| = 0 \quad \mathbb{P}_{\theta^*}\text{-a.s.}
\end{aligned}
$$

This holds by a similar argument to part (a), and part (b) follows. □

## References


Ang, A. and Bekaert, G. (2002), "International Asset Allocation with Regime Shifts," *Review of Financial Studies*, 15, 1137–1187.

Bickel, P. J., Ritov, Y., and Rydén, T. (1998), "Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models," *Annals of Statistics*, 26, 1614–1635.

Dai, Q., Singleton, K., and Yang, W. (2007), "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields." *Review of Financial Studies*, 20, 1669–1706.

Douc, R., Moulines, É., and Rydén, T. (2004), "Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime," *Annals of Statistics*, 32, 2254–2304.

Durrett, R. (2010), *Probability: Theory and Examples*, Cambridge University Press, 4th ed.

Edwards, S. and Susmel, R. (2001), "Volatility dependence and contagion in emerging equity markets," *Journal of Development Economics*, 66, 505–532.

Evans, M. and Wachtel, P. (1993), "Were Price Changes During the Great Depression Anticipated?: Evidence from Nominal Interest Rates," *Journal of Monetary Economics*, 32, 3–34.

Fong, W. M. (1997), "Volatility persistence and switching ARCH in Japanese stock return," *Financial Engeneering and the Japanese Market*, 4, 37–57.

Francq, C. and Roussignol, M. (1998), "Ergodicity of Autoregressive Processes with Markov-Switching and Consistency of the Maximum-Likelihood Estimator," *Statistics*, 32, 151–173.

Garcia, R. and Perron, P. (1996), "An Analysis of the Real Interest Rate Under Regime Shifts," *Review of Economics and Statisitcs*, 78.

Goodwin, T. H. (1993), "Business-Cycle Analysis with a Markov-Switching Model," *Journal of Business and Economic Statistics*, 11, 331–339.

Gourieroux, C. and Monfort, A. (1995), *Statistics and Econometric Models, Vol. II*, Cambridge, U.K.: Cambridge University Press.

Gray, S. (1996), "Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.

Hamilton, J. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Hamilton, J. and Susmel, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime." *Journal of Econometrics*, 64, 307–333.

Hamilton, J. D. (1996), "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics*, 70, 127–157.

Hamilton, J. D. and Lin, G. (1996), "Stock Market Volatility and the BUsiness Cycle," *Journal of Applied Econometrics*, 11, 573–593.

Hartigan, J. (1985), "Failure of Log-likelihood Ratio Test," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, eds. Le Cam, L. and Olshen, R., Berkeley: University of California Press, vol. 2, pp. 807–810.

Inoue, T. and Okimoto, T. (2008), "Were There Structural Breaks in the Effect of Japanese Monetary Policy?: Re-evaluating the Policy Effects of the Lost Decade," *Journal of the Japanese and International Economies*, 22, 320–342.

Jensen, J. L. and Petersen, N. V. (1999), "Asymptotic normality of the Maximum Likelihood Estimator in State Space Models," *Annals of Statistics*, 27, 514–535.

Krishnamurthy, V. and Rydén, T. (1998), "Consistent Estimation of Linear and Non-Linear Autoregressive Models with Markov Regime," *Journal of Time Series Analysis*, 19, 291–307.

Leroux, B. G. (1992), "Maximum-Likelihood Estimation for Hidden Markov Models," *Stochastic Processes and their Applications*, 40, 127–143.

Levin, D. A., Peres, Y., and Wilmer, E. L. (2009), *Markov Chains and Mixing Times*, American Mathematical Society.

Louis, T. A. (1982), "Finding the Observed Information Matrix When Using the EM Algorithm," *Journal of the Royal Statistical Society Series B*, 44, 226–233.

Maheu, J. M. and McCurdy, T. H. (2000), "Identifying Bull and Bear Markets in Stock Returns," *Journal of Business and Economic Statistics*, 18, 100–112.

McLeish, D. L. (1974), "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability*, 2, 620–628.

Meyn, S. and Tweedie, R. L. (2009), *Markov Chains and Stochastic Stability*, Cambridge University Press, 2nd ed.

Newey, W. K. and McFadden, D. L. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Amsterdam: North-Holland, vol. 4, pp. 2111–2245.

Okimoto, T. (2008), "New Evidence of Asymmetric Dependence Structures in International Equity Markets," *Journal of Financial and Quantitative Analysis*, 43, 787–815.

Ramchand, L. and Susmel, R. (1998), "Volatility and cross correlation across major stock markets," *Journal of Empirical Finance*, 5, 397–4416.

Rosenthal, J. S. (1995), “Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo,” *Journal of the American Statistical Association*, 90, 558–566.

Sims, C. and Zha, T. (2006), “Were There Regime Switches in U.S. Monetary Policy?” *American Economic Review*, 96, 54–81.