Testing the Number of Regimes in Markov Regime Switching 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

January 27, 2017

Abstract

Markov regime switching models have been widely used in numerous empirical applications in economics and finance. However, the asymptotic distribution of the likelihood ratio test statistic for testing the number of regimes in Markov regime switching models is an unresolved problem. This paper proposes the likelihood ratio test of the null hypothesis of M_0 regimes against the alternative hypothesis of $M_0 + 1$ regimes for any $M_0 \geq 1$ and derives its asymptotic distribution.

Key words: asymptotic distribution; DQM expansion; likelihood ratio test; loss of identifiability

1 Introduction

Since Hamilton (1989)'s seminal contribution, Markov regime switching models have been widely used in numerous empirical applications in economics and finance because it can capture many important features in 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).

The number of regimes is an important parameter in applications of Markov regime switching models. Despite its importance, testing for the number of regimes in Markov regime switching models has been an unsolved problem because the standard asymptotic analysis of the likelihood ratio test statistic (LRTS) breaks down due to problems such as non-identifiable parameters, the

*This research is support by the Natural Science and Engineering Research Council of Canada and JSPS Grant-in-Aid for Scientific Research (C) No. 26380267.

true parameter being on the boundary of the parameter space, and the degeneracy of Fisher information matrix. Analyzing the asymptotic properties of LRTS for testing the number of regimes for Markov regime switching models with normal component density, which is popular in empirical applications, is especially difficult because the normal density has an undesirable mathematical property that the second-order derivative with respect to the mean parameter is linearly dependent of the first derivative with respect to the variance parameter, leading to a further singularity.

The issue of non-identifiability under the null hypothesis and the degeneracy in Fisher information matrix in testing the presence of regime switching has been well recognized in the existing literature. Hansen (1992) derives a lower bound on the asymptotic distribution of the LRTS, and Garcia (1998) also studies this problem. Carrasco et al. (2014) propose an information matrix-type test for parameter stability based on the fourth order expansion of the likelihood and show that the contiguous local alternatives are of order $n^{-1/4}$, where n is the sample size. In contrast, in a closely related problem of testing the number of components in finite mixture normal regression models, Kasahara and Shimotsu (2015) shows that an eighth-order Taylor expansion is required to characterize the quadratic-form approximation of the log-likelihood function and, consequently, the contiguous local alternatives are of order $n^{-1/8}$ (see also Chen and Li, 2009; Chen et al., 2012; Ho and Nguyen, 2016).

Cho and White (2007) derive the asymptotic distribution of the quasi-likelihood ratio test statistic (Q-LRTS) for testing single regime against two regimes in the model with scalar switching parameter by rewriting the model as a two-component mixture models, thereby ignoring the temporal dependence of the regimes. Qu and Fan (2015) extend the analysis of Cho and White (2007) and derive the asymptotic distribution of the LRTS that properly takes into account the temporal dependence of the regimes while allowing for multiple switching parameters. Both Cho and White (2007) and Qu and Fan (2015) focus on testing single regime against two regimes. To the best of our knowledge, the asymptotic distribution of the LRTS for testing the null hypothesis of more than two regimes remains unknown. Dannemann and Holtzmann (2008) analyze the Q-LRTS for testing the null of two regimes against three.

This paper proposes a likelihood ratio test of the null hypothesis of M_0 regimes against the alternative hypothesis of M_0+1 regimes for any $M_0 \geq 1$. To this end, this paper develops a version of LaCam's differentiable in quadratic mean (DQM) expansion that expands likelihood ratio under loss of identifiability while adopting the reparameterization of Kasahara and Shimotsu (2015). We show that the log-likelihood function is locally approximated by a quadratic function of polynomials of reparameterized parameters, and derive the asymptotic null distribution of the LRTS using the results of Andrews (1999, 2001).

The DQM expansion under loss of identifiability was developed by Liu and Shao (2003) in an iid setting, and their expansion is based on a generalized score function. We extend Liu and Shao (2003) to accommodate dependent data and also modify it to fit our context of parametric regime switching model. Using the DQM-type expansion has advantage over the "classical" approach based on the Taylor expansion that expands up to the Hessian term in this context because deriving a higher-order expansion becomes tedious as the order of expansion increases in a Markov regime switching model.

Following Douc et al. (2004) [DMR, hereafter], we consider the conditional likelihood given an *arbitrary* distribution of initial unobserved regimes and show that the asymptotic distribution of the LRTS does not depend on the initial distribution of unobserved regimes. The latter result follows from the deterministic geometrically decaying bound on the conditional chain of unobserved regime at time k given the information up to the time $k-1$ given in equation (2). Applying Missing Information Principle (Woodbury, 1971; Louis, 1982) and extending the analysis of DMR, we express the higher-order derivatives of period density-ratios in terms of the conditional expectation of the derivatives of period *complete-data* log-density, and show that its sequence can be approximated by a stationary, ergodic and square integrable martingale difference sequence by conditioning on the infinite past. Bounding their moments and applying a law of large numbers and a central limit theorem uniformly over parameter values using Proposition 6 and the result from Hansen (1996), we show that the regularity conditions for our DQM expansion hold in Markov regime switching models.

We first derive the asymptotic null distribution of the LRTS for testing $H_0 : M = 1$ against $H_A : M = 2$. When the regime-specific density function is not normal, the log-likelihood function is locally approximated by a quadratic function of the *second-order* polynomials of reparameterized parameters. When the density function is normal, the required order of expansion depends on the value of unidentified parameter; in particular, when the latent regime variables are serially uncorrelated, the model reduces to a finite mixture normal model in which the fourth-order DQM expansion is necessary to derive a quadratic approximation of the log-likelihood function. We expand the log-likelihood with respect to a judiciously chosen polynomials of reparameterized parameters—which involves the *fourth-order* polynomials—to obtain a uniform approximation of the log-likelihood function in quadratic-form, and derive the asymptotic null distribution of LRTS by maximizing the quadratic form under a set of constraints, each of which is locally approximated by a cone.

To derive the asymptotic null distribution of the LRTS for testing $H_0 : M = M_0$ against $H_A : M = M_0 + 1$ for $M_0 \geq 2$, we partition a set of parameters that describes the true null model in the alternative model into M_0 subsets, each of which corresponds to a specific way of generating the null model. We show that the asymptotic distribution of the LRTS for testing $H_0 : M = M_0$ is characterized by the maximum of M_0 random variables, each of which represents the LRTS for testing each of M_0 subsets.

The remainder of this paper is organized as follows. After introducing notation and assumptions in section 2, we discuss the degeneracy of Fisher information matrix and the loss of identifiability in regime switching model in section 3. Section 4 establishes the DQM-type expansion. Section 5 presents the uniform convergence for the derivatives of density-ratios. Sections 6 and 7 derives the asymptotic distribution of the LRTS. Section 8 collects the proofs and the auxiliary results.

2 Notation and assumptions

Let := denote "equals by definition." Let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes indexed by π for some space Π. For a matrix B, let $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ be the smallest and the largest eigenvalue of B, respectively. For a k-dimensional vector $x = (x_1, \ldots, x_k)'$ and a matrix B, define $|x| := \sqrt{x'x}$ and $|B| := \sqrt{\lambda_{\max}(B'B)}$. For $k \times 1$ vector $a = (a_1, \ldots, a_k)'$ and a function $f(a)$, let $\nabla_a^j f(a)$ denote a collection of derivatives of the form $(\partial^j / \partial a_{i_1} \partial a_{i_2} \ldots \partial a_{i_j}) f(a)$. The notation $||\cdot||$ is used for the L^2 norm. Let $\mathbb{I}\{A\}$ denote an indicator function that takes value 1 when A is true and 0 otherwise. $\mathcal{C}$ denotes a generic nonnegative finite constant whose value may change from one expression to another. Let $\lfloor x \rfloor$ denote the largest integer less than or equal to x, and define $(x)_+ := \max\{x, 0\}$. The proof of all the propositions and lemmas is presented in the appendix.

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 values in a set $\mathcal{X}_M \times \mathcal{Y} \times \mathcal{W}$ with the associated Borel σ-field $\mathcal{B}(\mathcal{X}_M \times \mathcal{Y} \times \mathcal{W})$. For a stochastic process $\{Z_k\}$ and $a < b$, define $Z_a^b := (Z_a, Z_{a+1}, \ldots, Z_b)$. Denote $\overline{Y}_{k-1} := (Y_{k-1}, \ldots, Y_{k-s})$ for a fixed integer s and $\overline{Y}_a^b := (\overline{Y}_a, \overline{Y}_{a+1}, \ldots, \overline{Y}_b)$.

Assumption 1. *(a) $\{X_k\}_{k=0}^\infty$ is a first-order Markov chain with the state space $\mathcal{X}_M := \{1, 2, \ldots, M\}$. (b) For each $k \geq 1$, X_k is independent of $(X_0^{k-2}, \overline{Y}_0^{k-1}, W_0^\infty)$ given X_{k-1}. (c) For each $k \geq 1$, $Y_k \in \mathcal{Y} \subset \mathbb{R}^{q_y}$ is conditionally independent of $(X_0^{k-1}, \overline{Y}_0^{k-2}, W_0^{k-1}, W_{k+1}^\infty)$ given $(\overline{Y}_{k-1}, W_k, X_k)$. (d) W_1^∞ is conditionally independent of $(\overline{Y}_0, X_0)$ given W_0. (e) $\{(X_k, Y_k, W_k)\}_{k=0}^\infty$ is a strictly stationary ergodic process.*

The Markov chain $\{X_k\}$ is not observable and is called the *regime*. The integer M represents the number of regimes specified in the model. For each $\vartheta_M = (\vartheta_{M,x}', \vartheta_{M,y}')'$, we denote the transition probability of X_k by $q_{\vartheta_{M,x}}(x_{k-1}, x_k) := \mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1})$ and the conditional density function of Y_k given $(\overline{Y}_{k-1}, W_k, X_k)$ by $g_{\vartheta_{M,y}}(y_k | \overline{y}_{k-1}, w_k, x_k) = \sum_{j \in \mathcal{X}_M} \mathbb{I}\{x_k = j\} f(y_k | \overline{y}_{k-1}, w_k; \gamma, \theta_j)$ so that $f(y_k | \overline{y}_{k-1}, w_k; \gamma, \theta_j)$ is the conditional density of y_k given $(\overline{y}_{k-1}, w_k)$ when $x_k = j$. Here, $\vartheta_{M,x}$ contains the parameter $p_{ij} := q_{\vartheta_{M,x}}(i, j)$ for $i = 1, \ldots, M$ and $j = 1, \ldots, M-1$, and $q_{\vartheta_{M,x}}(i, M)$ is determined by $q_{\vartheta_{M,x}}(i, M) = 1 - \sum_{j=1}^{M-1} p_{ij}$. $\vartheta_{M,y} = (\theta_1', \ldots, \theta_M', \gamma')'$, where γ is the structural parameter that does not vary across regimes and θ_j is the regime-specific parameter that may vary across regimes. Let

$$
\begin{aligned}
p_\vartheta(y_k, x_k | \overline{y}_{k-1}, w_k, x_{k-1}) &:= q_{\vartheta_x}(x_{k-1}, x_k) g_{\vartheta_y}(y_k | \overline{y}_{k-1}, w_k, x_k) \\
&= q_{\vartheta_x}(x_{k-1}, x_k) \sum_{j \in \mathcal{X}_M} \mathbb{I}\{x_k = j\} f(y_k | \overline{y}_{k-1}, w_k; \gamma, \theta_j).
\end{aligned}
$$

The parameter ϑ_M belongs to $\Theta_M = \Theta_{M,x} \times \Theta_{M,y}$, a compact subset of $\mathbb{R}^{q_M}$, and the true parameter value is denoted by ϑ_M^*.

We make the following assumptions that correspond to (A1)-(A3) of DMR.

Assumption 2. *(a)* $0 < \sigma_- := \inf_{\vartheta_{M,x}\in\Theta_{M,x}} \min_{x,x'\in\mathcal{X}_M} q_{\vartheta_{M,x}}(x,x')$ *and* $\sigma_+ := \sup_{\vartheta_{M,x}\in\Theta_{M,x}} \max_{x,x'\in\mathcal{X}_M} q_{\vartheta_{M,x}}(x,x') < 1$ *for each* M. *(b) For all* $y' \in \mathcal{Y}$, $\overline{y} \in \mathcal{Y}^s$, *and* $w \in \mathcal{W}$, $0 < \inf_{\vartheta_y\in\Theta_{M,y}} \sum_{x\in\mathcal{X}_M} g_{\vartheta_{M,y}}(y'|\overline{y},w,x)$ *and* $\sup_{\vartheta_{M,y}\in\Theta_{M,y}} \sum_{x\in\mathcal{X}_M} g_{\vartheta_{M,y}}(y'|\overline{y},w,x) < \infty$. *(c)* $b_+ := \sup_{\vartheta_{M,y}\in\Theta_y} \sup_{\overline{y}_0,y_1,w,x} g_{\vartheta_{M,y}}(y_1|\overline{y}_0,w,x) < \infty$ *and* $\mathbb{E}_{\vartheta^*}(|\log b_-(\overline{Y}_0,W,Y_1)|) < \infty$, *where* $b_-(\overline{y}_0,w,y_1) := \inf_{\vartheta_{M,y}\in\Theta_{M,y}} \sum_{x\in\mathcal{X}_M} g_{\vartheta_{M,y}}(y_1|\overline{y}_0,w,x)$.

As discussed in p. 2260 of DMR, Assumption 2(a) implies that the Markov chain $\{X_k\}$ has a unique invariant distribution and uniformly ergodic for all $\theta_{M,x} \in \Theta_{M,x}$. For notational brevity, we drop the subscript M from $\mathcal{X}_M$, ϑ_M, Θ_M, etc., unless it is important to clarify the specific value of M. Assumption 1(b)(c) imply that $\{Z_k\}_{k=0}^\infty := \{(X_k,\overline{Y}_k)\}_{k=0}^\infty$ is a Markov chain on $\mathcal{Z} := \mathcal{X}\times\mathcal{Y}^s$ given $\{W_k\}_{k=0}^\infty$, and Z_k is conditionally independent of $(Z_0^{k-2}, W_0^{k-1}, W_{k+1}^\infty)$ given (Z_{k-1}, W_k). Consequently, Lemma 1, Corollary 1, and Lemma 9 of DRM go through even in the presence of $\{W_k\}_{k=0}^\infty$. Because $\{(Z_k,W_k)\}_{k=0}^\infty$ is stationary, we can and will extend $\{(Z_k,W_k)\}_{k=0}^\infty$ to a stationary process $\{(Z_k,W_k)\}_{k=-\infty}^\infty$ with doubly infinite time. DMR use $\overline{\mathbb{P}}_\vartheta$ and $\overline{\mathbb{E}}_\vartheta$ to denote probability and expectation under stationarity on $\{Z_k\}_{k=\infty}^\infty$, because their Section 7 deals with the case when Z_0 is drawn from an arbitrary distribution. Because we assume $\{Z_k\}_{k=\infty}^\infty$ is stationary throughout this paper, we use notations such as $\mathbb{P}_\vartheta$ and $\mathbb{E}_\vartheta$ without an overline for simplicity.

The density function of Y_1^n given $X_0 = x_0$, $\overline{Y}_0$ and W_1^n for the model with M regimes is

$$p_{\vartheta_M}(Y_1^n|\overline{Y}_0, W_1^n, x_0) = \sum_{x_1^n\in\mathcal{X}_M^n} \prod_{k=1}^n p_{\vartheta_M}(Y_k, x_k|\overline{Y}_{k-1}, W_k, x_{k-1}). \tag{1}$$

Define the conditional log-likelihood function and log-likelihood function under stationarity as

$$\ell_n(\vartheta,x_0) := \log p_\vartheta(Y_1^n|\overline{Y}_0,W_1^n,x_0) = \sum_{k=1}^n \log p_\vartheta(Y_k|\overline{Y}_0^{k-1},W_1^n,x_0),$$
$$\ell_n(\vartheta) := \log p_\vartheta(Y_1^n|\overline{Y}_0,W_1^n) = \sum_{k=1}^n \log p_\vartheta(Y_k|\overline{Y}_0^{k-1},W_1^n).$$

Note that

$$\begin{aligned} &p_\vartheta(Y_k|\overline{Y}_0^{k-1},W_1^n,x_0) - p_\vartheta(Y_k|\overline{Y}_0^{k-1},W_1^n) \\ &= \sum_{(x_{k-1},x_k)\in\mathcal{X}^2} p_\vartheta(Y_k,x_k|\overline{Y}_{k-1},W_k,x_{k-1}) \times \left(\mathbb{P}_\vartheta(x_{k-1}|\overline{Y}_0^{k-1},W_1^n,x_0) - \mathbb{P}_\vartheta(x_{k-1}|\overline{Y}_0^{k-1},W_1^n)\right). \end{aligned}$$

and $\mathbb{P}_\vartheta(x_{k-1}|\overline{Y}_0^{k-1},W_1^n) = \sum_{x_0\in\mathcal{X}} \mathbb{P}_\vartheta(x_{k-1}|\overline{Y}_0^{k-1},W_1^n,x_0)\mathbb{P}_\vartheta(x_0|\overline{Y}_0^{k-1},W_1^n)$. Let $\rho := 1-\sigma_-/\sigma_+ \in (0,1)$. Lemma 9 in the appendix implies that, for all probability measures μ_1 and μ_2 on $\mathcal{B}(\mathcal{X})$ and all $(\overline{Y}_0^{k-1}, W_1^n)$,

$$\sup_A \left| \sum_{x_0 \in \mathcal{X}} \mathbb{P}_\vartheta(X_{k-1} \in A | \overline{Y}_0^{k-1}, W_1^n, x_0)\mu_1(x_0) - \sum_{x_0 \in \mathcal{X}} \mathbb{P}_\vartheta(X_{k-1} \in A | \overline{Y}_0^{k-1}, W_1^n, x_0)\mu_2(x_0) \right| \leq \rho^{k-1}. \tag{2}$$

Consequently, $p_\vartheta(Y_k|\overline{Y}_0^{k-1}, W_1^n, x_0) - p_\vartheta(Y_k|\overline{Y}_0^{k-1}, W_1^n)$ goes to zero at an exponential rate as $k \to \infty$. Therefore, as shown in the following proposition, the difference between $\ell_n(\theta, x_0)$ and $\ell_n(\theta)$ is bounded by a deterministic constant, and the maximum of $\ell_n(\vartheta, x_0)$ and the maximum of $\ell_n(\vartheta)$ are asymptotically equivalent.

Proposition 1. *Under Assumptions 1-2, for all $x_0 \in \mathcal{X}$,*

$$\sup_{\vartheta \in \Theta} |\ell_n(\vartheta, x_0) - \ell_n(\vartheta)| \leq 1/(1-\rho)^2, \quad \overline{\mathbb{P}}_{\theta^*} - a.s.$$

As discussed on p. 2263 of DRM, the stationary density $p_\vartheta(Y_k|\overline{Y}_0^{k-1}, W_1^n)$ is not available in closed form for some models with autoregression. For this reason, we consider the log-likelihood function when the initial distribution of X_0 follows some arbitrary distribution $\xi_M \in \Xi_M := \{\xi(x_0)_{x_0 \in \mathcal{X}_M} : \xi(x_0) \geq 0 \text{ and } \sum_{x_0 \in \mathcal{X}_M} \xi(x_0) = 1\}$.

Define the Maximum Likelihood Estimator (MLE, hereafter), $\hat{\vartheta}_{M,\xi_M}$, by the maximizer of the conditional log likelihood

$$\ell_n(\vartheta_M, \xi_M) := \log \left(\sum_{x_0=1}^{M} p_{\vartheta_M}(Y_1^n | \overline{Y}_0, W_1^n, x_0)\xi_M(x_0) \right), \tag{3}$$

where $p_{\vartheta_M}(Y_1^n | \overline{Y}_0, W_1^n, x_0)$ is given in (1). We define the *number of regimes* by the smallest number M such that the data density admits the representation (3). Our objective is to test

$$H_0 : M = M_0 \quad \text{against} \quad H_A : M = M_0 + 1.$$

Define the likelihood ratio test statistic (LRTS, hereafter) for testing H_0 as $2[\max_{\vartheta_{M_0+1} \in \Theta_{M_0+1}} \ell_n(\vartheta_{M_0+1}, \xi_{M_0+1}) - \max_{\vartheta_{M_0} \in \Theta_{M_0}} \ell_n(\vartheta_{M_0}, \xi_{M_0})]$.

3 Degeneracy of Fisher information matrix and non-identifiability under the null hypothesis

Consider testing $H_0 : M = 1$ against $H_A : M = 2$ in a two-regime model based on the LRTS. The null hypothesis can be written as $H_0 : \theta_1^* = \theta_2^*$.[1] When $\theta_1 = \theta_2$, the parameter $\vartheta_{2,x}$ is not identified

[1] The null hypothesis of $H_0 : M = 1$ also holds when $p_{11} = 1$ or $p_{22} = 1$. We impose Assumption 2(a) to exclude $p_{11} = 1$ or $p_{22} = 1$ from the parameter space because the log likelihood function is unbounded as p_{11} or p_{22} tends to zero (Gassiat and Keribin, 2000).

because Y_k has the same distribution across regimes. Furthermore, Section 6 shows that, when $\theta_1 = \theta_2$, the scores with respect to θ_1 and θ_2 are linearly dependent so that the Fisher information matrix is degenerate

The log-likelihood function of Markov switching models with normal density has further degeneracy. For example, in a two-regime model where Y_k in the j-th regime follows $N(\mu_j, \sigma_j^2)$, the model reduces to a heteroskedastic normal mixture model when $\mathbb{P}(X_k = 1|X_{k-1} = 1) = \mathbb{P}(X_k = 1|X_{k-1} = 2)$, i.e., $p_{11} = 1 - p_{22}$. Kasahara and Shimotsu (2015) show that, in a heteroskedastic normal mixture model, the first and second derivatives of the log-likelihood function are linearly dependent and the score function is a function of the fourth-order derivative. Consequently, one needs to expand the log-likelihood function four times to derive the score function.

4 Quadratic expansion under loss of identifiability

When testing the number of regimes by the LRT, a part of ϑ is not identified under the null hypothesis. Let π denote the part of ϑ that is not identified under the null, split ϑ as $\vartheta = (\psi', \pi')'$, and write $\ell_n(\vartheta, \xi) = \ell_n(\psi, \pi, \xi)$ and $\ell_n(\vartheta) = \ell_n(\psi, \pi)$. For example, in testing $H_0 : M = 1$ against $H_A : M = 2$, we have $\psi = \vartheta_{2,y}$ and $\pi = \vartheta_{2,x}$. We also use p_ϑ and $p_{\psi\pi}$ interchangeably.

Denote the true parameter value of ψ by ψ^*, and denote the set of (ψ, π) corresponding to the null hypothesis by $\Gamma^* = \{(\psi, \pi) \in \Theta : \psi = \psi^*\}$. Let t_ϑ be a continuous function of ϑ such that $t_\vartheta = 0$ if and only if $\psi = \psi^*$. For $\varepsilon > 0$, define a neighborhood of Γ^* by

$$\mathcal{N}_\varepsilon := \{\vartheta \in \Theta : |t_\vartheta| < \varepsilon\}.$$

When the MLE is consistent, the asymptotic distribution of the LRTS is determined by the local properties of the likelihood functions in $\mathcal{N}_\varepsilon$.

We establish a general quadratic expansion of the log-likelihood function $\ell_n(\psi, \pi, \xi)$ around $\ell_n(\psi^*, \pi, \xi)$ that expresses $\ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi)$ as a quadratic function of t_ϑ. Once we derive a quadratic expansion, the asymptotic distribution of the LRTS can be characterized by taking its supremum with respect to t_ϑ under an appropriate constraint and using the results of Andrews (1999, 2001).

Denote the conditional density-ratio by

$$l_{\vartheta k x_0} := \frac{p_{\psi\pi}(Y_k|\overline{Y}_0^{k-1}, W_1^n, x_0)}{p_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}, W_1^n, x_0)}, \tag{4}$$

so that $\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) = \sum_{k=1}^n \log l_{\vartheta k x_0}$. We assume that $l_{\vartheta k x_0}$ can be expanded around $l_{\vartheta^* k x_0} = 1$ as follows. With slight abuse of notation, let $P_n(f_k) := n^{-1} \sum_{k=1}^n f_k$ and $\nu_n(f_k) := n^{-1/2} \sum_{k=1}^n [f_k - \mathbb{E}_{\vartheta^*}(f_k)]$.

Assumption 3. *For all $k = 1, \ldots, n$, $l_{\vartheta k x_0} - 1$ admits an expansion*

$$l_{\vartheta k x_0} - 1 = t_\vartheta' s_{\pi k} + r_{\vartheta k} + u_{\vartheta k x_0}, \tag{5}$$

where $(s_{\pi k}, r_{\vartheta k}, u_{\vartheta k x_0})$ satisfy, for some $C \in (0, \infty)$, $\varepsilon > 0$, and $\rho \in (0, 1)$, (a) $\mathbb{E}_{\vartheta^*} \sup_{\pi \in \Theta_\pi} |s_{\pi k}|^2 < C$, (b) $\sup_{\pi \in \Theta_\pi} |P_n(s_{\pi k} s_{\pi k}') - \mathcal{I}_\pi| = o_p(1)$, where $\sup_{\pi \in \Theta_\pi} \lambda_{max}(\mathcal{I}_\pi) < \infty$, (c) $\mathbb{E}_{\vartheta^*}[\sup_{\vartheta \in \mathcal{N}_\varepsilon} |r_{\vartheta k}/(|t_\vartheta||\psi - \psi^*|)|^2] < \infty$, (d) $\sup_{\vartheta \in \mathcal{N}_\varepsilon}[\nu_n(r_{\vartheta k})/(|t_\vartheta||\psi - \psi^*|)] = O_p(1)$, (e) $\max_{x_0 \in \mathcal{X}} \mathbb{E}_{\vartheta^*}[\sup_{\vartheta \in \mathcal{N}_\varepsilon}(|u_{\vartheta k x_0}|/|\psi - \psi^*|)^2] \leq C\rho^{k-1}$, (f) $0 < \inf_{\pi \in \Theta_\pi} \lambda_{min}(\mathcal{I}_\pi)$, (g) there exists a stochastic process $Z_n(\vartheta)$ such that $\sup_{\vartheta \in \mathcal{N}_\varepsilon} |\nu_n(s_{\pi k}) - Z_n(\vartheta)| = o_p(1)$ and $\sup_{\vartheta \in \mathcal{N}_\varepsilon} |Z_n(\vartheta)| = O_p(1)$.

In Section 6, we derive an expansion (5) for various regime switching models that involves the higher order derivatives of density-ratios, $\nabla^j l_{\vartheta k x_0}$, and derive the asymptotic distribution of the LRTS.

We first establish an expansion $\ell_n(\psi, \pi, x_0)$ that holds for any $x_0 \in \mathcal{X}$. The following proposition expands $\ell_n(\psi, \pi, x_0)$ in a neighborhood $\mathcal{N}_{c/\sqrt{n}}$ for any $c > 0$.

Proposition 2. *Suppose that Assumption 3(a)-(e) holds. Then, for any $x_0 \in \mathcal{X}$ and for all $c > 0$,*

$$\sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} \left| \ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) - \sqrt{n} t_\vartheta' \nu_n(s_{\pi k}) + n t_\vartheta' \mathcal{I}_\pi t_\vartheta / 2 \right| = o_p(1).$$

The next proposition expands $\ell_n(\psi, \pi, x_0)$ in $A_n(x_0) := \{\vartheta \in \mathcal{N}_\varepsilon : \ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) \geq 0\}$. This proposition is useful for deriving the asymptotic distribution of the LRTS because a consistent MLE is in $A_n(x_0)$ by definition and it is difficult to find a uniform approximation of $\ell_n(\psi, \pi, x_0)$ up to an $o_p(1)$ term in $\mathcal{N}_\varepsilon$.

Proposition 3. *Suppose that Assumption 3 holds. Then, for any $x_0 \in \mathcal{X}$, (a) $\sup_{\vartheta \in A_n(x_0)} |t_\vartheta| = O_p(n^{-1/2})$; (b)*

$$\sup_{\vartheta \in A_n(x_0)} \left| \ell_n(\psi, \pi^*, x_0) - \ell_n(\psi, \pi, x_0) - \sqrt{n} t_\vartheta' Z_n(\vartheta) + n t_\vartheta' \mathcal{I}_\pi t_\vartheta / 2 \right| = o_p(1).$$

The following corollaries of Propositions 2 and 3 show that $\ell_n(\vartheta, \xi)$ defined in (3) admits a similar expansion to $\ell_n(\vartheta, x_0)$ for all ξ. Consequently, the asymptotic distribution of the LRTS does not depend on ξ, and $\ell_n(\vartheta, \xi)$ may be maximized in ϑ while fixing ξ or jointly in ϑ and ξ. Let $A_n := \{\vartheta \in \mathcal{N}_\varepsilon : \max_{x_0 \in \mathcal{X}}(\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0)) \geq 0\}$, which includes a consistent MLE with any ξ.

Corollary 1. *(a) Under the assumptions of Proposition 2, we have $\sup_{\xi \in \Xi} \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} |\ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n} t_\vartheta' \nu_n(s_{\pi k}) + n t_\vartheta' \mathcal{I}_\pi t_\vartheta / 2| = o_p(1)$ for all $c > 0$. (b) Under the assumptions of Proposition 3, we have $\sup_{\xi \in \Xi} \sup_{\vartheta \in A_n} |t_\vartheta| = O_p(n^{-1/2})$ and $\sup_{\xi \in \Xi} \sup_{\vartheta \in A_n} |\ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n} t_\vartheta' Z_n(\vartheta) + n t_\vartheta' \mathcal{I}_\pi t_\vartheta / 2| = o_p(1)$.*

5 Uniform convergence of the derivatives of the log-density and the density-ratios

In this section, we establish approximations that enable us to apply Propositions 2 and 3 and Corollary 1 to the log-likelihood function of regime switching models. Because of the presence of singularity, the expansion (5) involves higher-order derivatives of the density-ratios $\nabla_\psi^j l_{\vartheta k x_0}$ with $j \geq 2$. First, we express $\nabla_\psi^j l_{\vartheta k x_0}$ in terms of the conditional expectation of the derivatives of the *complete data* log-density by extending the Missing Information Principle (Woodbury, 1971; Louis, 1982) and the analysis of DMR to higher-order derivatives. We then show that a sequence $\{\nabla_\psi^j l_{\vartheta k x_0}\}_{k=0}^\infty$ can be approximated by a stationary martingale difference sequence by conditioning on the infinite past $\overline{Y}_{-\infty}^{k-1}$ in place of $\overline{Y}_0^{k-1}$. The leading term satisfies the assumptions on $s_{\pi k}$ in (5) because it is a stationary martingale difference sequence, and the resulting approximation error is sufficiently small to satisfy the assumptions on the remainder terms $r_{\vartheta k}$ and $u_{\vartheta k x_0}$.

For notational brevity, we assume ϑ is scalar in this section. Adaptations to vector-valued ϑ are straightforward but need more tedious notation. We first collect notations. Define $\overline{Z}_{k-1}^k := (X_{k-1}, \overline{Y}_{k-1}, W_k, X_k, Y_k)$ and denote the derivative of the complete data log-density by

$$\phi^i(\vartheta, \overline{Z}_{k-1}^k) := \nabla^i \log p_\vartheta(Y_k, X_k | \overline{Y}_{k-1}, W_k, X_{k-1}), \quad i \geq 1. \tag{6}$$

We use a short-handed notation $\phi_{\vartheta k}^i := \phi^i(\vartheta, \overline{Z}_{k-1}^k)$. We also suppress the superscript 1 from $\phi_{\vartheta k}^1$, so that $\phi_{\vartheta k} = \phi_{\vartheta k}^1$. For random variables $V_1, \ldots, V_q$ and a conditioning set $\mathcal{F}$, define the central conditional moment of $(V_1, \ldots, V_q)$ as

$$\mathbb{E}_\vartheta^c\left[V_1, \ldots, V_q | \mathcal{F}\right] := \mathbb{E}_\vartheta\left[(V_1 - \mathbb{E}_\vartheta[V_1|\mathcal{F}]) \cdots (V_q - \mathbb{E}_\vartheta[V_q|\mathcal{F}]) | \mathcal{F}\right],$$

so that $\mathbb{E}_\vartheta^c\left[\phi_{\vartheta k_1}\phi_{\vartheta k_2}\phi_{\vartheta k_3}|\mathcal{F}\right] := \mathbb{E}_\vartheta\left[\left(\phi_{\vartheta k_1} - \mathbb{E}_\vartheta[\phi_{\vartheta k_1}|\mathcal{F}]\right)\left(\phi_{\vartheta k_2} - \mathbb{E}_\vartheta[\phi_{\vartheta k_2}|\mathcal{F}]\right)\left(\phi_{\vartheta k_3} - \mathbb{E}_\vartheta[\phi_{\vartheta k_3}|\mathcal{F}]\right)|\mathcal{F}\right]$.

Let $\mathcal{I}(j) = (i_1, \ldots, i_j)$ denote a sequence of positive integer with j elements, let $\sigma(\mathcal{I}(j))$ denote all the unique permutations of $(i_1, \ldots, i_j)$, and let $|\sigma(\mathcal{I}(j))|$ denote the number of such unique permutations. For example, if $\mathcal{I}(3) = (2,1,1)$, then $\sigma(\mathcal{I}(3)) = \{(2,1,1),(1,2,1),(1,1,2)\}$ and $|\sigma(\mathcal{I}(3))| = 3$; if $\mathcal{I}(3) = (1,1,1)$, then $\sigma(\mathcal{I}(3)) = (1,1,1)$ and $|\mathcal{I}(3)| = 1$. Let $\mathcal{T}(j) = (t_1, \ldots, t_j)$ for

$j = 1, \ldots, 6$. For a conditioning set $\mathcal{F}$, define symmetrized central conditional moments as

$$\Phi^{\mathcal{I}(1)}_{\vartheta\mathcal{T}(1)}[\mathcal{F}] := \mathbb{E}_\vartheta\left[\phi^{i_1}_{\vartheta t_1}\middle|\mathcal{F}\right], \quad \Phi^{\mathcal{I}(2)}_{\vartheta\mathcal{T}(2)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(2))|}\sum_{(\ell_1,\ell_2)\in\sigma(\mathcal{I}(2))}\mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\middle|\mathcal{F}\right],$$

$$\Phi^{\mathcal{I}(3)}_{\vartheta\mathcal{T}(3)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(3))|}\sum_{(\ell_1,\ell_2,\ell_3)\in\sigma(\mathcal{I}(3))}\mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}\middle|\mathcal{F}\right],$$

$$\begin{aligned}\Phi^{\mathcal{I}(4)}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(4))|}\sum_{(\ell_1,\ldots,\ell_4)\in\sigma(\mathcal{I}(4))}\Big(&\mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}\phi^{\ell_4}_{\vartheta t_4}\middle|\mathcal{F}\right] - \mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\middle|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi^{\ell_3}_{\vartheta t_3}\phi^{\ell_4}_{\vartheta t_4}\middle|\mathcal{F}\right]\\ &-\mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_3}_{\vartheta t_3}\middle|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_4}_{\vartheta t_4}\middle|\mathcal{F}\right] - \mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_4}_{\vartheta t_4}\middle|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}\middle|\mathcal{F}\right]\Big),\end{aligned} \tag{7}$$

and Section 8.1 in appendix defines $\Phi^{\mathcal{I}(5)}_{\vartheta\mathcal{T}(5)}[\mathcal{F}]$ and $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$. Note that these moments are symmetric with respect to $(t_1, \ldots, t_j)$. Define, for $j = 1, 2, \ldots, 6$, $k \geq 1$, $m \geq 0$ and $x \in \mathcal{X}$,

$$\begin{aligned}\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta) := &\sum_{\mathcal{T}(j)\in\{-m+1,\ldots,k\}^j}\Phi^{\mathcal{I}(j)}_{\vartheta\mathcal{T}(j)}\left[\overline{Y}^k_{-m}, W^n_{-m+1}, X_{-m} = x\right]\\ &- \sum_{\mathcal{T}(j)\in\{-m+1,\ldots,k-1\}^j}\Phi^{\mathcal{I}(j)}_{\vartheta\mathcal{T}(j)}\left[\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x\right],\end{aligned} \tag{8}$$

where $\sum_{\mathcal{T}(j)\in\{-m+1,\ldots,k\}^j}$ denotes $\sum_{t_1=-m+1}^k\sum_{t_2=-m+1}^k\cdots\sum_{t_j=-m+1}^k$, and $\sum_{\mathcal{T}\in\{-m+1,\ldots,k-1\}^j}$ is defined similarly. Define $\overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\theta)$ analogously to $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$ by replacing the conditioning variables $\{\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x\}$ in (8) with $\{\overline{Y}^{k-1}_{-m}, W^n_{-m+1}\}$.

For $1 \leq k \leq n$ and $m \geq 0$, let

$$\overline{p}_\vartheta(Y^k_{-m}|\overline{Y}_{-m}, W^n_{-m+1}) := \sum_{x^k_{-m}\in\mathcal{X}^{k+m+1}}\prod_{t=-m+1}^k p_\vartheta(Y_t, x_t|\overline{Y}_{t-1}, W_t, x_{t-1})\mathbb{P}_{\vartheta^*}(x_{-m}|\overline{Y}_{-m}, W^n_{-m+1}), \tag{9}$$

denote the stationary density of Y^k_{-m} conditional on $\{\overline{Y}_{-m}, W^n_{-m+1}\}$ where X_{-m} is drawn from its true conditional stationary distribution $\mathbb{P}_{\vartheta^*}(X_{-m}|\overline{Y}^{k-1}_{-m}, W^n_{-m+1})$. Let $\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}) := \overline{p}_\vartheta(Y^k_{-m}|\overline{Y}_{-m}, W^n_{-m+1})/\overline{p}_\vartheta(Y^{k-1}_{-m}|\overline{Y}_{-m}, W^n_{-m+1})$ denote the associated conditional density of Y_k given $(\overline{Y}^{k-1}_{-m}, W^n_{-m+1})$.[2]

For $j = 1, 2\ldots, 6$, $1 \leq k \leq n$, $m \geq 0$ and $x \in \mathcal{X}$, define the derivatives of log densities and density-ratios by

$$\nabla^j\ell_{k,m,x}(\vartheta) := \nabla^j\log p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x), \quad \nabla^j l_{k,m,x}(\vartheta) := \frac{\nabla^j p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x)}{p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x)},$$

$$\nabla^j\overline{\ell}_{k,m}(\vartheta) := \nabla^j\log\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}), \quad \text{and} \quad \nabla^j\overline{l}_{k,m}(\vartheta) := \frac{\nabla^j\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1})}{\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1})}.$$

[2] Note that DRM use the same notation $\overline{p}_\vartheta(\cdot|\overline{Y}^{k-1}_{-m})$ for a different purpose. On p. 2263 and in some other (but not all) places, DRM use $\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_0)$ to denote an (ordinary) stationary conditional distribution of Y_k.

The following proposition expresses the derivatives of log densities, $\nabla^j \ell_{k,m,x}(\vartheta)$'s, in terms of the conditional expectation of the central moments of derivatives of the complete data log-density. The first two equations are also given in DMR (p. 2272 and pp. 2276-7).

Proposition 4. *For all $1 \leq k \leq n$, $m \geq 0$, and $x \in \mathcal{X}$,*

$$
\begin{aligned}
&\nabla^1 \ell_{k,m,x}(\vartheta) = \Delta^1_{1,k,m,x}(\vartheta), \quad \nabla^2 \ell_{k,m,x}(\vartheta) = \Delta^2_{1,k,m,x}(\vartheta) + \Delta^{1,1}_{2,k,m,x}(\vartheta),\\
&\nabla^3 \ell_{k,m,x}(\vartheta) = \Delta^3_{1,k,m,x}(\vartheta) + 3\Delta^{2,1}_{2,k,m,x}(\vartheta) + \Delta^{1,1,1}_{3,k,m,x}(\vartheta),\\
&\nabla^4 \ell_{k,m,x}(\vartheta) = \Delta^4_{1,k,m,x}(\vartheta) + 4\Delta^{3,1}_{2,k,m,x}(\vartheta) + 3\Delta^{2,2}_{2,k,m,x}(\vartheta) + 6\Delta^{2,1,1}_{3,k,m,x}(\vartheta) + \Delta^{1,1,1,1}_{4,k,m,x}(\vartheta),\\
&\nabla^5 \ell_{k,m,x}(\vartheta) = \Delta^5_{1,k,m,x}(\vartheta) + 5\Delta^{4,1}_{2,k,m,x}(\vartheta) + 10\Delta^{3,2}_{2,k,m,x}(\vartheta) + 10\Delta^{3,1,1}_{3,k,m,x}(\vartheta) + 15\Delta^{2,2,1}_{3,k,m,x}(\vartheta)\\
&\quad + 10\Delta^{2,1,1,1}_{4,k,m,x}(\vartheta) + \Delta^{1,1,1,1,1}_{5,k,m,x}(\vartheta),\\
&\nabla^6 \ell_{k,m,x}(\vartheta) = \Delta^6_{1,k,m,x}(\vartheta) + 6\Delta^{5,1}_{2,k,m,x}(\vartheta) + 15\Delta^{4,2}_{2,k,m,x}(\vartheta) + 10\Delta^{3,3}_{2,k,m,x}(\vartheta) + 15\Delta^{4,1,1}_{3,k,m,x}(\vartheta)\\
&\quad + 60\Delta^{3,2,1}_{3,k,m,x}(\vartheta) + 5\Delta^{2,2,2}_{3,k,m,x}(\vartheta) + 20\Delta^{3,1,1,1}_{4,k,m,x}(\vartheta) + 45\Delta^{2,2,1,1}_{4,k,m,x}(\vartheta) + 15\Delta^{2,1,1,1,1}_{5,k,m,x}(\vartheta) + \Delta^{1,1,1,1,1,1}_{6,k,m,x}(\vartheta).
\end{aligned}
$$

Further, the above holds when $\nabla^j \ell_{k,m,x}(\vartheta)$ and $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$ are replaced with $\nabla^j \overline{\ell}_{k,m}(\vartheta)$ and $\overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta)$.

The following assumption corresponds to (A6)-(A8) of DMR and are tailored to our setting where some elements of ϑ^*_x are not identified. Note that Assumptions (A6)-(A7) of DMR pertaining to $q_{\vartheta_x}(x, x')$ hold in our case because p_{ij}'s are bounded away from 0 and 1. Let $G_{\vartheta k} := \sum_{x_k \in \mathcal{X}} g_{\vartheta_y}(Y_k|\overline{Y}_{k-1}, W_k, x_k)$. $G_{\vartheta k}$ satisfies Assumption 4(b) in general when $\mathcal{N}^*$ is sufficiently small.

Assumption 4. *There exists a positive real δ such that on $\mathcal{N}^* := \{\vartheta \in \Theta : |\vartheta_y - \vartheta^*_y| \leq \delta\}$ the following conditions hold: (a) For all $(\overline{y}, y', w, x) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W} \times \mathcal{X}$, the function $\vartheta_y \mapsto g_{\vartheta_y}(y'|\overline{y}, w, x)$ is six times continuously differentiable on $\mathcal{N}^*$. (b) $\mathbb{E}_{\vartheta^*}[\sup_{\vartheta \in \mathcal{N}^*} \sup_{x \in \mathcal{X}} |\nabla^j_{\vartheta_y} \log g_{\vartheta_y}(Y_1|\overline{Y}_0, W, x)|^{2q_j}] < \infty$ for $j = 1, 2, \ldots, 6$ and $\mathbb{E}_{\vartheta^*} \sup_{\vartheta \in \mathcal{N}^*} |G_{\vartheta k}/G_{\vartheta^* k}|^{q_g} < \infty$ with $q_1 = 6q_0, q_2 = 5q_0, \ldots, q_6 = q_0$, where $q_0 = (1 + \varepsilon)\max\{2, dim(\vartheta)\}$ and $q_g = (1 + \varepsilon)\max\{2, dim(\vartheta)\}/\varepsilon$ for some $\varepsilon > 0$. (c) For almost all $(\overline{y}, y', w_1^n) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}^n$, there exists a function $f_{\overline{y}, y', w_1^n} : \mathcal{X} \to \mathbb{R}^+$ such that $\sup_{\vartheta \in \mathcal{N}^*} g_{\vartheta_y}(y'|\overline{y}, w_1^n, x) \leq f_{\overline{y}, y', w_1^n}(x) < \infty$ and, for almost all $(\overline{y}, w_1^n, x) \in \mathcal{Y}^s \times \mathcal{W}^n \times \mathcal{X}$, for $j = 1, 2, \ldots, 6$, there exist functions $f^j_{\overline{y}, w_1^n, x} : \mathcal{Y} \to \mathbb{R}^+$ in L^1 such that $|\nabla^j_\vartheta g_{\vartheta_y}(y'|\overline{y}, w_1^n, x)| \leq f^j_{\overline{y}, w_1^n, x}(y')$ for all $\vartheta \in \mathcal{N}^*$.*

Lemma 3 in the appendix shows that, for all $x \in \mathcal{X}$ and $1 \leq k \leq n$ and a suitably defined $r_{\mathcal{I}(j)}$, $\{\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)\}_{m \geq 0}$ is a uniform Cauchy sequence in $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ that converges uniformly with respect to $\vartheta \in \mathcal{N}^*$ $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ to a random variable that does not depend on x. Then, in view of Proposition 4 and Lemma 3, $\{\nabla^j \ell_{k,m,x}(\vartheta)\}_{m \geq 0}$ and $\{\nabla^j \overline{\ell}_{k,m}(\vartheta)\}_{m \geq 0}$ converge to $\nabla^j \ell_{k,\infty}(\vartheta)$ uniformly with respect to $\vartheta \in \mathcal{N}^*$ $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{r_j}(\mathbb{P}_{\vartheta^*})$ as the following proposition shows. Define $\rho := 1 - \sigma_-/\sigma_+$.

Proposition 5. *Under Assumptions 1, 2, and 4, for $j = 1, \ldots, 6$, there exist random variables $K_j, M_j \in L^{r_j}(\mathbb{P}_{\vartheta^*})$ such that, for all $1 \leq k \leq n$ and $m' \geq m \geq 0$,*

$$(a) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta) - \nabla^j \bar{\ell}_{k,m}(\vartheta)| \leq K_j (k+m)^6 \rho^{(k+m-1)/12}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(b) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta) - \nabla^j \ell_{k,m',x}(\vartheta)| \leq K_j (k+m)^6 \rho^{(k+m-1)/531}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(c) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta)| + \sup_{\vartheta} |\nabla^j \bar{\ell}_{k,m}(\vartheta)| + \sup_{\vartheta} |\nabla^j \ell_{k,\infty}(\vartheta)| \leq M_j, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

where $r_1 = 6q_0$, $r_2 = 3q_0$, $r_3 = 2q_0$, $r_4 = 3q_0/2$, $r_5 = 6q_0/5$, and $r_6 = q_0$.

Finally, we prove the uniform convergence of the derivatives of density-ratios by expressing them as polynomial functions of the derivatives of log-density and applying Proposition 5 and the Hölder's inequality. The following Proposition 6(a)(b) imply that $\{\nabla^j l_{k,m,x}(\vartheta)\}_{m \geq 0}$ and $\{\nabla^j \bar{l}_{k,m}(\vartheta)\}_{m \geq 0}$ converge to $\nabla^j l_{k,\infty}(\vartheta)$ uniformly with respect to $x \in \mathcal{X}$ and $\vartheta \in \mathcal{N}^*$ $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{\max\{2, \dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$.

Proposition 6. *Under Assumptions 1, 2, and 4, for $j = 1, \ldots, 6$, there exist random variables $K_j, M_j \in L^{\max\{2, \dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$ and $\rho_* \in (0,1)$ such that, for all $1 \leq k \leq n$ and $m' \geq m \geq 0$,*

$$(a) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta) - \nabla^j \bar{l}_{k,m}(\vartheta)| \leq K_j (k+m)^6 \rho_*^{k+m-1}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(b) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta) - \nabla^j l_{k,m',x}(\vartheta)| \leq K_j (k+m)^6 \rho_*^{k+m-1}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(c) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j l_{k,0,x}(\vartheta) - \nabla^j l_{k,\infty}(\vartheta)| \leq K_j k^6 \rho_*^{k-1}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(d) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta)| + \sup_{\vartheta} |\nabla^j \bar{l}_{k,m}(\vartheta)| + \sup_{\vartheta} |\nabla^j l_{k,\infty}(\vartheta)| \leq M_j, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

When we apply Propositions 2 and 3 and Corollary 1 to regime switching models, $l_{k,0,x}(\vartheta)$ corresponds to $l_{\vartheta k x_0}$ on the left hand side of (5), and $s_{\pi k}$ and $r_{\vartheta k}$ in (5) are functions of $\nabla^j \bar{l}_{k,m}(\vartheta)$'s. The term $u_{\vartheta k x_0}$ in (5) that depends on x_0 satisfies Assumption 3(e) because of Proposition 6(a)(c). Proposition 6 and dominated convergence theorem for conditional expectations (Durrett, 2010, Theorem 5.5.9) imply that $\mathbb{E}_{\vartheta^*}[\nabla^j l_{k,\infty}(\vartheta) | \overline{Y}_{-\infty}^{k-1}] = 0$ for all $\vartheta \in \mathcal{N}^*$ and $\{\nabla^j l_{k,\infty}(\vartheta)\}_{k=-\infty}^{\infty}$ is a stationary, ergodic and square integrable martingale difference sequence so that we can apply a martingale central limit theorem. $\nabla^j l_{k,\infty}(\vartheta)$ for $j = 1, \ldots, 5$ satisfies Assumption 3(g) from Theorem 2 of Hansen (1996) because $\nabla^j l_{k,\infty}(\vartheta)$ is Lipschitz continuous with Lipschitz coefficient $\nabla^{j+1} l_{k,\infty}(\vartheta) \in L^{\max\{2, \dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$ from Proposition 6.

6 Testing homogeneity

Before developing the LRT of M_0 components, we analyze a simpler case of testing the null hypothesis $H_0 : M = 1$ against $H_A : M = 2$ when the data is from H_0. We assume that the parameter space for $\vartheta_{2,x} = (p_{11}, p_{22})'$ satisfies Assumption 2(a) with restriction $p_{11}, p_{22} \in [\epsilon, 1-\epsilon]$ for a small $\epsilon \in (0, 1/2)$, and let $\Theta_{2\epsilon}$ denote Θ_2 with this restriction. This assumption is necessary because the LRTS is unbounded under the null hypothesis when p_{11} or p_{22} tends to 1 (Gassiat and Keribin, 2000). Denote the true parameter in a one-regime model by $\vartheta_1^* := ((\theta^*)', (\gamma^*)')'$. The two-regime model gives rise to the true density $p_{\vartheta_1^*}(Y_1^n|\overline{Y}_0, x_0)$ if the parameter $\vartheta_2 = (\theta_1, \theta_2, \gamma, p_{11}, p_{22})'$ lies in a subset of the parameter space

$$\Gamma^* := \{(\theta_1, \theta_2, \gamma, p_{11}, p_{22}) \in \Theta_{2\epsilon} : \theta_1 = \theta_2 = \theta^* \text{ and } \gamma = \gamma^*\}.$$

Note that (p_{11}, p_{22}) is not identified under H_0.

Let $\ell_n(\vartheta_2, \xi_2) := \log\left(\sum_{x_0=1}^{2} p_{\vartheta_2}(Y_1^n|\overline{Y}_0, W_1^n, x_0)\xi_2(x_0)\right)$ denote the two-regime log-likelihood for a given initial distribution $\xi_2(x_0) \in \Xi_2$, and let $\hat{\vartheta}_2 := \arg\max_{\vartheta_2 \in \Theta_{2\epsilon}} \ell_n(\vartheta_2, \xi_2)$ denote the maximum likelihood estimator (MLE) of ϑ_2 given ξ_2. Because ξ_2 does not matter asymptotically, we treat ξ_2 fixed for simplicity and suppress ξ_2 from $\hat{\vartheta}_2$. Let $\hat{\vartheta}_1$ denote the one-regime MLE that maximizes the one-regime log-likelihood function $\ell_{0,n}(\vartheta_1) := \sum_{k=1}^{n} \log f(Y_k|\overline{Y}_{k-1}, W_k; \gamma, \theta)$ under the constraint $\vartheta_1 = (\theta', \gamma')' \in \Theta_1$.

We introduce the following assumption for consistency of $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$. Assumption 5(b) corresponds to Assumption (A4) of DMR. Assumption 5(c) is a standard identification condition for the one-regime model. Assumption 5(d) implies that the Kullback-Leibler divergence between $p_{\vartheta_1^*}(Y_1|\overline{Y}_{-m}^0, W_{-m}^0)$ and $p_{\vartheta_2}(Y_1|\overline{Y}_{-m}^0, W_{-m}^0)$ is 0 if and only if $\vartheta_2 \in \Gamma^*$. This assumption is similar to Assumption (A5′) in DMR.[3]

Assumption 5. *(a) Θ_1 and Θ_2 are compact. (b) For all $(x, x') \in \mathcal{X}$ and all $(\overline{y}, y', w) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}$, the function $(\theta, \gamma) \mapsto f(y'|\overline{y}_0, w; \gamma, \theta)$ is continuous. (c) If $\vartheta_1 \neq \vartheta_1^*$, then $\mathbb{P}_{\vartheta_1^*}\left(f(Y_1|\overline{Y}_0, W_1; \gamma, \theta) \neq f(Y_1|\overline{Y}_0, W_1; \gamma^*, \theta^*)\right) > 0$. (d) $\mathbb{E}_{\vartheta_1^*}[\log p_{\vartheta_2}(Y_1|\overline{Y}_{-m}^0, W_{-m}^n)]$ $= \mathbb{E}_{\vartheta_1^*}[\log p_{\vartheta_1^*}(Y_1|\overline{Y}_{-m}^0, W_{-m}^n)]$ for all $m \geq 0$ if and only if $\vartheta_2 \in \Gamma^*$.*

The following proposition shows that the MLE of ϑ_1^* and ϑ_2^* are consistent under this condition.

Proposition 7. *Suppose that Assumptions 1, 2, and 5 hold. Then, under the null hypothesis of $M = 1$, $\hat{\vartheta}_1 \to_p \vartheta_1^*$ and $\inf_{\vartheta_2 \in \Gamma^*} |\hat{\vartheta}_2 - \vartheta_2| \to_p 0$.*

We proceed to derive the asymptotic distribution of the LRTS building on the results in Sections 4 and 5. Following the notation of Section 4, we split ϑ_2 as $\vartheta_2 = (\psi, \pi)$, where π is the part of ϑ that is not identified under the null hypothesis, and the elements of ψ will be delineated later. In the current setting, $\vartheta_{2,x} = (p_{11}, p_{22})'$ is not identified under the null. Define $\varrho := \text{corr}_{\vartheta_{2,x}}(X_k, X_{k+1}) = p_{11} + p_{22} - 1$ and $\alpha := \mathbb{P}_{\vartheta_{2,x}}(X_k = 1) = (1 - p_{22})/(2 - p_{11} - p_{22})$ (see Lemma 12). The parameter spaces for ϱ and α under restriction $p_{11}, p_{22} \in [\epsilon, 1-\epsilon]$ are given by $\Theta_{\varrho\epsilon} := [-1 + 2\epsilon, 1 - 2\epsilon]$ and $\Theta_{\alpha\epsilon} := [\epsilon, 1 - \epsilon]$, respectively. Because the mapping from (p_{11}, p_{22}) to (ϱ, α) is one-to-one, we

[3] DMR derive their Assumption (A5′) from their "minimal assumption" (Assumption (A5)). We do not repeat DMR here because extending their proof (in particular, the proof of their Lemmas 5 and 6) to models with W_k is not very straightforward.

reparameterize π as $\pi := (\varrho, \alpha)' \in \Theta_{\pi\epsilon} := \Theta_{\varrho\epsilon} \times \Theta_{\alpha\epsilon}$, and let $p_{\psi\pi}(\cdot|\cdot) := p_{\vartheta_2}(\cdot|\cdot)$. Henceforth, we suppress W_1^n for notational brevity and write, for example, $p_{\psi\pi}(Y_1^n|\overline{Y}_0, W_1^n, x_0)$ as $p_{\psi\pi}(Y_1^n|\overline{Y}_0, x_0)$ and $p_{\psi\pi}(y_k, x_k|\overline{y}_{k-1}, w_k, x_{k-1})$ as $p_{\psi\pi}(y_k, x_k|\overline{y}_{k-1}, x_{k-1})$ unless confusion might arise. We apply Corollary 1 to $\ell_n(\psi, \pi, \xi_2)$ by finding a representation of $(t_\vartheta, s_{\pi k}, r_{\vartheta k}, u_{\vartheta k x_0})$ in (5) in terms of ϑ, $p_{\psi\pi}(\cdot|\cdot)$, and derivatives of $p_{\psi\pi}(\cdot|\cdot)$ and then showing that $(t_\vartheta, s_{\pi k}, r_{\vartheta k}, u_{\vartheta k x_0})$ satisfy Assumption 3. Because of the degeneracy of Fisher information matrix, $s_{\pi k}$ involves higher-order derivatives, and t_ϑ consists of functions of polynomials of (reparameterized) ϑ.

The remainder of this section derives $s_{\pi k}$ as a function of $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ with $\overline{p}_{\psi\pi}(Y_k|\overline{Y}_0^{k-1})$ defined in (9). This approximation is valid because Proposition 6 implies that $\nabla^j p_{\psi\pi}(Y_k|\overline{Y}_0^{k-1}, x_0)/p_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}, x_0) - \nabla^j \overline{p}_{\psi\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ goes to zero at an exponential rate as $k \to \infty$. Section 6.1 analyzes the case when the regime-specific distribution of y_k is not normal distribution with unknown variance. Section 6.2 analyzes the case when the regime-specific distribution y_k is normal distribution with regime-specific and unknown variance, and Section 6.3 handles normal distribution where the variance is unknown and common across regimes.

Note that, because $\overline{Y}_{-\infty}^{\infty}$ and $X_{-\infty}^{\infty}$ are independent when $\psi = \psi^*$,

$$\mathbb{P}_{\psi^*\pi}(X_{-\infty}^{\infty}|\overline{Y}_{-\infty}^{\infty}) = \mathbb{P}_{\psi^*\pi}(X_{-\infty}^{\infty}). \tag{10}$$

Define $q_k := \mathbb{I}\{X_k = 1\}$ so that $\alpha = \mathbb{E}_{\psi^*\pi}[q_k]$.

6.1 Non-normal distribution

In this section, we derive $s_{\pi k}$ when the conditional distribution of Y_k is not normal with unknown variance. We find a representation of $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ in terms of $\{\nabla^j f(Y_t|X_t; \gamma^*, \theta^*)\}_{t=1}^k$ via Louis Information Principle (Lemma 1 in the appendix). To this end, we first derive the derivatives of the complete data conditional density $p_{\vartheta_2}(y_k, x_k|\overline{y}_{k-1}, x_{k-1}) = g_{\vartheta_{2,y}}(y_k|\overline{y}_{k-1}, x_k) q_{\vartheta_{2,x}}(x_{k-1}, x_k) = \sum_{j=1}^2 \mathbb{I}\{x_k = j\} f(y_k|\overline{y}_{k-1}; \gamma, \theta_j) q_{\vartheta_{2,x}}(x_{k-1}, x_k)$.

Consider the following reparameterization. Let

$$\begin{pmatrix} \lambda \\ \nu \end{pmatrix} := \begin{pmatrix} \theta_1 - \theta_2 \\ \alpha\theta_1 + (1-\alpha)\theta_2 \end{pmatrix}, \quad \text{so that} \quad \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \nu + (1-\alpha)\lambda \\ \nu - \alpha\lambda \end{pmatrix}. \tag{11}$$

Let $\eta := (\gamma', \nu')'$ and $\psi_\alpha := (\eta', \lambda')' \in \Theta_\eta \times \Theta_\lambda$. Under the null hypothesis of one regime, the true value of ψ_α is given by $\psi_\alpha^* := (\gamma^*, \theta^*, 0)'$. Henceforth, we suppress the subscript α from ψ_α. Using this definition of ψ, let $\vartheta_2 := (\psi', \pi')' \in \Theta_\psi \times \Theta_\pi$. Using reparameterization (11) and noting that $q_k = \mathbb{I}\{x_k = 1\}$, we have $p_{\psi\pi}(y_k, x_k|\overline{y}_{k-1}, x_{k-1}) = g_\psi(y_k|\overline{y}_{k-1}, x_k) q_\pi(x_{k-1}, x_k)$ and

$$g_\psi(y_k|\overline{y}_{k-1}, x_k) = f(y_k|\overline{y}_{k-1}; \gamma, \nu + (q_k - \alpha)\lambda). \tag{12}$$

Expanding $g_\psi(y_k|\overline{y}_{k-1}, x_k)$ twice with respect to $\psi = (\gamma', \nu', \lambda')'$ and evaluating at ψ^* gives, with suppressing $\overline{y}_{k-1}$,

$$\begin{aligned}
\nabla_{\gamma} g_{\psi^*}(y_k|x_k) &= \nabla_{\gamma} f(y_k;\gamma^*,\theta^*), \quad \nabla_{\nu} g_{\psi^*}(y_k|x_k) = \nabla_{\theta} f(y_k;\gamma^*,\theta^*),\\
\nabla_{\lambda} g_{\psi^*}(y_k|x_k) &= (q_k-\alpha)\nabla_{\theta} f(y_k;\gamma^*,\theta^*),\\
\nabla_{\lambda\eta'} g_{\psi^*}(y_k|x_k) &= (q_k-\alpha)\nabla_{\theta(\gamma',\theta')} f(y_k;\gamma^*,\theta^*),\\
\nabla_{\lambda\lambda'} g_{\psi^*}(y_k|x_k) &= (q_k-\alpha)^2\nabla_{\theta\theta'} f(y_k;\gamma^*,\theta^*).
\end{aligned} \tag{13}$$

Recall $\varrho := \mathrm{corr}_{\vartheta_2^*}(q_k,q_{k+1})$. Then, it follows from (10) and Lemma 12(a) that

$$\mathbb{E}_{\vartheta^*}[q_k-\alpha|\overline{Y}_{-\infty}^n]=0, \quad \mathbb{E}_{\vartheta^*}[(q_{t_1}-\alpha)(q_{t_2}-\alpha)|\overline{Y}_{-\infty}^n]=\alpha(1-\alpha)\varrho^{t_2-t_1}, \quad t_2\ge t_1. \tag{14}$$

Henceforth, let f_k^*, ∇f_k^*, g_k^*, and ∇g_k^* denote $f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*)$, $\nabla f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*)$, $g_{\psi^*}(Y_k|\overline{Y}_{k-1},X_k)$, and $\nabla g_{\psi^*}(Y_k|\overline{Y}_{k-1},X_k)$, respectively, and similarly for $\log f_k^*$,$\nabla \log f_k^*$, $\log g_k^*$, and $\nabla \log g_k^*$.

From Louis Information Principle (Lemma 1), $\log p_{\psi\pi}(y_k,x_k|\overline{y}_{k-1},x_{k-1}) = \log g_{\psi}(y_k|\overline{y}_{k-1},x_k) + \log q_{\pi}(x_{k-1},x_k)$, and the definition of $\overline{p}_{\psi\pi}(Y_k|\overline{Y}_0^{k-1})$ in (9), we obtain

$$\frac{\nabla_{\psi}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \nabla_{\psi}\log\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \sum_{t=1}^{k}\mathbb{E}_{\vartheta^*}\left[\nabla_{\psi}\log g_t^*\middle|\overline{Y}_0^{k}\right] - \sum_{t=1}^{k-1}\mathbb{E}_{\vartheta^*}\left[\nabla_{\psi}\log g_t^*\middle|\overline{Y}_0^{k-1}\right].$$

Applying (13), (14), and $g_k^*=f_k^*$ to the right hand side gives

$$\begin{aligned}
\nabla_{\eta}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) &= \nabla_{(\gamma',\theta')'}\log f_k^*,\\
\nabla_{\lambda}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) &= 0.
\end{aligned} \tag{15}$$

Similarly, it follows from Lemma 1, (13), (14), (15), and $g_k^*=f_k^*$ that

$$\nabla_{\lambda\eta'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \nabla_{\lambda\eta'}\log\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0, \tag{16}$$

$$\begin{aligned}
&\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})\\
&= \nabla_{\lambda\lambda'}\log\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})\\
&= \sum_{t=1}^{k}\mathbb{E}_{\vartheta^*}\left[\nabla_{\lambda\lambda'}\log g_t^*\middle|\overline{Y}_0^{k}\right] - \sum_{t=1}^{k-1}\mathbb{E}_{\vartheta^*}\left[\nabla_{\lambda\lambda'}\log g_t^*\middle|\overline{Y}_0^{k-1}\right]\\
&\quad + \sum_{t_1=1}^{k}\sum_{t_2=1}^{k}\mathbb{E}_{\vartheta^*}\left[\frac{\nabla_{\lambda}g_{t_1}^*}{g_{t_1}^*}\frac{\nabla_{\lambda'}g_{t_2}^*}{g_{t_2}^*}\middle|\overline{Y}_0^{k}\right] - \sum_{t_1=1}^{k-1}\sum_{t_2=1}^{k-1}\mathbb{E}_{\vartheta^*}\left[\frac{\nabla_{\lambda}g_{t_1}^*}{g_{t_1}^*}\frac{\nabla_{\lambda'}g_{t_2}^*}{g_{t_2}^*}\middle|\overline{Y}_0^{k-1}\right]\\
&= \alpha(1-\alpha)\left[\frac{\nabla_{\theta\theta'}f_k^*}{f_k^*} + \sum_{t=1}^{k-1}\varrho^{k-t}\left(\frac{\nabla_{\theta}f_t^*}{f_t^*}\frac{\nabla_{\theta'}f_k^*}{f_k^*} + \frac{\nabla_{\theta}f_k^*}{f_k^*}\frac{\nabla_{\theta'}f_t^*}{f_t^*}\right)\right].
\end{aligned} \tag{17}$$

Note that ϱ is bounded away from -1 and 1 in $\Theta_{\varrho\epsilon}$. Because the first-order derivative with respect to λ is identically equal to zero in (15), the information on λ is provided by the second-order derivative with respect to λ in (17). Consequently, the unique elements of $\nabla_{\eta}\log\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ and

$\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \nabla_{\lambda\lambda'}\log\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ constitute the generalized score $s_{\pi k}$ in Corollary 1. Because this score is approximated by a stationary martingale difference sequence and the remainder term satisfies Assumption 4 from Lemmas 6, we can apply Corollary 1 to the likelihood ratio to derive the asymptotic distribution of the LRTS.

We collect some notations. Recall $\psi = (\eta', \lambda')'$ and $\eta = (\gamma', \nu')'$. For a $q \times 1$ vector λ and a $q \times q$ matrix s, define $q_\lambda \times 1$ vectors $v(\lambda)$ and $V(s)$ as

$$\begin{aligned} v(\lambda) &:= (\lambda_1^2, \ldots, \lambda_q^2, \lambda_1\lambda_2, \ldots, \lambda_1\lambda_q, \lambda_2\lambda_3, \ldots, \lambda_2\lambda_q, \ldots, \lambda_{q-1}\lambda_q)', \\ V(s) &:= (s_{11}/2, \ldots, s_{qq}/2, s_{12}, \ldots, s_{1q}, s_{23}, \ldots, s_{2q}, \ldots, s_{q,q-1})'. \end{aligned} \tag{18}$$

Noting that $\alpha(1-\alpha) > 0$ for $\alpha \in \Theta_{\alpha\epsilon}$, define

$$t(\psi, \pi) := \begin{pmatrix} \eta - \eta^* \\ t_\lambda(\lambda, \pi) \end{pmatrix}, \ s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda \varrho k} \end{pmatrix}, \text{ where } t_\lambda(\lambda, \pi) := \alpha(1-\alpha)v(\lambda), \ s_{\eta k} := \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix}, \tag{19}$$

and $s_{\lambda\varrho k} := V(s_{\lambda\lambda\varrho k})$ with $s_{\lambda\lambda\varrho k} := \nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/[\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})\alpha(1-\alpha)]$, namely,

$$s_{\lambda\lambda\varrho k} = \frac{\nabla_{\theta\theta'} f_k^*}{f_k^*} + \sum_{t=1}^{k-1} \varrho^{k-t} \left(\frac{\nabla_\theta f_t^*}{f_t^*} \frac{\nabla_{\theta'} f_k^*}{f_k^*} + \frac{\nabla_\theta f_k^*}{f_k^*} \frac{\nabla_{\theta'} f_t^*}{f_t^*} \right). \tag{20}$$

Here, $s_{\varrho k}$ in (19) depends on ϱ but not on α and corresponds to $s_{\pi k}$ in Corollary 1. The following proposition shows that the log-likelihood function is approximated by a quadratic function of $\sqrt{n}t(\psi, \pi)$. Let $\mathcal{N}_\varepsilon := \{\vartheta_2 \in \Theta_{2\epsilon} : |t(\psi, \pi)| < \varepsilon\}$. Let $A_{nc}(\xi) := \{\vartheta \in \mathcal{N}_\varepsilon : \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) \geq 0\} \cup \mathcal{N}_{c/\sqrt{n}}$, where we suppress the subscript 2 from ξ_2. We use this definition of $A_{nc}(\xi)$ through Sections 6.1-6.3.

Assumption 6. *$0 < \inf_{\varrho \in \Theta_{\varrho\epsilon}} \lambda_{\min}(\mathcal{I}_\varrho) \leq \sup_{\varrho \in \Theta_{\varrho\epsilon}} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ for $\mathcal{I}_\varrho = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho k} s_{\varrho k}')$, where $s_{\varrho k}$ is given in (19).*

Proposition 8. *Suppose Assumptions 1, 2, 4, 5 and 6 hold. Then, under the null hypothesis of $M = 1$, (a) $\sup_\xi \sup_{\vartheta \in A_n(\xi)} |t(\psi, \pi)| = O_p(n^{-1/2})$; and (b) for all $c > 0$,*

$$\sup_{\xi \in \Xi} \sup_{\vartheta \in A_{nc}(\xi)} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n}t(\psi, \pi)'\nu_n(s_{\varrho k}) + nt(\psi, \pi)'\mathcal{I}_\varrho t(\psi, \pi)/2 \right| = o_p(1). \tag{21}$$

As shown in Sections 6.2 and 6.3, Assumption 6 does not hold for regime switching models with normal distribution. We proceed to derive the asymptotic distribution of the LRTS. With $s_{\varrho k}$ defined in (19), define

$$\begin{aligned} &\mathcal{I}_\eta := \mathbb{E}_{\vartheta^*}(s_{\eta k} s_{\eta k}'), \quad \mathcal{I}_{\lambda\varrho_1\varrho_2} := \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\lambda\varrho_1 k} s_{\lambda\varrho_2 k}'), \quad \mathcal{I}_{\lambda\eta\varrho} := \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\lambda\varrho k} s_{\eta k}'), \\ &\mathcal{I}_{\eta\lambda\varrho} := \mathcal{I}_{\lambda\eta\varrho}', \quad \mathcal{I}_{\lambda.\eta\varrho_1\varrho_2} := \mathcal{I}_{\lambda\varrho_1\varrho_2} - \mathcal{I}_{\lambda\eta\varrho_1}\mathcal{I}_\eta^{-1}\mathcal{I}_{\eta\lambda\varrho_2}, \quad \mathcal{I}_{\lambda.\eta\varrho} := \mathcal{I}_{\lambda.\eta\varrho\varrho}, \quad Z_{\lambda\varrho} := (\mathcal{I}_{\lambda.\eta\varrho})^{-1} G_{\lambda.\eta\varrho}, \end{aligned} \tag{22}$$

where $G_{\lambda.\eta\varrho}$ is a q_λ-vector mean zero Gaussian process indexed by ϱ with $\text{cov}(G_{\lambda.\eta\varrho_1}, G_{\lambda.\eta\varrho_2}) = \mathcal{I}_{\lambda.\eta\varrho_1\varrho_2}$. Define the set of admissible values for $\sqrt{n}\alpha(1-\alpha)v(\lambda)$ when $n \to \infty$ by $v(\mathbb{R}^q) := \{x \in \mathbb{R}^{q_\lambda} : x = v(\lambda) \text{ for some } \lambda \in \mathbb{R}^q\}$. Define $\hat{t}_{\lambda\varrho}$ by

$$g_{\lambda\varrho}(\hat{t}_{\lambda\varrho}) = \inf_{t_\lambda \in v(\mathbb{R}^q)} g_{\lambda\varrho}(t_\lambda), \quad g_{\lambda\varrho}(t_\lambda) := (t_\lambda - Z_{\lambda\varrho})' \mathcal{I}_{\lambda.\eta\varrho}(t_\lambda - Z_{\lambda\varrho}). \tag{23}$$

The following proposition establishes the asymptotic null distribution of the LRTS.

Proposition 9. *Suppose Assumptions 1, 2, 4, 5 and 6 hold. Then, under the null hypothesis of* $M = 1$, $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \Rightarrow \sup_{\varrho \in \Theta_{\varrho\epsilon}} \left(\hat{t}_{\lambda\varrho}' \mathcal{I}_{\lambda.\eta\varrho} \hat{t}_{\lambda\varrho} \right)$.

In proposition 9, the LRTS and its asymptotic distribution depends on the choice of ϵ. It is possible to develop a version of EM test (Chen and Li, 2009; Chen et al., 2012; Kasahara and Shimotsu, 2015) in this context which does not impose an explicit restriction on the parameter space for p_{11} and p_{22} but we leave such an extension for future research.

6.2 Heteroscedastic normal distribution

Suppose that $Y_k \in \mathbb{R}$ in the j-th regime follows a normal distribution with regime-specific intercept μ_j and variance σ_j^2. We split θ_j into $\theta_j = (\zeta_j, \sigma_j^2)' = (\mu_j, \beta_j', \sigma_j^2)'$, and write the density for the j-th regime as

$$f(y_k | \overline{y}_{k-1}; \gamma, \theta_j) = f(y_k | \overline{y}_{k-1}; \gamma, \zeta_j, \sigma_j^2) = \frac{1}{\sigma_j} \phi \left(\frac{y_k - \mu_j - \varpi(\overline{y}_{k-1}; \gamma, \beta_j)}{\sigma_j} \right), \tag{24}$$

for some function ϖ. In many applications, ϖ is a linear function of γ and β_j, e.g., $\varpi(\overline{y}_{k-1}, w_k; \gamma, \beta_j) = (\overline{y}_{k-1})'\beta_j + w_k'\gamma$. Consider the following reparameterization introduced in Kasahara and Shimotsu (2015) (θ in Kasahara and Shimotsu corresponds to ζ here):

$$\begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \nu_\zeta + (1-\alpha)\lambda_\zeta \\ \nu_\zeta - \alpha\lambda_\zeta \\ \nu_\sigma + (1-\alpha)(2\lambda_\sigma + C_1\lambda_\mu^2) \\ \nu_\sigma - \alpha(2\lambda_\sigma + C_2\lambda_\mu^2) \end{pmatrix}, \tag{25}$$

where $\nu_\zeta = (\nu_\mu, \nu_\beta')'$, $\lambda_\zeta = (\lambda_\mu, \lambda_\beta')'$, $C_1 := -(1/3)(1+\alpha)$, and $C_2 := (1/3)(2-\alpha)$, so that $C_1 = C_2 - 1$. Collect the reparameterized parameters, except for α, into one vector ψ_α. As in Section 6.1, we suppress the subscript α from ψ_α. Let the reparameterized density be

$$g_\psi(y_k | \overline{y}_{k-1}, x_k) = f\left(y_k | \overline{y}_{k-1}; \gamma, \nu_\zeta + (q_k - \alpha)\lambda_\zeta, \nu_\sigma + (q_k - \alpha)(2\lambda_\sigma + (C_2 - q_k)\lambda_\mu^2)\right). \tag{26}$$

Let $\psi := (\eta', \lambda')' \in \Theta_\psi = \Theta_\eta \times \Theta_\lambda$, where $\eta := (\gamma', \nu_\zeta', \nu_\sigma)'$ and $\lambda := (\lambda_\zeta', \lambda_\sigma)'$. Because the likelihood function of a normal mixture model is unbounded when $\sigma_j \to 0$ (Hartigan, 1985), we impose $\sigma_j \geq \epsilon_\sigma$ for a small $\epsilon_\sigma > 0$ in Θ_ψ. Henceforth, we suppress $\overline{y}_{k-1}$ from $g_\psi(y_k | \overline{y}_{k-1}, x_k)$. We proceed to derive the derivatives of $g_\psi(y_k|x_k)$ with respect to ψ. $\nabla_\psi g_\psi(y_k|x_k)$, $\nabla_{\lambda\eta'} g_\psi(y_k|x_k)$, and $\nabla_{\lambda\lambda'} g_\psi(y_k|x_k)$ are the same as those given in (13) except for $\nabla_{\lambda_\mu^2} g_\psi(y_k|x_k)$ and that those with respect to λ_σ^j are multiplied by 2^j. Higher-order derivatives of $g_\psi(y_k|x_k)$ with respect to λ_μ are derived by following Kasahara and Shimotsu (2015). From Lemma 5 and the fact that the normal density $f(\mu,\sigma^2)$ satisfies

$$\begin{aligned}&\nabla_{\mu^2} f(\mu,\sigma^2) = 2\nabla_{\sigma^2} f(\mu,\sigma^2), \quad \nabla_{\mu^3} f(\mu,\sigma^2) = 2\nabla_{\mu\sigma^2} f(\mu,\sigma^2), \quad \text{and} \\ &\nabla_{\mu^4} f(\mu,\sigma^2) = 2\nabla_{\mu^2\sigma^2} f(\mu,\sigma^2) = 4\nabla_{\sigma^2\sigma^2} f(\mu,\sigma^2),\end{aligned} \tag{27}$$

we have

$$\nabla_{\lambda_\mu^i} g_k^* = h_{ik} \nabla_{\mu^i} f(Y_k|\gamma^*,\theta^*), \quad i = 0,\dots,4, \tag{28}$$

where

$$\begin{aligned}&h_{0k} := 1, \quad h_{1k} := q_k - \alpha, \quad h_{2k} := (q_k-\alpha)(C_2-\alpha), \quad h_{3k} := 2(q_k-\alpha)^2(1-\alpha-q_k), \\ &h_{4k} := -2(q_k-\alpha)^4 + 3(q_k-\alpha)^2(\alpha - C_2)^2.\end{aligned}$$

It follows from $\mathbb{E}_{\vartheta^*}[q_k|\overline{Y}_{-\infty}^n] = \alpha$, Lemma 12(a), and elementary calculation that

$$\begin{aligned}\mathbb{E}_{\vartheta^*}[h_{ik}|\overline{Y}_{-\infty}^n] &= 0, \quad \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^i} g_k^*|\overline{Y}_{-\infty}^k] = 0, \quad i=1,2,3, \\ \mathbb{E}_{\vartheta^*}[h_{4k}|\overline{Y}_{-\infty}^n] &= \alpha(1-\alpha)b(\alpha), \\ \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^4} g_k^*|\overline{Y}_{-\infty}^k] &= \alpha(1-\alpha)b(\alpha)\nabla_{\mu^4} f(Y_k;\gamma^*,\theta^*) \\ &= \alpha(1-\alpha)b(\alpha)4\nabla_{\sigma^2\sigma^2} f(Y_k;\gamma^*,\theta^*) = b(\alpha)\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2} g_k^*|\overline{Y}_{-\infty}^k],\end{aligned} \tag{29}$$

with $b(\alpha) := -(2/3)(\alpha^2-\alpha+1) < 0$. Hence, $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2} g_k^*|\overline{Y}_{-\infty}^k]$ and $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^4} g_k^*|\overline{Y}_{-\infty}^k]$ are linearly dependent.

We proceed to derive $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$. Repeating the calculation leading to (15)-(17) and using (29) gives the following: first, (15) and (16) still hold; second, the elements of $\nabla_{\lambda\lambda'} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ except for the $(1,1)$th element are given by (17) after adjusting that the derivative with respect to λ_σ must be multiplied by 2 (e.g., $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma} g_k^*|\overline{Y}_{-\infty}^n] = 2\nabla_{\sigma^2} f_k^*$ and $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma\lambda_\mu} g_k^*|\overline{Y}_{-\infty}^n] = 2\nabla_{\sigma^2\mu} f_k^*$); third,

$$\nabla_{\lambda_\mu^2} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \alpha(1-\alpha)\sum_{t=1}^{k-1} \varrho^{k-t}\left(2\frac{\nabla_\mu f_t^*}{f_t^*}\frac{\nabla_\mu f_k^*}{f_k^*}\right). \tag{30}$$

When $\varrho \neq 0$, $\nabla_{\lambda_\mu^2} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ is a non-degenerate random variable as in the non-normal case. When $\varrho = 0$, however, $\nabla_{\lambda_\mu^2} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ becomes identically equal to 0, and indeed the first non-zero derivative with respect to λ_μ is the fourth derivative.

Because of this degeneracy, we derive the asymptotic distribution of the LRTS by expanding

$\ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi)$ four times. It is not correct, however, to simply approximate $\ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi)$ by a quadratic function of λ_μ^2 (and other terms) when $\varrho \neq 0$ and a quadratic function of λ_μ^4 when $\varrho = 0$. This results in discontinuity at $\varrho = 0$ and fails to provide a valid uniform approximation. We establish a uniform approximation by expanding $\ell_n(\psi,\pi,\xi)$ four times but expressing $\ell_n(\psi,\pi,\xi)$ in terms of $\varrho\lambda_\mu^2$, λ_μ^4, and other terms.

For $m \geq 0$, define $\zeta_{k,m}(\varrho) := \sum_{t=-m+1}^{k-1} \varrho^{k-t-1} 2\nabla_\mu f_t^* \nabla_\mu f_k^* / f_t^* f_k^*$. Then, we can write (30) as

$$\frac{\nabla_{\lambda_\mu^2} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\alpha(1-\alpha)\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \sum_{t=1}^{k-1} \varrho^{k-t}\left(2\frac{\nabla_\mu f_t^*}{f_t^*}\frac{\nabla_\mu f_k^*}{f_k^*}\right) = \varrho\zeta_{k,0}(\varrho). \tag{31}$$

Note that $\zeta_{k,m}(\varrho)$ satisfies $\mathbb{E}_{\vartheta^*}[\zeta_{k,m}(\varrho)|\overline{Y}_{-m}^{k-1}] = 0$ and is non-degenerate even when $\varrho = 0$.

Define $v(\lambda_\beta)$ as $v(\lambda)$ in (18) but replacing λ with λ_β. Collect the relevant parameters as

$$t(\psi,\pi) := \begin{pmatrix} \eta - \eta^* \\ t_\lambda(\lambda,\pi) \end{pmatrix}, \tag{32}$$

where

$$t_\lambda(\lambda,\pi) := \begin{pmatrix} \alpha(1-\alpha)\varrho\lambda_\mu^2 \\ \alpha(1-\alpha)\lambda_\mu\lambda_\sigma \\ \alpha(1-\alpha)[\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12] \\ \alpha(1-\alpha)\lambda_\beta\lambda_\mu \\ \alpha(1-\alpha)\lambda_\beta\lambda_\sigma \\ \alpha(1-\alpha)v(\lambda_\beta) \end{pmatrix}, \tag{33}$$

with $b(\alpha) := -(2/3)(\alpha^2 - \alpha + 1) < 0$. Recall $\theta_j = (\zeta_j', \sigma_j^2)' = (\mu_j, \beta_j', \sigma_j^2)'$. Similarly to (20), define the elements of the generalized score by

$$\begin{pmatrix} * & s_{\lambda_{\mu\beta}\varrho k} & s_{\lambda_{\mu\sigma}\varrho k} \\ s_{\lambda_{\beta\mu}\varrho k} & s_{\lambda_{\beta\beta}\varrho k} & s_{\lambda_{\beta\sigma}\varrho k} \\ s_{\lambda_{\sigma\mu}\varrho k} & s_{\lambda_{\beta\sigma}\varrho k} & s_{\lambda_{\sigma\sigma}\varrho k} \end{pmatrix} = \frac{\nabla_{\theta\theta'} f_k^*}{f_k^*} + \sum_{t=1}^{k-1} \varrho^{k-t}\left(\frac{\nabla_\theta f_t^*}{f_t^*}\frac{\nabla_{\theta'} f_k^*}{f_k^*} + \frac{\nabla_\theta f_k^*}{f_k^*}\frac{\nabla_{\theta'} f_t^*}{f_t^*}\right). \tag{34}$$

Define the generalized score as

$$s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\varrho k} \end{pmatrix}, \quad \text{where} \quad s_{\eta k} := \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix} \quad \text{and} \quad s_{\lambda\varrho k} := \begin{pmatrix} \zeta_{k,0}(\varrho)/2 \\ 2s_{\lambda_{\mu\sigma}\varrho k} \\ 2s_{\lambda_{\sigma\sigma}\varrho k} \\ s_{\lambda_{\beta\mu}\varrho k} \\ 2s_{\lambda_{\beta\sigma}\varrho k} \\ V(s_{\lambda_{\beta\beta}\varrho k}) \end{pmatrix}. \tag{35}$$

The following proposition establishes a uniform approximation of the log-likelihood ratio.

Assumption 7. *(a)* $0 < \inf_{\varrho \in \Theta_{\varrho\epsilon}} \lambda_{\min}(\mathcal{I}_\varrho) \leq \sup_{\varrho \in \Theta_{\varrho\epsilon}} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ *for* $\mathcal{I}_\varrho = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho k} s_{\varrho k}')$, *where* $s_{\varrho k}$ *is given in (35). (b)* $\sigma_1^*, \sigma_2^* > \epsilon_\sigma$.

Proposition 10. *Suppose Assumptions 1, 2, 4, 5 and 7 hold, and the density for the j-th regime is given by (24). Then, under the null hypothesis of* $M = 1$, *(a)* $\sup_{\vartheta \in A_n(\xi)} |t(\psi, \pi)| = O_p(n^{-1/2})$; *and (b) for all* $c > 0$,

$$\sup_{\xi \in \Xi} \sup_{\vartheta \in A_{nc}(\xi)} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n} t(\psi, \pi)' \nu_n(s_{\varrho k}) + n t(\psi, \pi)' \mathcal{I}_\varrho t(\psi, \pi)/2 \right| = o_p(1). \quad (36)$$

Let $\Lambda_{\lambda\varrho n}$ be the set of **possible** values of $\sqrt{n} t_\lambda(\lambda, \pi)$ defined in (32). The asymptotic null distribution of $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)]$ is characterized by the supremum of $2t_\lambda' G_{\lambda.\eta\varrho} - t_\lambda' \mathcal{I}_{\lambda.\eta\varrho} t_\lambda$, where $G_{\lambda.\eta\varrho}$ and $\mathcal{I}_{\lambda.\eta\varrho}$ are defined analogously to those in (22) but with $s_{\varrho k}$ defined in (35), and the supremum is taken with respect to t_λ and $\varrho \in \Theta_{\varrho\epsilon}$ under the constraint implied by the limit of $\Lambda_{\lambda\varrho n}$ as $n \to \infty$. This constraint is given by Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$, where $d_\beta := \dim(\beta)$, $q_\lambda := 3 + 2d_\beta + d_\beta(d_\beta + 1)/2$, and

$$\begin{aligned}
\Lambda_\lambda^1 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}', t_{\beta\sigma}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : \\
&\qquad (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}')' \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_- \times \mathbb{R}^{d_\beta}, t_{\beta\sigma} = 0, t_{v(\beta)} = 0\}, \\
\Lambda_{\lambda\varrho}^2 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}', t_{\beta\sigma}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : t_{\varrho\mu^2} = \varrho\lambda_\mu^2, t_{\mu\sigma} = \lambda_\mu\lambda_\sigma, \\
&\qquad t_{\sigma^2} = \lambda_\sigma^2, t_{\beta\mu} = \lambda_\beta\lambda_\mu, t_{\beta\sigma} = \lambda_\beta\lambda_\sigma, t_{v(\beta)} = v_\beta(\lambda_\beta) \text{ for some } \lambda \in \mathbb{R}^{2+d_\beta}\}.
\end{aligned} \quad (37)$$

Note that $\Lambda_{\lambda\varrho}^2$ depends on ϱ, whereas Λ_λ^1 does not depend on ϱ. Heuristically, Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$ correspond to the the limits of $\sqrt{n} t_\lambda(\lambda, \pi)$ when $\liminf_{n\to\infty} n^{1/8}|\lambda_\mu| > 0$ and $\lambda_\mu = o(n^{-1/8})$, respectively. When $\liminf_{n\to\infty} n^{1/8}|\lambda_\mu| > 0$, we have $(\hat{\lambda}_\sigma, \hat{\lambda}_\beta) = O_p(n^{-3/8})$ because $t_\lambda(\hat{\lambda}, \pi) = O_p(n^{-1/2})$. Further, the **set of possible** values of $\sqrt{n}\varrho\lambda_\mu^2$ converges to $\mathbb{R}$ because ϱ can be arbitrary small. Consequently, the limit of $\sqrt{n} t_\lambda(\lambda, \pi)$ is characterized by Λ_λ^1.

Define $Z_{\lambda\varrho}$ and $\mathcal{I}_{\lambda.\eta\varrho}$ as in (22) but with $s_{\pi k}$ defined in (35). Let $Z_{\lambda 0}$ and $\mathcal{I}_{\lambda.\eta 0}$ denote $Z_{\lambda\varrho}$ and $\mathcal{I}_{\lambda.\eta\varrho}$ evaluated at $\varrho = 0$. Define $\hat{t}_\lambda^1$ and $\hat{t}_{\lambda\varrho}^2$ by

$$\begin{aligned}
r_\lambda(\hat{t}_\lambda^1) &= \inf_{t_\lambda \in \Lambda_\lambda^1} r_\lambda(t_\lambda), \quad r_\lambda(t_\lambda) := (t_\lambda - Z_{\lambda 0})' \mathcal{I}_{\lambda.\eta 0} (t_\lambda - Z_{\lambda 0}) \\
r_{\lambda\varrho}(\hat{t}_{\lambda\varrho}^2) &= \inf_{t_\lambda \in \Lambda_{\lambda\varrho}^2} r_{\lambda\varrho}(t_\lambda), \quad r_{\lambda\varrho}(t_\lambda) := (t_\lambda - Z_{\lambda\varrho})' \mathcal{I}_{\lambda.\eta\varrho} (t_\lambda - Z_{\lambda\varrho}).
\end{aligned} \quad (38)$$

The following proposition establishes the asymptotic null distribution of the LRT statistic. Define a stochastic process $W(\varrho)$ as $W(\varrho) := \max\{\mathbb{I}\{\varrho = 0\}(\hat{t}_\lambda^1)' \mathcal{I}_{\lambda.\eta 0} \hat{t}_\lambda^1, (\hat{t}_{\lambda\varrho}^2)' \mathcal{I}_{\lambda.\eta\varrho} \hat{t}_{\lambda\varrho}^2\}$; $W(\varrho)$ is the maximum of two random variables if $\varrho = 0$.

Proposition 11. *Suppose that assumptions in Proposition 10 hold. Then, under the null hypothesis of* $M = 1$, $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \Rightarrow \sup_{\varrho \in \Theta_{\varrho\epsilon}} W(\varrho)$.

Even though the current model reduces to a heteroskedastic normal mixture when $\varrho = 0$, the asymptotic distribution of the LRTS is different from a union of that in Proposition 9 and that of the

LRTS of a heteroskedastic normal mixture given in Kasahara and Shimotsu (2015). This is because $W(\varrho)$ at $\varrho = 0$ is distributed differently from the LRTS in Kasahara and Shimotsu (2015) in that $t_\lambda \in \Lambda_\lambda^1$ here has an additional term $t_{\varrho\mu^2}$ that corresponds to $\varrho\lambda_\mu^2$ in the expansion of $\ell_n(\psi, \pi, \xi)$. This term signifies the importance of the behavior of the LRTS when ϱ is in a neighborhood of 0.

6.3 Homoscedastic normal distribution

Suppose that $Y_k \in \mathbb{R}$ in the j-th regime follows a normal distribution with regime specific intercept μ_j but with common variance σ^2. We split γ and θ_j into $\gamma = (\tilde{\gamma}, \sigma^2)$ and $\theta_j = (\mu_j, \beta_j)$, and write the density for the j-th regime as

$$f(y_k|\overline{y}_{k-1}; \gamma, \theta_j) = f(y_k|\overline{y}_{k-1}; \tilde{\gamma}, \theta_j, \sigma^2) = \frac{1}{\sigma}\phi\left(\frac{y_k - \mu_j - \varpi(\overline{y}_{k-1}; \tilde{\gamma}, \beta_j)}{\sigma}\right). \tag{39}$$

for some function ϖ. Consider the following reparameterization:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \nu_\theta + (1-\alpha)\lambda \\ \nu_\theta - \alpha\lambda \\ \nu_\sigma - \alpha(1-\alpha)\lambda_\mu^2 \end{pmatrix}, \tag{40}$$

where $\nu_\theta = (\nu_\mu, \nu_\beta')'$ and $\lambda = (\lambda_\mu, \lambda_\beta')'$. Collect the reparameterized parameters, except for α, into one vector ψ_α. Suppressing α from ψ_α, let the reparameterized density be

$$g_\psi(y_k|\overline{y}_{k-1}, x_k) = f\left(y_k|\overline{y}_{k-1}; \tilde{\gamma}, \nu_\theta + (q_k - \alpha)\lambda, \nu_\sigma - \alpha(1-\alpha)\lambda_\mu^2\right). \tag{41}$$

Let $\eta = (\tilde{\gamma}', \nu_\theta', \nu_\sigma)'$, then the first and second derivatives of $g_\psi(y_k|\overline{y}_{k-1}, x_k)$ with respect to η and λ are the same as those given in (13) except for $\nabla_{\lambda_\mu^2} g_\psi(y_k|\overline{y}_{k-1}, x_k)$. We derive higher-order derivatives of $g_\psi(y_k|\overline{y}_{k-1}, x_k)$ with respect to λ_μ. From Lemma 5 and (27), we obtain

$$\begin{aligned} \nabla_{\lambda\eta^i} g_k^* &= h_{1k} \nabla_{\theta\eta^i} f(Y_k|\overline{Y}_{k-1}; \gamma^*, \theta^*) \quad \text{for } i = 0, 1, \ldots, \\ \nabla_{\lambda_\mu^i} g_k^* &= h_{ik} \nabla_{\mu^i} f(Y_k|\overline{Y}_{k-1}; \gamma^*, \theta^*) \quad \text{for } i = 0, 1, \ldots, 4, \end{aligned} \tag{42}$$

where $h_{0k} := 1$, $h_{1k} := q_k - \alpha$, $h_{2k} := (q_k - \alpha)^2 - \alpha(1-\alpha)$, $h_{3k} := (q_k - \alpha)^3 - 3(q_k - \alpha)\alpha(1-\alpha)$, and $h_{4k} := (q_k - \alpha)^4 - 6(q_k - \alpha)^2\alpha(1-\alpha) + 3\alpha^2(1-\alpha)^2$. It follows from $\mathbb{E}_{\vartheta^*}[q_k|\overline{Y}_{-\infty}^n] = \alpha$, Lemma 12(a), and elementary calculation that

$$\begin{aligned} \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^i} g_k^*|\overline{Y}_0^k] &= \mathbb{E}_{\vartheta^*}[h_{ik}|\overline{Y}_0^k] = 0, \quad i = 1, 2, \\ \mathbb{E}_{\vartheta^*}[h_{3k}|\overline{Y}_0^k] &= \alpha(1-\alpha)(1-2\alpha), \quad \mathbb{E}_{\vartheta^*}[h_{4k}|\overline{Y}_0^k] = \alpha(1-\alpha)(1-6\alpha+6\alpha^2). \end{aligned} \tag{43}$$

Repeating the calculation leading to (15)–(17) and using (43) gives the following: first, (15) and (16) still hold; second, the elements of $\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ are given by (17) except for the $(1,1)$th element; third, $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ is given by (30). Further, Lemma

7 in the Appendix shows that, when $\varrho = 0$, $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \alpha(1-\alpha)(1-2\alpha)\nabla_{\mu^3} f_k^*/f_k^*$ and $\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\nabla_{\mu^4} f_k^*/f_k^*$. Because $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$ when $\alpha = 1/2$ and $\varrho = 0$, we expand $\ell_n(\psi,\pi,\xi)$ four times and express it in terms of $\varrho\lambda_\mu^2$, $(1-2\alpha)\lambda_\mu^3$, λ_μ^4, and other terms to establish a uniform approximation.

Collect the relevant parameters as

$$t(\psi,\pi) := \begin{pmatrix} \eta-\eta^* \\ t_\lambda(\lambda,\pi) \end{pmatrix} \quad \text{and} \quad t_\lambda(\lambda,\pi) := \begin{pmatrix} \alpha(1-\alpha)\varrho\lambda_\mu^2 \\ \alpha(1-\alpha)(1-2\alpha)\lambda_\mu^3 \\ \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\lambda_\mu^4 \\ \alpha(1-\alpha)\lambda_\beta\lambda_\mu \\ \alpha(1-\alpha)v(\lambda_\beta) \end{pmatrix}. \tag{44}$$

Define the generalized score as

$$s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\varrho k} \end{pmatrix}, \quad \text{where} \quad s_{\eta k} := \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix} \quad \text{and} \quad s_{\lambda\varrho k} := \begin{pmatrix} \zeta_{k,0}(\varrho)/2 \\ s_{\lambda_\mu^3 k}/3! \\ s_{\lambda_\mu^4 k}/4! \\ s_{\lambda_{\beta\mu}\varrho k} \\ V(s_{\lambda_{\beta\beta}\varrho k}) \end{pmatrix}, \tag{45}$$

where $\zeta_{k,m}(\varrho)$ is defined as in (31), $s_{\lambda_\mu^i k} := \nabla_{\mu^i} f_k^*/f_k^*$ for $i = 3,4$, and $s_{\lambda_{\beta\mu}\varrho k}$ and $s_{\lambda_{\beta\beta}\varrho k}$ are defined as in (34) but using the density function (39) in place of (24). Define, with $d_\beta := \dim(\beta)$ and $q_\lambda := 3 + d_\beta + d_\beta(d_\beta+1)/2$,

$$\begin{aligned} \Lambda_\lambda^1 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}')' \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}_-\times\mathbb{R}^{d_\beta}, t_{v(\beta)} = 0\}, \\ \Lambda_{\lambda\varrho}^2 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : t_{\varrho\mu^2} = \varrho\lambda_\mu^2, t_{\mu^3} = t_{\mu^4} = 0, t_{\beta\mu} = \lambda_\beta\lambda_\mu, \\ &\qquad t_{v(\beta)} = v_\beta(\lambda_\beta) \text{ for some } \lambda \in \mathbb{R}^{1+d_\beta}\}. \end{aligned} \tag{46}$$

The following two propositions correspond to Proposition 10 and 11, establishing a uniform approximation of the log-likelihood ratio and the asymptotic distribution of the LRT statistic.

Assumption 8. *$0 < \inf_{\varrho\in\Theta_{\varrho\epsilon}} \lambda_{\min}(\mathcal{I}_\varrho) \le \sup_{\varrho\in\Theta_{\varrho\epsilon}} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ for $\mathcal{I}_\varrho = \lim_{k\to\infty}\mathbb{E}_{\vartheta^*}(s_{\varrho k}s_{\varrho k}')$, where $s_{\varrho k}$ is given in (45).*

Proposition 12. *Suppose Assumptions 1, 2, 4, 5 and 8 hold, and the density for the j-th regime is given by (39). Then, statements (a) and (b) of Proposition 10 hold.*

Proposition 13. *Suppose that assumptions in Proposition 12 hold. Then, under the null hypothesis of $M = 1$, $2[\ell_n(\hat\vartheta_2,\xi_2) - \ell_{0,n}(\hat\vartheta_1)] \Rightarrow \sup_{\varrho\in\Theta_\varrho}\max\{\mathbb{I}\{\varrho=0\}(\hat t_\lambda^1)'\mathcal{I}_{\lambda.\eta 0}\hat t_\lambda^1, (\hat t_{\lambda\varrho}^2)'\mathcal{I}_{\lambda.\eta\varrho}\hat t_{\lambda\varrho}^2\}$, where $\hat t_\lambda^1$ and $\hat t_{\lambda\varrho}^2$ are defined as in (38) but in terms of $(Z_{\lambda\varrho}, \mathcal{I}_{\lambda.\eta\varrho}, Z_{\lambda 0}, \mathcal{I}_{\lambda.\eta 0})$ constructed with $s_{\varrho k}$ defined in (45) and Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$ defined in (46).*

7 Testing $H_0 : M = M_0$ against $H_A : M = M_0 + 1$ for $M_0 \geq 2$

In this section, we derive the asymptotic distribution of the LRTS for testing the null hypothesis of M_0 regimes against the alternative of $M_0 + 1$ regimes for general $M_0 \geq 2$.

Let $\vartheta^*_{M_0} = ((\vartheta^*_{M_0,x})', (\vartheta^*_{M_0,y})')'$ denote the parameter of the M_0-regime model, where $\vartheta^*_{M_0,x}$ contains $p^*_{ij} = q_{\vartheta^*_{M_0,x}}(i,j) > 0$ for $i = 1, \ldots, M_0$ and $j = 1, \ldots, M_0-1$, and $\vartheta^*_{M_0,y} = ((\theta^*_1)', \ldots, (\theta^*_{M_0})', (\gamma^*)')'$. We assume $\max_i \sum_{j=1}^{M_0-1} p^*_{ij} < 1$, and we assume $\theta^*_1 < \ldots < \theta^*_{M_0}$ for identification. The true M_0-regime conditional density function of Y_1^n given $\overline{Y}_0$ and x_0 is

$$p_{\vartheta^*_{M_0}}(Y_1^n|\overline{Y}_0, x_0) = \sum_{x_1^n \in \mathcal{X}^n_{M_0}} \prod_{k=1}^n p_{\vartheta^*_{M_0}}(Y_k, x_k|\overline{Y}_{k-1}, x_{k-1}), \tag{47}$$

where $p_{\vartheta^*_{M_0}}(y_k, x_k|\overline{y}_{k-1}, x_{k-1}) = g_{\vartheta^*_{M_0,y}}(y_k|\overline{y}_{k-1}, x_k) q_{\vartheta^*_{M_0,x}}(x_{k-1}, x_k)$ with $g_{\vartheta^*_{M_0,y}}(y_k|\overline{y}_{k-1}, x_k) = \sum_{j=1,\ldots,M_0} \mathbb{I}\{x_k = j\} f(y_k|\overline{y}_{k-1}; \gamma, \theta^*_j)$.

Let the conditional density of Y_1^n of an (M_0+1)-regime model be

$$p_{\vartheta_{M_0+1}}(Y_1^n|\overline{Y}_0, x_0) := \sum_{x_1^n \in \mathcal{X}^n_{M_0+1}} \prod_{k=1}^n p_{\vartheta_{M_0+1}}(Y_k, x_k|\overline{Y}_{k-1}, x_{k-1}), \tag{48}$$

where $p_{\vartheta_{M_0+1}}(y_k, x_k|\overline{y}_{k-1}, x_{k-1})$ is defined similarly to $p_{\vartheta^*_{M_0}}(y_k, x_k|\overline{y}_{k-1}, x_{k-1})$ with $\vartheta_{M_0+1,x} := \{p_{ij}\}_{i=1,\ldots,M_0+1, j=1,\ldots,M_0}$ and $\vartheta_{M_0+1,y} := (\theta'_1, \ldots, \theta'_{M_0+1}, \gamma')'$. We assume that $\min_{i,j} p_{ij} \geq \epsilon$ for some $\epsilon \in (0, 1/2)$, and let $\Theta_{M_0+1,\epsilon}$ denote Θ_{M_0+1} with this restriction.

Write the null hypothesis as $H_0 = \cup_{h=1}^{M_0} H_{0h}$ with

$$H_{0h} : \theta_1 < \cdots < \theta_h = \theta_{h+1} < \cdots < \theta_{M_0+1}.$$

Define the set of values of ϑ_{M_0+1} that yields the true density (47) under $\mathbb{P}_{\vartheta^*_{M_0}}$ as $\Upsilon^* := \{\vartheta_{M_0+1} \in \Theta_{M_0+1,\epsilon} : p_{\vartheta_{M_0+1}}(Y_1^n|\overline{Y}_0, x_0) = p_{\vartheta^*_{M_0}}(Y_1^n|\overline{Y}_0, x_0)\ \mathbb{P}_{\vartheta^*_{M_0}}\text{-a.s.}\}$. Under H_{0h}, the (M_0+1)-regime model (48) generates the true M_0-regime density (47) if $\theta_h = \theta_{h+1} = \theta^*_h$ and the transition matrix of X_k reduces to that of the true M_0-regime model. Define $J_h := \{h, h+1\}$ and $\overline{J}_h := \{1, \ldots, M_0+1\} \setminus J_h$, and let p_j and p^*_j denote $\mathbb{P}_{\vartheta_{M_0+1}}(X_k = j)$ and $\mathbb{P}_{\vartheta^*_{M_0}}(X_k = j)$, respectively. Define the subset of Υ^* that corresponds to H_{0h} as

$$\begin{aligned}\Upsilon^*_h := \big\{ & \vartheta_{M_0+1} \in \Theta_{M_0+1,\epsilon} : \theta_j = \theta^*_j \text{ for } 1 \leq j < h;\ \theta_h = \theta_{h+1} = \theta^*_h; \\ & \theta_j = \theta^*_{j-1} \text{ for } h+1 < j \leq M_0+1;\ \gamma = \gamma^*;\ \vartheta_{M_0+1,x} \in \Theta^*_{ph} \times \Pi^*_h \big\},\end{aligned}$$

where, with $\wedge$ and $\vee$ denoting "and" and "or",[4]

$$\Theta^*_{ph} := \left\{ \{p_{ij}\}_{i\in \overline{J}_h \wedge j \in \overline{J}_h \setminus \{M_0+1\}} : p_{ij} = p^*_{ij} \text{ for } 1 \le j < h \text{ and } p_{ij} = p^*_{i,j-1} \text{ for } h+1 < j \le M_0 \right\},$$
$$\Pi^*_h := \big\{ \{p_{ij}\}_{i\in J_h \vee j \in J_h \setminus \{M_0+1\}} : p_{ih} + p_{i,h+1} = p^*_{ih} \text{ for } i \in \bar{J}_h;$$
$$p_j = p^*_j \text{ for } 1 \le j < h \text{ and } p_j = p^*_{j-1} \text{ for } h+1 < j \le M_0+1 \big\},$$

then $\Upsilon^* = \Upsilon^*_1 \cup \cdots \cup \Upsilon^*_{M_0}$ holds. Lemma 17 in the Appendix shows that Υ^*_h gives the parameter set that corresponds to H_{0h}.

For $m = M_0, M_0+1$, let $\ell_n(\vartheta_m, \xi_m) := \log\left(\sum_{x_0=1}^m p_{\vartheta_m}(Y_1^n|\overline{Y}_0, x_0)\xi_m(x_0)\right)$ denote the m-regime log-likelihood for a given initial distribution $\xi_m(x_0) \in \Xi_m$. We treat $\xi_m(x_0)$ fixed. Let $\hat{\vartheta}_{M_0} := \arg\max_{\vartheta_{M_0}\in\Theta_{M_0}} \ell_n(\vartheta_{M_0}, \xi_{M_0})$ and $\hat{\vartheta}_{M_0+1} := \arg\max_{\vartheta_{M_0+1}\in\Theta_{M_0+1,\epsilon}} \ell_n(\vartheta_{M_0+1}, \xi_{M_0+1})$. The following proposition shows that the MLE is consistent in the sense that the distance between $\hat{\vartheta}_{M_0+1}$ and Υ^* tends to 0 in probability. The proof of Proposition 14 is essentially the same as the proof of Proposition 7 and hence is omitted.

Assumption 9. *(a) Θ_{M_0} and Θ_{M_0+1} are compact. (b) For all $(x,x') \in \mathcal{X}$ and all $(\overline{y}, y', w) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}$, the function $(\theta,\gamma) \mapsto f(y'|\overline{y}_0, w; \gamma, \theta)$ is continuous. (c) $\mathbb{E}_{\vartheta^*_{M_0}}[\log(p_{\vartheta^*_{M_0}}(Y_1|\overline{Y}^0_{-m}, W^0_{-m})] = \mathbb{E}_{\vartheta^*_{M_0}}[\log p_{\vartheta_{M_0}}(Y_1|\overline{Y}^0_{-m}, W^0_{-m})]$ for all $m \ge 0$ if and only if $\vartheta_{M_0} = \vartheta^*_{M_0}$. (d) $\mathbb{E}_{\vartheta^*_{M_0}}[\log(p_{\vartheta^*_{M_0}}(Y_1|\overline{Y}^0_{-m}, W^0_{-m})] = \mathbb{E}_{\vartheta^*_{M_0}}[\log p_{\vartheta_{M_0+1}}(Y_1|\overline{Y}^0_{-m}, W^0_{-m})]$ for all $m \ge 0$ if and only if $\vartheta_{M_0+1} \in \Upsilon^*$. (e) $p^*_{ij} > 2\epsilon$ for all $(i,j) \in \{1,\ldots,M_0\}^2$.*

Proposition 14. *Suppose Assumptions 1, 2, and 9 hold. Then, under the null hypothesis of $M = M_0$, $\hat{\vartheta}_{M_0} \to_p \vartheta^*_{M_0}$ and $\inf_{\vartheta_{M_0+1}\in\Upsilon^*} |\hat{\vartheta}_{M_0+1} - \vartheta_{M_0+1}| \to_p 0$.*

We proceed to derive the asymptotic distribution of the LRTS by analyzing the behavior of LRTS when $\vartheta_{M_0+1} \in \Upsilon^*_h$ for each h. Split $\vartheta_{M_0+1,x}$ as $\vartheta_{M_0+1,x} = (\vartheta'_{xh}, (\vartheta^c_{xh})')'$, where $\vartheta_{xh} := \{p_{ij}\}_{i\in\overline{J}_h\wedge j\in\overline{J}_h\setminus\{M_0+1\}}$ is identified under H_{0h}, and $\vartheta^c_{xh} := \{p_{ij}\}_{i\in J_h \vee j\in J_h\setminus\{M_0+1\}}$ is not point identified under H_{0h}. We reparameterize $(p_{hh}, p_{h,h+1}, p_{h+1,h}, p_{h,h+1}) \in \vartheta^c_{xh}$ to $(p_{hJ}, p_{hh|J}, p_{h+1,J}, p_{h+1,h+1|J})$, where $p_{hJ} := p_{hh} + p_{h,h+1}$, $p_{hh|J} := p_{hh}/p_{hJ}$, $p_{h+1,J} := p_{h+1,h} + p_{h+1,h+1}$, and $p_{h+1,h+1|J} := p_{h+1,h+1}/p_{h+1,J}$. If $X_1^k \in J_h^k$, then X_1^k follows a two-state Markov chain on J_h with transition matrix $P_J = \begin{pmatrix} p_{hh|J} & 1-p_{hh|J} \\ 1-p_{h+1,h+1|J} & p_{h+1,h+1|J} \end{pmatrix}$. We reparameterize $p_{hh|J}$ and $p_{h+1,h+1|J}$ to $\alpha_h := \mathbb{P}_{\vartheta^*_{M_0}}(X_k = h|X_k \in J_h)$ and $\varrho_h := \text{corr}_{\vartheta^*_{M_0}}(X_{k-1}, X_k|(X_{k-1}, X_k) \in J_h^2)$ and collect reparameterized ϑ^c_{xh} into $\pi_h := (\alpha'_h, \varrho'_h, \phi'_h)'$, where ϕ_h collects the elements of ϑ^c_{xh} that are not point identified under H_{0h} and that do not affect the transition probability of X_1^k when $X_1^k \in J_h^k$.

Define $q_{kj} := \mathbb{I}\{X_k = j\}$, then we can write α_h and ϱ_h as $\alpha_h = \mathbb{E}_{\vartheta^*_{M_0}}(q_{kh}|X_k \in J_h)$ and $\varrho_h = \text{corr}_{\vartheta^*_{M_0}}(q_{k-1,h}, q_{kh}|(X_{k-1}, X_k) \in J_h^2)$. Because $\overline{Y}^{\infty}_{-\infty}$ provides no information for distinguishing between $X_k = h$ and $X_k = h+1$ if $\theta_h = \theta_{h+1}$, we can write α_h and ϱ_h as

$$\alpha_h = \mathbb{E}_{\vartheta^*_{M_0}}(q_{kh}|X_k \in J_h, \overline{Y}^{\infty}_{-\infty}) \quad \text{and} \quad \varrho_h = \text{corr}_{\vartheta^*_{M_0}}(q_{k-1,h}, q_{k,h}|(X_{k-1}, X_k) \in J_h^2, \overline{Y}^{\infty}_{-\infty}). \tag{49}$$

[4] Strictly speaking, when $h = M_0$, we need to redefine $\vartheta_{M_0+1,x}$ so that $\vartheta_{M_0+1,x}$ contains $p_{i2}, \ldots, p_{i,M_0+1}$ and p_{i1} is determined by $p_{i1} = 1 - \sum_{j=2}^{M_0} p_{ij}$. We suppress this technical detail.

7.1 Non-normal distribution

For non-normal component distributions, consider the following reparameterization similar to (11):

$$\begin{pmatrix} \theta_h \\ \theta_{h+1} \end{pmatrix} = \begin{pmatrix} \nu + (1-\alpha_h)\lambda \\ \nu - \alpha_h \lambda \end{pmatrix}.$$

Collect the reparameterized identified parameters into one vector $\psi_h := (\vartheta'_{xh}, \eta', \lambda')'$, where $\eta = (\gamma', \{\theta'_j\}_{j=1}^{h-1}, \nu', \{\theta'_j\}_{j=h+2}^{M_0+1})'$, so that the reparameterized (M_0+1)-regime log-likelihood function is $\ell_n(\psi_h, \pi_h, \xi_{M_0+1})$. Let ψ_h^* denote the value of ψ_h under H_{0h}. Define the reparameterized conditional density of y_k as

$$g^h_{\psi_h}(y_k|\overline{y}_{k-1}, x_k) := \mathbb{I}\{x_k \in J_h\} f(y_k|\overline{y}_{k-1}; \gamma, \nu + (q_{kh} - \alpha_h)\lambda) + \sum_{j \in \overline{J}_h} q_{kj} f(y_k|\overline{y}_{k-1}; \gamma, \theta_j).$$

Let f^*_{hk} denote $f(Y_k|\overline{Y}_{k-1}; \gamma^*, \theta^*_h)$. It follows from (49) and the law of iterated expectations that

$$\begin{aligned}
&\mathbb{E}_{\vartheta^*_{M_0}}\left[\frac{\mathbb{I}\{X_k \in J_h\}(q_{kh} - \alpha_h)}{g^h_{\psi^*_h}(Y_k|\overline{Y}_{k-1}, X_k)} \middle| \overline{Y}^n_{-\infty} \right] \\
&= \mathbb{E}_{\vartheta^*_{M_0}}\left[\mathbb{E}_{\vartheta^*_{M_0}}\left[\frac{q_{kh} - \alpha_h}{f^*_{hk}} \middle| X_k \in J_h, \overline{Y}^n_{-\infty} \right] \mathbb{I}\{X_k \in J_h\} \middle| \overline{Y}^n_{-\infty} \right] = 0, \\
&\mathbb{E}_{\vartheta^*_{M_0}}\left[\frac{\mathbb{I}\{X_{t_1} \in J_h\}\mathbb{I}\{X_{t_2} \in J_h\}(q_{t_1 h} - \alpha_h)(q_{t_2 h} - \alpha_h)}{g^h_{\psi^*_h}(Y_{t_1}|\overline{Y}_{t_1-1}, X_{t_1}) g^h_{\psi^*_h}(Y_{t_2}|\overline{Y}_{t_2-1}, X_{t_2})} \middle| \overline{Y}^n_{-\infty} \right] \\
&= \mathbb{E}_{\vartheta^*_{M_0}}\left[\mathbb{E}_{\vartheta^*_{M_0}}\left[\frac{(q_{t_1 h} - \alpha_h)(q_{t_2 h} - \alpha_h)}{f^*_{ht_1} f^*_{ht_2}} \middle| X^{t_2}_{t_1} \in J_h^{t_2 - t_1 + 1}, \overline{Y}^n_{-\infty} \right] \mathbb{I}\{(X_{t_1}, X_{t_2}) \in J_h^2\} \middle| \overline{Y}^n_{-\infty} \right] \\
&= \frac{\alpha_h(1-\alpha_h)\varrho_h^{t_2 - t_1}}{f^*_{ht_1} f^*_{ht_2}} \mathbb{P}_{\vartheta^*_{M_0}}((X_{t_1}, X_{t_2}) \in J_h^2 | \overline{Y}^n_{-\infty}), \quad t_2 \geq t_1,
\end{aligned} \tag{50}$$

where the second equality holds because $g^h_{\psi^*_h}(Y_k|\overline{Y}_{k-1}, X_k) = f^*_{hk}$ if $X_k \in J_h$, and last equality holds because, conditional on $\{X^{t_2}_{t_1} \in J_h^{t_2 - t_1 + 1}, \overline{Y}^n_{-\infty}\}$, $X^{t_2}_{t_1}$ is a two-state stationary Markov process with transition probability P_J.

Let q^*_k denote $q_{\vartheta^*_{M_0,x}}(X_{k-1}, X_k)$. Repeating a derivation similar to (13)–(17) but using (50) in place of (14), we obtain

$$\begin{aligned}
&\nabla_{\vartheta_{xh}} \overline{p}_{\psi^*_h \pi}(Y_k|\overline{Y}_0^{k-1}) / \overline{p}_{\psi^*_h \pi}(Y_k|\overline{Y}_0^{k-1}) \\
&= \nabla_{\vartheta_{xh}} \log \overline{p}_{\psi^*_h \pi}(Y_k|\overline{Y}_0^{k-1}) = \sum_{t=1}^{k} \mathbb{E}_{\vartheta^*}\left[\nabla_{\vartheta_{M_0,x}} \log q^*_t \middle| \overline{Y}_0^k \right] - \sum_{t=1}^{k-1} \mathbb{E}_{\vartheta^*}\left[\nabla_{\vartheta_{M_0,x}} \log q^*_t \middle| \overline{Y}_0^{k-1} \right], \\
&\nabla_{\eta} \overline{p}_{\psi^*_h \pi}(Y_k|\overline{Y}_0^{k-1}) / \overline{p}_{\psi^*_h \pi}(Y_k|\overline{Y}_0^{k-1}) \\
&= \sum_{t=1}^{k} \mathbb{E}_{\vartheta^*}\left[\nabla_{(\gamma', \theta'_1, \ldots, \theta'_{M_0})'} \log g^*_t \middle| \overline{Y}_0^k \right] - \sum_{t=1}^{k-1} \mathbb{E}_{\vartheta^*}\left[\nabla_{(\gamma', \theta'_1, \ldots, \theta'_{M_0})'} \log g^*_t \middle| \overline{Y}_0^{k-1} \right],
\end{aligned} \tag{51}$$

$$\nabla_{\lambda}\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})=0,\quad \nabla_{\lambda\eta'}\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})=0,\quad \text{and} \tag{52}$$

$$\begin{aligned}\nabla_{\lambda\lambda'}\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1}) &= \alpha_h(1-\alpha_h)\frac{\nabla_{\theta\theta'}f_{hk}^*}{f_{hk}^*}\mathbb{P}_{\vartheta_{M_0}^*}(X_k\in J_h|\overline{Y}_0^k)\\ &+\alpha_h(1-\alpha_h)\sum_{t=1}^{k-1}\varrho_h^{k-t}\left(\frac{\nabla_\theta f_{ht}^*}{f_{ht}^*}\frac{\nabla_{\theta'}f_{hk}^*}{f_{hk}^*}+\frac{\nabla_\theta f_{hk}^*}{f_{hk}^*}\frac{\nabla_{\theta'}f_{ht}^*}{f_{ht}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_h^2|\overline{Y}_0^k).\end{aligned} \tag{53}$$

Collect the score for testing $H_{01},\ldots,H_{0M_0}$ into one vector as, with $\tilde{\varrho}:=(\varrho_1,\ldots,\varrho_{M_0})'$,

$$\tilde{s}_{\tilde{\varrho}k}:=\begin{pmatrix}\tilde{s}_{\eta k}\\ \tilde{s}_{\lambda\tilde{\varrho}k}\end{pmatrix},\quad \text{where}\ \ \tilde{s}_{\eta k}:=\frac{\nabla_{(\vartheta_{xh}',\eta')'}\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})},\quad \tilde{s}_{\lambda\tilde{\varrho}k}:=\begin{pmatrix}s_{\lambda\varrho_1k}^1\\ \vdots\\ s_{\lambda\varrho_{M_0}k}^{M_0}\end{pmatrix}, \tag{54}$$

and $s_{\lambda\varrho_hk}^h:=V(s_{\lambda\lambda\varrho_hk}^h)$, where $s_{\lambda\lambda\varrho_hk}^h$ is defined similarly to (20) as

$$\begin{aligned}s_{\lambda\lambda\varrho_hk}^h:=&\frac{\nabla_{\theta\theta'}f_{hk}^*}{f_{hk}^*}\mathbb{P}_{\vartheta_{M_0}^*}(X_k\in J_h|\overline{Y}_0^k)\\ &+\sum_{t=1}^{k-1}\varrho_h^{k-t}\left(\frac{\nabla_\theta f_{ht}^*}{f_{ht}^*}\frac{\nabla_{\theta'}f_{hk}^*}{f_{hk}^*}+\frac{\nabla_\theta f_{hk}^*}{f_{hk}^*}\frac{\nabla_{\theta'}f_{ht}^*}{f_{ht}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_h^2|\overline{Y}_0^k).\end{aligned} \tag{55}$$

Similarly to (22), define

$$\begin{aligned}&\tilde{\mathcal{I}}_\eta:=\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\eta k}\tilde{s}_{\eta k}'),\quad \tilde{\mathcal{I}}_{\lambda\tilde{\varrho}_1\tilde{\varrho}_2}:=\lim_{k\to\infty}\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\lambda\tilde{\varrho}_1k}\tilde{s}_{\lambda\tilde{\varrho}_2k}'),\quad \tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}}:=\lim_{k\to\infty}\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\lambda\tilde{\varrho}k}\tilde{s}_{\eta k}'),\\ &\tilde{\mathcal{I}}_{\eta\lambda\tilde{\varrho}}:=\tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}}',\quad \tilde{\mathcal{I}}_{\lambda.\eta\tilde{\varrho}_1\tilde{\varrho}_2}:=\tilde{\mathcal{I}}_{\lambda\tilde{\varrho}_1\tilde{\varrho}_2}-\tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}_1}\tilde{\mathcal{I}}_\eta^{-1}\tilde{\mathcal{I}}_{\eta\lambda\tilde{\varrho}_2},\quad \tilde{\mathcal{I}}_{\lambda.\eta\varrho_h}^h:=\mathbb{E}_{\vartheta_{M_0}^*}[G_{\lambda.\eta\varrho_h}^h(G_{\lambda.\eta\varrho_h}^h)'],\\ &Z_{\lambda\varrho_h}^h:=(\tilde{\mathcal{I}}_{\lambda.\eta\varrho_h}^h)^{-1}G_{\lambda.\eta\varrho_h}^h,\end{aligned} \tag{56}$$

where $G_{\lambda.\eta\tilde{\varrho}}=((G_{\lambda.\eta\varrho_1}^1)',\ldots,(G_{\lambda.\eta\varrho_{M_0}}^{M_0})')'$ is an M_0q_λ-vector mean zero Gaussian process with $\mathrm{cov}(G_{\lambda.\eta\tilde{\varrho}_1},G_{\lambda.\eta\tilde{\varrho}_2})=\tilde{\mathcal{I}}_{\lambda.\eta\tilde{\varrho}_1\tilde{\varrho}_2}$. Note that $G_{\lambda.\eta\tilde{\varrho}}$ corresponds to the residuals from projecting $\tilde{s}_{\lambda\tilde{\varrho}k}$ on $\tilde{s}_{\eta k}$. Define $\hat{t}_{\lambda\varrho_h}^h$ by

$$g_{\lambda\varrho_h}^h(\hat{t}_{\lambda\varrho_h}^h)=\inf_{t_\lambda\in v(\mathbb{R}^q)}g_{\lambda\varrho_h}^h(t_\lambda),\quad g_{\lambda\varrho_h}^h(t_\lambda):=(t_\lambda-Z_{\lambda\varrho_h}^h)'\tilde{\mathcal{I}}_{\lambda.\eta\varrho_h}^h(t_\lambda-Z_{\lambda\varrho_h}^h).$$

The following proposition gives the asymptotic null distribution of the LRTS for testing $H_0: M=M_0$. In the neighborhood of Υ_h^*, the log-likelihood function permits a similar quadratic approximation to the one we derived in Section 6. Consequently, the LRTS is asymptotically distributed as the maximum of M_0 random variables, each of which represents the asymptotic distribution of the LRTS that tests H_{0h}. Denote the parameter space for ϱ_h under restriction $p_{ij}\geq\epsilon$ by $\Theta_{\varrho_h\epsilon}$, and let $\tilde{\Theta}_{\varrho\epsilon}:=\Theta_{\varrho_1\epsilon}\times\ldots\times\Theta_{\varrho_{M_0}\epsilon}$.

Assumption 10. *$0<\inf_{\tilde{\varrho}\in\tilde{\Theta}_{\varrho\epsilon}}\lambda_{\min}(\tilde{\mathcal{I}}_{\tilde{\varrho}})\leq\sup_{\tilde{\varrho}\in\tilde{\Theta}_{\varrho\epsilon}}\lambda_{\max}(\tilde{\mathcal{I}}_{\tilde{\varrho}})<\infty$ for $\tilde{\mathcal{I}}_{\tilde{\varrho}}:=\lim_{k\to\infty}\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\tilde{\varrho}k}\tilde{s}_{\tilde{\varrho}k}')$, where $\tilde{s}_{\tilde{\varrho}k}$ is given in (54).*

Proposition 15. *Suppose Assumptions 1, 2, 4, 9, and 10 hold. Then, under* $H_0 : m = M_0$, $2[\ell_n(\hat{\vartheta}_{M_0+1}, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})] \Rightarrow \max_{h=1,\dots,M_0} \left\{ \sup_{\varrho_h \in \Theta^h_{\varrho\epsilon}} \left((\hat{t}^h_{\lambda\varrho_h})' \tilde{\mathcal{I}}^h_{\lambda.\eta\varrho_h} \hat{t}^h_{\lambda\varrho_h} \right) \right\}$.

7.2 Heteroscedastic normal distribution

As in Section 6.2, we assume that $Y_k \in \mathbb{R}$ in the j-th regime follows a normal distribution with regime-specific intercept and variance of which density is given by (24). Consider the following reparameterization similar to (25):

$$\begin{pmatrix} \zeta_h \\ \zeta_{h+1} \\ \sigma_h^2 \\ \sigma_{h+1}^2 \end{pmatrix} = \begin{pmatrix} \nu_\zeta + (1-\alpha_h)\lambda_\zeta \\ \nu_\zeta - \alpha_h \lambda_\zeta \\ \nu_\sigma + (1-\alpha_h)(2\lambda_\sigma + C_1\lambda_\mu^2) \\ \nu_\sigma - \alpha_h(2\lambda_\sigma + C_2\lambda_\mu^2) \end{pmatrix},$$

where $\nu_\zeta = (\nu_\mu, \nu_\beta')'$, $\lambda_\zeta = (\lambda_\mu, \lambda_\beta')'$, $C_1 := -(1/3)(1+\alpha_h)$, and $C_2 := (1/3)(2-\alpha_h)$. As in Section 7.1, we collect the reparameterized identified parameters into $\psi_h := (\vartheta_{xh}', \eta', \lambda')'$, where $\eta = (\gamma', \{\theta_j'\}_{j=1}^{h-1}, \nu_\zeta', \nu_\sigma, \{\theta_j'\}_{j=h+2}^{M_0+1})'$ and $\lambda := (\lambda_\zeta', \lambda_\sigma)'$. Similar to (26), define the reparameterized conditional density of y_k as

$$\begin{aligned} g^h_{\psi_h}(y_k|\overline{y}_{k-1}, x_k) &= \sum_{j \in \overline{J}_h} q_{kj} f(y_k|\overline{y}_{k-1}; \gamma, \theta_j) \\ &+ \mathbb{I}\{x_k \in J_h\} f\left(y_k|\overline{y}_{k-1}; \gamma, \nu_\zeta + (q_{kh}-\alpha_h)\lambda_\zeta, \nu_\sigma + (q_{kh}-\alpha_h)(2\lambda_\sigma + (C_2 - q_{kh})\lambda_\mu^2)\right). \end{aligned}$$

Let g_k^{h*}, ∇g_k^{h*}, and ∇f_{hk}^* denote $g^h_{\psi_h^*}(Y_k|\overline{Y}_{k-1}, X_k)$, $\nabla g^h_{\psi_h^*}(Y_k|\overline{Y}_{k-1}, X_k)$, and $\nabla f(Y_k|\overline{Y}_{k-1}; \gamma^*, \theta_h^*)$. It follows from (28) and a derivation similar to (50) that

$$\begin{aligned} \mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda_\mu^i} g_k^{h*}/g_k^{h*} \middle| \overline{Y}^k_{-\infty}\right] &= 0, \quad i = 1,2,3, \\ \mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda_\mu^4} g_k^{h*}/g_k^{h*} \middle| \overline{Y}^k_{-\infty}\right] &= \alpha_h(1-\alpha_h)b(\alpha_h)(\nabla_{\mu^4} f^*_{hk}/f^*_{hk})\mathbb{P}_{\vartheta^*_{M_0}}(X_k \in J_h|\overline{Y}^k_{-\infty}) \\ &= b(\alpha_h)\mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda_\sigma^2} g_k^{h*}/g_k^{h*} \middle| \overline{Y}^k_{-\infty}\right], \end{aligned} \tag{57}$$

which corresponds to (29) in testing homogeneity. Repeating the calculation leading to (51)–(53) and using (57) gives the following: first, (51) and (52) still hold; second, the elements of $\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ except for the $(1,1)$th element are given by (53) while adjusting the derivative with respect to λ_σ by multiplying by 2; third,

$$\frac{\nabla_{\lambda_\mu^2}\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi_h^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \alpha_h(1-\alpha_h)\sum_{t=1}^{k-1}\varrho_h^{k-t}\left(2\frac{\nabla_\mu f^*_{ht}}{f^*_{ht}}\frac{\nabla_\mu f^*_{hk}}{f^*_{hk}}\right)\mathbb{P}_{\vartheta^*_{M_0}}((X_t, X_k) \in J_h^2|\overline{Y}_0^k).$$

For $m \geq 0$, define $\zeta^h_{k,m}(\varrho_h) := \sum_{t=-m+1}^{k-1}\varrho_h^{k-t-1}2(\nabla_\mu f^*_{ht}\nabla_\mu f^*_{hk}/f^*_{ht}f^*_{hk})\mathbb{P}_{\vartheta^*_{M_0}}((X_t, X_k) \in J_h^2|\overline{Y}_0^k)$.

Define $s^h_{\lambda\lambda\varrho_h k}$ as in (55) but using density (24), and denote each element of $s^h_{\lambda\lambda\varrho_h k}$ as

$$s^h_{\lambda\lambda\varrho_h k} := \begin{pmatrix} * & s^h_{\lambda_{\mu\beta}\varrho_h k} & s^h_{\lambda_{\mu\sigma}\varrho_h k} \\ s^h_{\lambda_{\beta\mu}\varrho_h k} & s^h_{\lambda_{\beta\beta}\varrho_h k} & s^h_{\lambda_{\beta\sigma}\varrho_h k} \\ s^h_{\lambda_{\sigma\mu}\varrho_h k} & s^h_{\lambda_{\beta\sigma}\varrho_h k} & s^h_{\lambda_{\sigma\sigma}\varrho_h k} \end{pmatrix}.$$

Define $\tilde{s}_{\tilde{\varrho}k}$ as in (54) with redefining $s^h_{\lambda\varrho_h k}$ in (54) as

$$s^h_{\lambda\varrho_h k} := \left(\zeta^h_{k,0}(\varrho_h)'/2 \quad 2s^h_{\lambda_{\mu\sigma}\varrho_h k} \quad 2s^h_{\lambda_{\sigma\sigma}\varrho_h k} \quad (s^h_{\lambda_{\beta\mu}\varrho_h k})' \quad 2(s^h_{\lambda_{\beta\sigma}\varrho_h k})' \quad V(s^h_{\lambda_{\beta\beta}\varrho_h k})'\right)'. \tag{58}$$

Define $\mathcal{I}^h_{\lambda.\eta\varrho_h}$ and $Z^h_{\lambda\varrho_h}$ as in (56) with $s^h_{\lambda\varrho_h k}$ defined in (58). Let $Z^h_{\lambda 0}$ and $\mathcal{I}^h_{\lambda.\eta 0}$ denote $Z^h_{\lambda\varrho_h}$ and $\mathcal{I}^h_{\lambda.\eta\varrho_h}$ evaluated at $\varrho_h = 0$. Define Λ^1_λ as in (37), and define $\Lambda^2_{\lambda\varrho_h}$ as in (37) with replacing ϱ with ϱ_h. Similar to (38), define $\hat{t}^{h1}_\lambda$ and $\hat{t}^{h2}_{\lambda\varrho_h}$ by $r_\lambda(\hat{t}^{h1}_\lambda) = \inf_{t_\lambda \in \Lambda^1_\lambda} r^h_\lambda(t_\lambda)$ and $r_{\lambda\varrho_h}(\hat{t}^{h2}_{\lambda\varrho_h}) = \inf_{t_\lambda \in \Lambda^2_{\lambda\varrho_h}} r^h_{\lambda\varrho_h}(t_\lambda)$, where $r^h_\lambda(t_\lambda) := (t_\lambda - Z^h_{\lambda 0})'\mathcal{I}^h_{\lambda.\eta 0}(t_\lambda - Z^h_{\lambda 0})$ and $r^h_{\lambda\varrho_h}(t_\lambda) := (t_\lambda - Z^h_{\lambda\varrho_h})'\mathcal{I}^h_{\lambda.\eta\varrho_h}(t_\lambda - Z^h_{\lambda\varrho_h})$. Define a stochastic process $W^h(\varrho_h)$ as $W^h(\varrho_h) := \max\{\mathbb{I}\{\varrho_h = 0\}(\hat{t}^{h1}_\lambda)'\mathcal{I}^h_{\lambda.\eta 0}\hat{t}^{h1}_\lambda, (\hat{t}^{h2}_{\lambda\varrho_h})'\mathcal{I}^h_{\lambda.\eta\varrho_h}\hat{t}^{h2}_{\lambda\varrho_h}\}$.

The following proposition establishes the asymptotic null distribution of the LRT statistic. As in the non-normal case, the LRTS is asymptotically distributed as the maximum of M_0 random variables.

Assumption 11. *Assumption 10 holds when $\tilde{s}_{\tilde{\varrho}_h k}$ is given in (58).*

Proposition 16. *Suppose Assumptions 1, 2, 4, 9, and 11 hold and the component density for the j-th regime is given by (24). Then, under $H_0 : m = M_0$,*
$2[\ell_n(\hat{\vartheta}_{M_0+1}, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})] \Rightarrow \max_{h=1,\ldots,M_0}\{\sup_{\varrho_h \in \Theta^h_{\varrho\epsilon}} W^h(\varrho_h)\}$.

7.3 Homoscedastic normal distribution

As in Section 6.3, we assume that $Y_k \in \mathbb{R}$ in the j-th regime follows a normal distribution with regime-specific intercept and common variance whose density is given by (39).

The asymptotic distribution of the LRTS is derived by using a reparameterization

$$\begin{pmatrix} \theta_h \\ \theta_{h+1} \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \nu_\theta + (1-\alpha_h)\lambda \\ \nu_\theta - \alpha_h\lambda \\ \nu_\sigma - \alpha_h(1-\alpha_h)\lambda^2_\mu \end{pmatrix},$$

similar to (40) and following the derivation in Sections 6.3 and 7.2. For brevity, we omit details in derivation. Define $s^h_{\lambda\lambda\varrho_h k}$ as in (55), and denote each element of $s^h_{\lambda\lambda\varrho_h k}$ as

$$s^h_{\lambda\lambda\varrho_h k} = \begin{pmatrix} * & s^h_{\lambda_{\mu\beta}\varrho_h k} \\ s^h_{\lambda_{\beta\mu}\varrho_h k} & s^h_{\lambda_{\beta\beta}\varrho_h k} \end{pmatrix}.$$

Define $\tilde{s}_{\tilde{\varrho}k}$ as in (54) with redefining $s^h_{\lambda\varrho_h k}$ in (54) as

$$s^h_{\lambda\varrho_h k} := \left(\zeta^h_{k,0}(\varrho_h)'/2 \quad s^h_{\lambda^3_\mu k}/3! \quad s^h_{\lambda^4_\mu k}/4! \quad (s^h_{\lambda_{\beta\mu}\varrho k})' \quad V(s^h_{\lambda_{\beta\beta}\varrho k})'\right)', \tag{59}$$

where $s^h_{\lambda^i_\mu k} := \nabla_{\mu^i} f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*_h)/f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*_h)$ for $i = 3,4$.

The following proposition establishes the asymptotic null distribution of the LRT statistic.

Assumption 12. *Assumption 10 holds when $\tilde{s}_{\tilde{\varrho}_h k}$ is given in (59).*

Proposition 17. *Suppose Assumptions 1, 2, 4, 9, and 12 hold and the component density for the j-th regime is given by (39). Then, under $H_0 : m = M_0$, $2[\ell_n(\hat{\vartheta}_{M_0+1},\xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0},\xi_{M_0})] \Rightarrow \max_{h=1,\ldots,M_0}\{\sup_{\varrho_h\in\Theta_{\varrho_h\epsilon}} \max\{\mathbb{I}\{\varrho_h = 0\}(\hat{t}^{h1}_\lambda)'\mathcal{I}^h_{\lambda.\eta 0}\hat{t}^{h1}_\lambda, (\hat{t}^{h2}_{\lambda\varrho_h})'\mathcal{I}^h_{\lambda.\eta\varrho_h}\hat{t}^{h2}_{\lambda\varrho_h}\}\}$, where $\hat{t}^{h1}_\lambda$ and $\hat{t}^{h2}_{\lambda\varrho_h}$ are defined as in Proposition 16 but in terms of $(Z^h_{\lambda\varrho_h},\mathcal{I}^h_{\lambda.\eta\varrho_h},Z^h_{\lambda 0},\mathcal{I}^h_{\lambda.\eta 0})$ constructed with $s^h_{\lambda\varrho_h k}$ given in (59) and Λ^1_λ and $\Lambda^2_{\lambda\varrho_h}$ defined as in (46) but replacing ϱ with ϱ_h.*

8 Appendix

8.1 Definition of $\Phi^{\mathcal{I}(5)}_{\vartheta\mathcal{T}(5)}[\mathcal{F}]$ and $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$

Define

$$\begin{aligned}
\Phi^{\mathcal{I}(5)}_{\vartheta\mathcal{T}(5)}[\mathcal{F}] &:= \frac{1}{|\sigma(\mathcal{I}(5))|}\sum_{(\ell_1,\ldots,\ell_5)\in\sigma(\mathcal{I}(5))}\Bigg(\mathbb{E}^c_\vartheta\left[\phi^{\ell_1}_{\theta t_1}\phi^{\ell_2}_{\theta t_2}\phi^{\ell_3}_{\theta t_3}\phi^{\ell_4}_{\theta t_4}\phi^{\ell_5}_{\theta t_5}\middle|\mathcal{F}\right] \\
&\qquad - \sum_{(\{a,b,c\},\{d,e\})\in\sigma_5}\mathbb{E}^c_\vartheta\left[\phi^{\ell_a}_{\theta t_a}\phi^{\ell_b}_{\theta t_b}\phi^{\ell_c}_{\theta t_c}\middle|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi^{\ell_d}_{\theta t_d}\phi^{\ell_e}_{\theta t_e}\middle|\mathcal{F}\right]\Bigg), \\
\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}] &:= \mathbb{E}^c_\vartheta\left[\phi_{\theta t_1}\phi_{\theta t_2}\phi_{\theta t_3}\phi_{\theta t_4}\phi_{\theta t_5}\phi_{\theta t_6}|\mathcal{F}\right] - \sum_{(\{a,b,c,d\},\{e,f\})\in\sigma_{61}}\mathbb{E}^c_\vartheta\left[\phi_{\theta t_a}\phi_{\theta t_b}\phi_{\theta t_c}\phi_{\theta t_d}|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi_{\theta t_e}\phi_{\theta t_f}\middle|\mathcal{F}\right] \\
&\qquad - \sum_{(\{a,b,c\},\{d,e,f\})\in\sigma_{62}}\mathbb{E}^c_\vartheta\left[\phi_{\theta t_a}\phi_{\theta t_b}\phi_{\theta t_c}|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi_{\theta t_d}\phi_{\theta t_e}\phi_{\theta t_f}\middle|\mathcal{F}\right] \\
&\qquad + 2\sum_{(\{a,b\},\{c,d\},\{e,f\})\in\sigma_{63}}\mathbb{E}^c_\vartheta\left[\phi_{\theta t_a}\phi_{\theta t_b}|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi_{\theta t_c}\phi_{\theta t_d}|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi_{\theta t_e}\phi_{\theta t_f}\middle|\mathcal{F}\right],
\end{aligned} \tag{60}$$

where

$$\begin{aligned}
\sigma_5 &:= \text{the set of } \tbinom{5}{3} = 10 \text{ partitions of } \{1,2,3,4,5\} \text{ of the form } \{a,b,c\},\{d,e\}, \\
\sigma_{61} &:= \text{the set of } \tbinom{6}{4} = 15 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b,c,d\},\{e,f\}, \\
\sigma_{62} &:= \text{the set of } \tbinom{6}{3}/2 = 10 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b,c\},\{d,e,f\}, \\
\sigma_{63} &:= \text{the set of } \tbinom{6}{2}\tbinom{4}{2}/6 = 15 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b\},\{c,d\},\{e,f\}.
\end{aligned} \tag{61}$$

8.2 Proof of Propositions and Corollaries

For notational brevity, we suppress W_{-m}^n from the conditioning variable unless confusion might arise.

Proof of Proposition 1. The proof is essentially identical to the proof of Lemma 2 in DMR. Therefore, the details are omitted. The only difference from DMR is (i) we do not impose Assumption (A2) of DMR, but this does not affect the proof because Assumption (A2) is not used in the proof of Lemma 2 in DMR, and (ii) we have W_1^n, but our Lemma 9 extends Corollary 1 of DMR to accommodate W_k's. Consequently, the argument of the proof of DMR goes through. □

Proof of Proposition 2. Let $m_{\vartheta k} := t_\vartheta' s_{\pi k} + r_{\vartheta k}$, so that $l_{\vartheta k x_0} - 1 = m_{\vartheta k} + u_{\vartheta k x_0}$. Observe that

$$\max_{1\le k\le n} \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} |m_{\vartheta k}| = \max_{1\le k\le n} \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} |t_\vartheta' s_{\pi k} + r_{\vartheta k}| = o_p(1), \tag{62}$$

from Assumption 3(a)(c) and Lemma 8. Define $h_{\vartheta k x_0} := \sqrt{l_{\vartheta k x_0}} - 1$. We first show

$$\sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} \left| nP_n(h_{\vartheta k x_0}^2) - n t_\vartheta' \mathcal{I}_\pi t_\vartheta/4 \right| = o_p(1). \tag{63}$$

To show (63), write $4P_n(h_{\vartheta k x_0}^2)$ as

$$4P_n(h_{\vartheta k x_0}^2) = P_n\left(\frac{4(l_{\vartheta k x_0}-1)^2}{(\sqrt{l_{\vartheta k x_0}}+1)^2}\right) = P_n(l_{\vartheta k x_0}-1)^2 - P_n\left((l_{\vartheta k x_0}-1)^3 \frac{(\sqrt{l_{\vartheta k x_0}}+3)}{(\sqrt{l_{\vartheta k x_0}}+1)^3}\right). \tag{64}$$

It follows from Assumption 3(a)(b)(c)(e) and $(E|XY|)^2 \le E|X|^2 E|Y|^2$ that, uniformly for $\vartheta \in \mathcal{N}_\varepsilon$,

$$\begin{aligned} P_n(l_{\vartheta k x_0}-1)^2 &= t_\vartheta' P_n(s_{\pi k} s_{\pi k}') t_\vartheta + 2 t_\vartheta' P_n[s_{\pi k}(r_{\vartheta k} + u_{\vartheta k x_0})] + P_n(r_{\vartheta k} + u_{\vartheta k x_0})^2 \\ &= t_\vartheta' \mathcal{I}_\pi t_\vartheta + o_p(|t_\vartheta|^2) + O_p(n^{-1}|t_\vartheta||\psi-\psi^*|) + O_p(n^{-1}|\psi-\psi^*|^2). \end{aligned} \tag{65}$$

Then, (63) holds because the second term on the right of (64) is bounded by, from (62), $P_n(m_{\vartheta k}^2) = t_\vartheta' \mathcal{I}_\pi t_\vartheta + o_p(|t_\vartheta|^2)$, and Assumption 3(e),

$$\begin{aligned} &\mathcal{C} \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} P_n\left[|m_{\vartheta k}|^3 + 3m_{\vartheta k}^2|u_{\vartheta k x_0}| + 3|m_{\vartheta k}|u_{\vartheta k x_0}^2\right] + \mathcal{C} \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} P_n(|u_{\vartheta k x_0}|^3) \\ &\le o_p(1) \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} P_n\left[m_{\vartheta k}^2 + |m_{\vartheta k}||u_{\vartheta k x_0}| + u_{\vartheta k x_0}^2\right] + \mathcal{C} \sup_{\vartheta\in\mathcal{N}_{c/\sqrt{n}}} P_n(|u_{\vartheta k x_0}|^3) = o_p(n^{-1}). \end{aligned}$$

Consider the following expansion of $h_{\vartheta k x_0}$:

$$h_{\vartheta k x_0} = (l_{\vartheta k x_0}-1)/2 - h_{\vartheta k x_0}^2/2 = (t_\vartheta' s_{\pi k} + r_{\vartheta k} + u_{\vartheta k x_0})/2 - h_{\vartheta k x_0}^2/2. \tag{66}$$

It follows from (63), (66), and Assumption 3(d)(e) that $nP_n(h_{\vartheta k x_0}) = \sqrt{n} t_\vartheta' \nu_n(s_{\pi k})/2 - n t_\vartheta \mathcal{I}_\pi t_\vartheta'/8 +$

$o_p(1)$ uniformly for $\vartheta \in \mathcal{N}_{c/\sqrt{n}}$. Using the Taylor expansion of $2\log(1+x) = 2x - x^2(1+o(1))$ for small x, we have, uniformly for $\vartheta \in \mathcal{N}_{c/\sqrt{n}}$,

$$
\begin{aligned}
\ell_n(\psi,\pi,x_0) - \ell_n(\psi^*,\pi,x_0) &= 2\sum_{k=1}^n \log(1+h_{\vartheta k x_0}) = nP_n(2h_{\vartheta k x_0} - [1+o_p(1)]h_{\vartheta k x_0}^2) \\
&= \sqrt{n}t_\vartheta' \nu_n(s_{\pi k}) - t_\vartheta \mathcal{I}_\pi t_\vartheta'/4 + nP_n(h_{\vartheta k x_0}^2) + o_p(1) \\
&= \sqrt{n}t_\vartheta' \nu_n(s_{\pi k}) - t_\vartheta \mathcal{I}_\pi t_\vartheta'/2 + o_p(1),
\end{aligned}
$$

giving the stated result. □

Proof of Proposition 3. For part (a), applying the inequality $\log(1+x) \le x$ to the log-likelihood ratio function and using (66) give

$$
\ell_n(\psi,\pi,x_0) - \ell_n(\psi^*,\pi,x_0) = 2\sum_{k=1}^n \log(1+h_{\vartheta k x_0}) \le 2nP_n(h_{\vartheta k x_0}) = \sqrt{n}\nu_n(l_{\vartheta k x_0} - 1) - nP_n(h_{\vartheta k x_0}^2). \tag{67}
$$

From the first equality in (64), for some $\delta > 0$,

$$
\begin{aligned}
P_n(h_{\vartheta k x_0}^2) &\ge P_n\left(\frac{(l_{\vartheta k x_0} - 1)^2}{l_{\vartheta k x_0} + 1}\right) \\
&\ge \frac{1}{\delta+1} P_n\left(\mathbb{I}\{l_{\vartheta k x_0} \le \delta\}(l_{\vartheta k x_0} - 1)^2\right) \\
&\ge \frac{1}{\delta+1}\left[P_n((l_{\vartheta k x_0} - 1)^2) - P_n\left(\mathbb{I}\{l_{\vartheta k x_0} > \delta\}(l_{\vartheta k x_0} - 1)^2\right)\right].
\end{aligned}
$$

From Assumption 3(a)(b)(c)(e), we have $B := \sup_{\vartheta\in\mathcal{N}_\varepsilon} |l_{\vartheta k x_0} - 1| \in L^2(\mathbb{P}_{\vartheta^*})$, and hence $\lim_{\delta\to\infty} \sup_{\vartheta\in\mathcal{N}_\epsilon} P_n\left(\mathbb{I}\{l_{\vartheta k x_0} > \delta\}(l_{\vartheta k x_0} - 1)^2\right) \le \lim_{\delta\to\infty} P_n\left(\mathbb{I}\{B+1 > \delta\}B^2\right) = 0$ almost surely. Therefore, by choosing δ sufficiently large, it follows from (65) and Assumption 3(f) that, uniformly for $\vartheta \in \mathcal{N}_\varepsilon$,

$$
P_n(h_{\vartheta k x_0}^2) \ge \kappa t_\vartheta' \mathcal{I}_\pi t_\vartheta + o_p(|t_\vartheta|^2) + O_p(n^{-1}|t_\vartheta||\psi - \psi^*|) + O_p(n^{-1}|\psi-\psi^*|^2), \tag{68}
$$

for $\kappa = (2(\delta+1))^{-1} > 0$.

Because $\sqrt{n}\nu_n(l_{\vartheta k x_0} - 1) = \sqrt{n}t_\vartheta'[Z_n(\vartheta) + O_p(1)] + O_p(1)$ from Assumption 3(d)(e)(g), it follows from (67) and (68) that

$$
0 \le \ell_n(\psi,\pi,x_0) - \ell_n(\psi^*,\pi,x_0) \le \sqrt{n}t_\vartheta'[Z_n(\vartheta) + O_p(1)] - \kappa n t_\vartheta' \mathcal{I}_\pi t_\vartheta + o_p(n|t_\vartheta|^2) + O_p(1). \tag{69}
$$

The rest of the proof follows the proof of Theorem 1 of Andrews (1999). Let $T_n := \mathcal{I}_\pi^{1/2}\sqrt{n}t_\vartheta$. In view of Assumption 3(f)(g), we can write (69) as $0 \le ||T_n||O_p(1) - \kappa||T_n||^2 + o_p(||T_n||^2) + O_p(1)$. Rearranging this equation gives $||T_n||^2 \le 2||T_n||\varsigma_n + O_p(1)$ with $\varsigma_n = O_p(1)$. Then, $(||T_n|| - \varsigma_n)^2 \le \varsigma_n^2 + O_p(1) = O_p(1)$, and taking square roots gives $||T_n|| \le O_p(1)$, giving part (a). Part (b) follows from part (a) and Proposition 2. □

Proof of Corollary 1. Because logarithm is monotone, we have $\min_{x_0 \in \mathcal{X}} \ell_n(\psi, \pi, x_0) \leq \ell_n(\psi, \pi, \xi) \leq \max_{x_0 \in \mathcal{X}} \ell_n(\psi, \pi, x_0)$. Part (a) then follows from Proposition 2. For part (b), note that we have $\vartheta \in A_n$ only if $\vartheta \in A_n(x_0)$ for some x_0. Consequently, part (b) follows from Proposition 3. □

Proof of Proposition 4. The stated result follows from writing $\nabla^j \ell_{k,m,x}(\vartheta) = \nabla^j \log p_\vartheta(Y_{-m+1}^k | \overline{Y}_{-m}, X_{-m} = x) - \nabla^j \log p_\vartheta(Y_{-m+1}^{k-1} | \overline{Y}_{-m}, X_{-m} = x)$, applying Lemma 1 to the right hand side, and noting that $\nabla^j \log p_\vartheta(Y_{-m+1}^k, X_{-m+1}^k | \overline{Y}_{-m}, X_{-m}) = \sum_{t=-m+1}^k \phi^j(\vartheta, \overline{Z}_{t-1}^t)$ (see (1) and (6)). The result for $\nabla^j \ell_{k,m,x}(\vartheta)$ with $j = 1, 2$ is also given in DMR (p. 2272 and pp. 2276-7). For $j = 3$, the term $\Delta_{2,k,m,x}^{2,1}(\vartheta)$ follows from $\sum_{t_1=-m+1}^k \sum_{t_2=-m+1}^k \mathbb{E}_\vartheta^c [\phi_{\vartheta t_1}^2 \phi_{\vartheta t_2}^1 | \overline{Y}_{-m}^k, X_{-m} = x] = \sum_{t_1=-m+1}^k \sum_{t_2=-m+1}^k \Phi_{\vartheta t_1 t_2}^{2,1} [\overline{Y}_{-m}^k, X_{-m} =$ $x]$. For $j = 4$, note that when we apply Lemma 1 to $\nabla^4 \log p_\vartheta(Y_{-m+1}^k | \overline{Y}_{-m}, X_{-m} = x)$, the last two terms on the right hand side can be written as $\sum_{\mathcal{T}(4) \in \{-m+1,\ldots,k\}^4} \Phi_{\vartheta \mathcal{T}(4)}^{1,1,1,1} [\overline{Y}_{-m}^k, X_{-m} = x]$. The result for $j = 5$ follows from a similar argument. For $j = 6$, note that when we apply Lemma 1 to $\nabla^6 \log p_\vartheta(Y_{-m+1}^k | \overline{Y}_{-m}, X_{-m} = x)$, the last four terms on the right hand side can be written as $\sum_{\mathcal{T}(6) \in \{-m+1,\ldots,k\}^6} \Phi_{\vartheta \mathcal{T}(6)}^{\mathcal{I}(6)} [\overline{Y}_{-m}^k, X_{-m} = x]$. □

Proof of Proposition 5. The stated result follows from Proposition 4 and Lemma 3 and noting $q_1 = 6q_0, q_2 = 5q_0, q_3 = 4q_0, \ldots, q_6 = q_0$. □

Proof of Proposition 6. First, we prove part (a). The proof of part (b) is essentially the same as that of part (a) and hence omitted. Observe that

$$\nabla l_{k,m,x}^j(\vartheta) - \overline{\nabla} l_{k,m}^j(\vartheta) = \Psi_{k,m,x}^j(\vartheta) \left(\frac{p_\vartheta(Y_k | \overline{Y}_{-m}^{k-1}, X_{-m} = x)}{p_{\vartheta^*}(Y_k | \overline{Y}_{-m}^{k-1}, X_{-m} = x)} - \frac{\overline{p}_\vartheta(Y_k | \overline{Y}_{-m}^{k-1})}{\overline{p}_{\vartheta^*}(Y_k | \overline{Y}_{-m}^{k-1})} \right) + \frac{\overline{p}_\vartheta(Y_k | \overline{Y}_{-m}^{k-1})}{\overline{p}_{\vartheta^*}(Y_k | \overline{Y}_{-m}^{k-1})} \left(\Psi_{k,m,x}^j(\vartheta) - \Psi_{k,m}^j(\vartheta) \right),$$

where

$$\Psi_{k,m,x}^j(\vartheta) := \frac{\nabla^j p_\vartheta(Y_k | \overline{Y}_{-m}^{k-1}, X_{-m} = x)}{p_\vartheta(Y_k | \overline{Y}_{-m}^{k-1}, X_{-m} = x)}, \quad \overline{\Psi}_{k,m}^j(\vartheta) := \frac{\nabla^j \overline{p}_\vartheta(Y_k | \overline{Y}_{-m}^{k-1})}{\overline{p}_\vartheta(Y_k | \overline{Y}_{-m}^{k-1})}.$$

In view of Lemma 4 and the Hölder's inequality, part (a) holds if we show, for $j = 1, \ldots, 6$, for some $A_j \in L^{q_0}(\mathbb{P}_{\vartheta^*})$ and uniformly in $x \in \mathcal{X}$ and $\vartheta \in \mathcal{N}^*$,

$$(A) \quad \Psi_{k,m,x}^j(\vartheta) \in L^{q_0}(\mathbb{P}_{\vartheta^*}), \quad \text{and} \quad (B) \quad |\Psi_{k,m,x}^j(\vartheta) - \overline{\Psi}_{k,m}^j(\vartheta)| \leq A_j (k+m)^6 \rho_*^{k+m-1}. \qquad (70)$$

Note that, from (87) we have, suppressing (ϑ) and superscript 1 from $\nabla^1\ell_{k,m,x}$,

$$\begin{aligned}
\Psi^1_{k,m,x} &= \nabla\ell_{k,m,x}, \quad \Psi^2_{k,m,x} = \nabla^2\ell_{k,m,x} + (\nabla\ell_{k,m,x})^2,\\
\Psi^3_{k,m,x} &= \nabla^3\ell_{k,m,x} + 3\nabla^2\ell_{k,m,x}\nabla\ell_{k,m,x} + (\nabla\ell_{k,m,x})^3,\\
\Psi^4_{k,m,x} &= \nabla^4\ell_{k,m,x} + 4\nabla^3\ell_{k,m,x}\nabla\ell_{k,m,x} + 3(\nabla^2\ell_{k,m,x})^2 + 6\nabla^2\ell_{k,m,x}(\nabla\ell_{k,m,x})^2 + (\nabla\ell_{k,m,x})^4,\\
\Psi^5_{k,m,x} &= \nabla^5\ell_{k,m,x} + 5\nabla^4\ell_{k,m,x}\nabla\ell_{k,m,x} + 10\nabla^3\ell_{k,m,x}\nabla^2\ell_{k,m,x} + 10\nabla^3\ell_{k,m,x}(\nabla\ell_{k,m,x})^2\\
&\quad + 15(\nabla^2\ell_{k,m,x})^2\nabla\ell_{k,m,x} + 10\nabla^2\ell_{k,m,x}(\nabla\ell_{k,m,x})^3 + (\nabla\ell_{k,m,x})^5,\\
\Psi^6_{k,m,x} &= \nabla^6\ell_{k,m,x} + 6\nabla^5\ell_{k,m,x}\nabla\ell_{k,m,x} + 15\nabla^4\ell_{k,m,x}\nabla^2\ell_{k,m,x} + 15\nabla^4\ell_{k,m,x}(\nabla\ell_{k,m,x})^2\\
&\quad + 10(\nabla^3\ell_{k,m,x})^2 + 60\nabla^3\ell_{k,m,x}\nabla^2\ell_{k,m,x}\nabla\ell_{k,m,x} + 20\nabla^3\ell_{k,m,x}(\nabla\ell_{k,m,x})^3\\
&\quad + 15(\nabla^2\ell_{k,m,x})^3 + 45(\nabla^2\ell_{k,m,x})^2(\nabla\ell_{k,m,x})^2 + 15\nabla^2\ell_{k,m,x}(\nabla\ell_{k,m,x})^4 + (\nabla\ell_{k,m,x})^6,
\end{aligned}$$

and $\overline{\Psi}^j_{k,m}$ is written analogously with $\nabla^j\overline{\ell}_{k,m}$ replacing $\nabla^j\ell_{k,m,x}$.

Consequently, (A) of (70) follows from Proposition 5(c), the Cauchy-Schwartz inequality, and the Hölder's inequality such as $E|ab|^{q_0} \le (E|a|^{3q_0/2})^{2/3}(E|b|^{3q_0})^{1/3} < \infty$ when $a \in L^{3q_0/2}(\mathbb{P}_{\vartheta^*})$ and $b \in L^{3q_0}(\mathbb{P}_{\vartheta^*})$, $E|ab|^{q_0} \le (E|a|^{6q_0/5})^{5/6}(E|b|^{6q_0})^{1/6} < \infty$ when $a \in L^{6q_0/5}(\mathbb{P}_{\vartheta^*})$ and $b \in L^{6q_0}(\mathbb{P}_{\vartheta^*})$, and $E|abc|^{q_0} \le (E|ab|^{6q_0/5})^{5/6}(E|c|^{6q_0})^{1/6} \le [(E|a|^{2q_0})^{3/5}(E|b|^{3q_0})^{2/5}]^{5/6}(E|c|^{6q_0})^{1/6} < \infty$ when $a \in L^{2q_0}(\mathbb{P}_{\vartheta^*})$, $b \in L^{3q_0}(\mathbb{P}_{\vartheta^*})$, $c \in L^{6q_0}(\mathbb{P}_{\vartheta^*})$. (B) of (70) follows from Proposition 5(a)(c), the relation $ab - cd = a(b-c) - c(a-d)$, $a^n - b^n = (a-b)\sum_{i=0}^{n-1}(a^{n-1-i}b^i)$, and the Hölder's inequality.

Part (c) follows from parts (a) and (b). For part (d), $\{\nabla^j l_{k,m,x}(\vartheta)\}_{m\ge0}$ is uniformly in $L^{\max\{2,\dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$ from writing $\nabla l^j_{k,m,x}(\vartheta) = [\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m} = x)/\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m} = x)]\Psi^j_{k,m,x}(\vartheta)$, and using (70) and Lemma 4. From a similar argument, $\{\overline{\nabla}^j l_{k,m}(\vartheta)\}_{m\ge0}$ is also uniformly in $L^{\max\{2,\dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$. Therefore, $\nabla^j l_{k,\infty}(\vartheta)$ is uniformly in $L^{\max\{2,\dim(\vartheta)\}}(\mathbb{P}_{\vartheta^*})$ from part (c) and the completeness of $L^q(\mathbb{P}_{\vartheta^*})$, and part (d) is proven. □

Proof of Proposition 7. Consistency of $\hat{\vartheta}_1$ follows from Theorem 2.1 of Newey and McFadden (1994), because (i) ϑ_1^* uniquely maximizes $\mathbb{E}_{\vartheta_1^*}\log f(Y_1|\overline{Y}_0, W_1;\gamma,\theta)$ from Assumption 5(c), and (ii) $\sup_{\vartheta_1\in\Theta_1}|n^{-1}\ell_{0,n}(\vartheta_1) - \mathbb{E}_{\vartheta_1^*}\log f(Y_1|\overline{Y}_0, W_1;\gamma,\theta)| \to_p 0$ and $\mathbb{E}_{\vartheta_1^*}\log f(Y_1|\overline{Y}_0, W_1;\gamma,\theta)$ is continuous because (Y_k, W_k) is strict stationary and ergodic from Assumption 1(e) and
$\mathbb{E}_{\vartheta_1^*}\sup_{\vartheta_1\in\Theta_1}|\log f(Y_1|\overline{Y}_0, W_1;\gamma,\theta)| \le \max\{\mathbb{E}_{\vartheta_1^*}|\log(b_-(\overline{Y}_0, W, Y_1)|, \log b_+\} < \infty$ from Assumption 2(c).

We proceed to prove the consistency of $\hat{\vartheta}_2$. Define, similarly to pp. 2265–2266 in DMR, $\Delta_{k,m,x}(\vartheta_2) := \log p_{\vartheta_2}(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1}, X_{-m} = x)$, $\Delta_{k,m}(\vartheta_2) := \log p_{\vartheta_2}(Y_k|\overline{Y}^{k-1}_{-m}, W^n_{-m+1})$, $\Delta_{k,\infty}(\vartheta_2) := \lim_{m\to\infty}\Delta_{k,m}(\vartheta_2)$, and $\ell(\vartheta_2) := \mathbb{E}_{\vartheta_1^*}[\Delta_{0,\infty}(\vartheta)]$. Observe that Lemmas 3, 4 and Proposition 2 of DMR hold for our $\{\Delta_{k,m,x}(\vartheta_2), \Delta_{k,m}(\vartheta_2), \Delta_{k,\infty}(\vartheta_2), \ell_n(\vartheta_2, x_0), \ell(\vartheta_2)\}$ under our assumptions because (i) their Assumption (A2), which we do not assume, is not used in the proof of their Lemmas 3 and 4 and Proposition 2, and (ii) our Lemma 9 extends Corollary 1 of DMR to accommodate W_k's. It follows that (i) $\ell(\vartheta_2)$ is maximized if and only if $\vartheta_2 \in \Gamma^*$ from Assumption 2(d) because $\mathbb{E}_{\vartheta_1^*}[\log p_{\vartheta_2}(Y_1|\overline{Y}^0_{-m}, W^n_{-m})]$ converges to $\ell(\vartheta_2)$ uniformly in ϑ_2 as $m \to \infty$ from Lemma

3 of DMR and the dominated convergence theorem, (ii) $\ell(\vartheta_2)$ is continuous from Lemma 4 of DMR, and (iii) $\sup_{\xi_2}\sup_{\vartheta_2\in\Theta_2}|n^{-1}\ell_n(\vartheta_2,\xi_2)-\ell(\vartheta_2)|=o_p(1)$ holds from Proposition 2 of DMR and $\ell_n(\vartheta_2,\xi_2)\in[\min_{x_0}\ell_n(\vartheta_2,x_0),\max_{x_0}\ell_n(\vartheta_2,x_0)]$. Consequently, $\inf_{\vartheta_2\in\Gamma^*}|\hat{\vartheta}_2-\vartheta_2|\to_p 0$ follows from Theorem 2.1 of Newey and McFadden (1994) with an adjustment for the fact that the maximizer of $\ell(\vartheta_2)$ is a set, not a singleton. □

Proof of Proposition 8. We prove the stated result by applying Corollary 1 to $l_{\vartheta k x_0}-1$ with $l_{\vartheta k x_0}$ defined in (4). Because the first and second derivatives of $l_{\vartheta k x_0}-1$ play the role of the score, we expand $l_{\vartheta k x_0}-1$ with respect to ψ up to the third order. Let $d=\dim(\psi)$. Recall that the p-th order Taylor expansion of $f(x)$ with $x\in\mathbb{R}^d$ around $x=x^*$ is given by

$$f(x^*)+\sum_{j=1}^{p}\sum_{(u_1,\dots,u_d)}^{j}\frac{1}{u_1!u_2!\cdots u_d!}\frac{\partial^j f(x^*)}{\partial x_1^{u_1}\partial x_2^{u_2}\cdots\partial x_d^{u_d}}(x_1-x_1^*)^{u_1}(x_2-x_2^*)^{u_2}\cdots(x_d-x_d^*)^{u_d},$$

where $\sum_{(u_1,\dots,u_d)}^{j}$ denotes the sum over the sets of nonnegative integers $(u_1,\dots,u_d)$ such that $u_1+u_2+\cdots u_d=j$. For $m\geq 0$ and $j=1,2,\dots$, let $\Lambda^j_{k,m,x_{-m}}(\psi,\pi)$ denote the vector that collects the terms of the form

$$\frac{1}{p_{\psi^*\pi}(Y_k|\overline{Y}_{-m}^{k-1},x_{-m})}\frac{1}{u_1!u_2!\cdots u_d!}\frac{\partial^j p_{\psi\pi}(Y_k|\overline{Y}_{-m}^{k-1},x_{-m})}{\partial\psi_1^{u_1}\partial\psi_2^{u_2}\cdots\partial\psi_d^{u_d}};\qquad \begin{array}{l}u_j=0,1,\dots,\\ u_1+u_2+\cdots+u_d=j,\end{array}\tag{71}$$

and let $\Delta^j\psi$ be the vector that collects the corresponding elements of $\psi-\psi^*$, i.e., the terms of the form $(\psi_1-\psi_1^*)^{u_1}(\psi_2-\psi_2^*)^{u_2}\cdots(\psi_d-\psi_d^*)^{u_d}$. Define $\Lambda^j_{k,m}(\psi,\pi)$ similarly to $\Lambda^j_{k,m,x_{-m}}(\psi,\pi)$ by replacing $p_{\psi\pi}(Y_k|\overline{Y}_{-m}^{k-1},x_{-m})$ in (71) with $\overline{p}_{\psi\pi}(Y_k|\overline{Y}_{-m}^{k-1})$. With this notation, expanding $l_{\vartheta k x_0}-1$ three times around ψ^* while fixing π gives, with $\overline{\psi}\in[\psi,\psi^*]$,

$$\begin{aligned}l_{\vartheta k x_0}-1&=(\Delta^1\psi)'\Lambda^1_{k,0,x_0}(\psi^*,\pi)+(\Delta^2\psi)'\Lambda^2_{k,0,x_0}(\psi^*,\pi)+(\Delta^3\psi)'\Lambda^3_{k,0,x_0}(\overline{\psi},\pi)\\&=(\Delta^1\psi)'\Lambda^1_{k,0}(\psi^*,\pi)+(\Delta^2\psi)'\Lambda^2_{k,0}(\psi^*,\pi)+(\Delta^3\psi)'\Lambda^3_{k,0}(\overline{\psi},\pi)+u_{kx_0}(\psi,\pi),\end{aligned}\tag{72}$$

where $\overline{\psi}$ may differ from element to element of the vector $\Lambda^3_{k,0,x_0}(\overline{\psi},\pi)$, and $u_{kx_0}(\psi,\pi):=\sum_{j=1}^2(\Delta^j\psi)'[\Lambda^j_{k,0,x_0}(\psi^*,\pi)-\Lambda^j_{k,0}(\psi^*,\pi)]+(\Delta^3\psi)'[\Lambda^3_{k,0,x_0}(\overline{\psi},\pi)-\Lambda^3_{k,0}(\overline{\psi},\pi)]$.

Define, for $m\geq 0$,

$$s_{k,m}(\pi):=\begin{pmatrix}\dfrac{\nabla_\eta\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_{-m}^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_{-m}^{k-1})}\\[2ex]\dfrac{\widetilde{\nabla}_{v(\lambda)}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_{-m}^{k-1})}{\alpha(1-\alpha)\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_{-m}^{k-1})}\end{pmatrix},$$

where $\widetilde{\nabla}_{v(\lambda)}:=(\nabla_{\lambda_1\lambda_1}/2,\dots,\nabla_{\lambda_q\lambda_q}/2,\nabla_{\lambda_1\lambda_2},\dots,\nabla_{\lambda_{q-1}\lambda_q})'$. Noting that $\nabla_\lambda\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})=0$

and $\nabla_{\lambda\eta'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$ from (15), we may rewrite (72) as

$$l_{k\vartheta x_0} - 1 = t(\psi,\pi)' s_{k,0}(\pi) + r_{k,0}(\psi,\pi) + u_{kx_0}(\psi,\pi), \tag{73}$$

with $r_{k,0}(\psi,\pi) := \tau(\psi)'\widetilde{\Lambda}_{k,0}(\pi) + (\Delta^3\psi)'\Lambda_{k,0}^3(\overline{\psi},\pi)$, where $\tau(\psi)$ is the vector that collects the elements of $\Delta^2\psi$ of the form $\eta_i\eta_j$, and $\widetilde{\Lambda}_{k,0}(\pi)$ denotes the vector of the corresponding elements of $\Lambda_{k,0}^2(\psi^*,\pi)$.

For $m \geq 0$, define $v_{k,m}(\vartheta) := (s_{k,m}(\pi)', \widetilde{\Lambda}_{k,m}(\pi)', \Lambda_{k,m}^3(\psi,\pi)')'$, and define $v_{k,\infty}(\vartheta) := \lim_{m\to\infty} v_{k,m}(\vartheta)$. Choose $\epsilon > 0$ sufficiently small that $\mathcal{N}_\epsilon$ is a subset of $\mathcal{N}^*$ in Assumption 4. Then, $v_{k,\infty}(\vartheta)$ is a $\overline{\mathbb{P}}_{\vartheta^*}$-stationary $L^2(\overline{\mathbb{P}}_{\vartheta^*})$ martingale difference sequence for all $\vartheta \in \mathcal{N}_\epsilon$ from Proposition 6. In order to apply Corollary 1 to $l_{\vartheta k x_0} - 1$, we first show

$$\sup_{\vartheta\in\mathcal{N}_\epsilon} \left| P_n[v_{k,0}(\vartheta)v_{k,0}(\vartheta)'] - \mathbb{E}_{\vartheta^*}[v_{k,\infty}(\vartheta)v_{k,\infty}(\vartheta)'] \right| = o_p(1), \tag{74}$$

$$\nu_n(v_{k,0}(\vartheta)) \Rightarrow W(\vartheta), \tag{75}$$

where $W(\vartheta)$ is a mean-zero continuous Gaussian process with $\mathbb{E}_{\vartheta^*}[W(\vartheta_1)W(\vartheta_2)'] = \mathbb{E}_{\vartheta^*}[v_{k,\infty}(\vartheta_1)v_{k,\infty}(\vartheta_2)']$. (74) holds because $P_n[v_{k,0}(\vartheta)v_{k,0}(\vartheta)' - v_{k,\infty}(\vartheta)v_{k,\infty}(\vartheta)'] = o_p(1)$ from Proposition 6, and $v_{k,\infty}(\vartheta)v_{k,\infty}(\vartheta)'$ satisfies a uniform law of large numbers (Lemma 2.4 and footnote 18 of Newey and McFadden (1994)) because $v_{k,\infty}(\vartheta)$ is continuous and $\mathbb{E}_{\vartheta^*}\sup_{\vartheta\in\mathcal{N}_\epsilon}|v_{k,\infty}(\vartheta)v_{k,\infty}(\vartheta)'| < \infty$ from Proposition 6. (75) holds because $\nu_n(v_{k,0}(\vartheta) - v_{k,\infty}(\vartheta)) = o_p(1)$ from Proposition 6 and $\nu_n(v_{k,\infty}(\vartheta)) \Rightarrow W(\vartheta)$ from Theorem 10.2 of Pollard (1990) because (i) the space of ϑ is totally bounded, (ii) the finite dimensional distributions of $\nu_n(v_{k,\infty}(\cdot))$ converge to those of $W(\cdot)$ from a multivariate martingale CLT, and (iii) $\{\nu_n(v_{k,\infty}(\cdot)) : n \geq 1\}$ is stochastically equicontinuous from Theorem 2 of Hansen (1996) because $v_{k,\infty}(\vartheta)$ is Lipschitz continuous in ϑ and both $v_{k,\infty}(\vartheta)$ and the Lipschitz coefficient are in $L^q(\mathbb{P}_{\vartheta^*})$ with $q > \dim(\vartheta)$ from Proposition 6.

We proceed to show that the terms on the right hand side of (73) satisfies Assumption 3(a)–(g). Observe that $t(\psi,\pi) = 0$ if and only if $\psi = \psi^*$. First, $s_{k,0}(\pi)$ satisfies Assumption 3(a)(b)(f)(g) by Proposition 6, (74), (75), and Assumption 6. Second, $r_{k,0}(\psi,\pi)$ satisfies assumptions (c)(d) from Proposition 6 and (75). Third, $u_{kx_0}(\psi,\pi)$ satisfies assumption (e) from Proposition 6(c). Therefore, the stated result follows from Corollary 1(b). □

Proof of Proposition 9. The proof is similar to that of Proposition 3 of Kasahara and Shimotsu (2015). Let $t_\eta := \eta - \eta^*$ and $t_\lambda := \alpha(1-\alpha)v(\lambda)$, so that $t(\psi,\pi) = (t_\eta', t_\lambda')'$. Let $\hat{\psi}_\pi := \arg\max_{\psi\in\Theta_\psi} \ell_n(\psi,\pi,\xi)$ denote the MLE of ψ, and split $t(\hat{\psi}_\pi,\pi)$ as $t(\hat{\psi}_\pi,\pi) = (\hat{t}_\eta', \hat{t}_\lambda')'$, where we suppress the dependence of $\hat{t}_\eta$ and $\hat{t}_\lambda$ on π. Define $G_{\varrho n} := \nu_n(s_{\varrho k})$. Let

$$G_{\varrho n} = \begin{bmatrix} G_{\eta n} \\ G_{\lambda\varrho n} \end{bmatrix}, \quad \begin{array}{c} G_{\lambda.\eta\varrho n} := G_{\lambda\varrho n} - \mathcal{I}_{\lambda\eta\varrho}\mathcal{I}_\eta^{-1}G_{\eta n}, \quad Z_{\lambda.\eta\varrho n} := \mathcal{I}_{\lambda.\eta\varrho}^{-1}G_{\lambda.\eta\varrho n}, \\ t_{\eta.\lambda\varrho} := t_\eta + \mathcal{I}_\eta^{-1}\mathcal{I}_{\eta\lambda\varrho}t_\lambda. \end{array}$$

Then, we can write (21) as

$$\sup_{\xi\in\Xi}\sup_{\vartheta\in A_{nc}(\xi)}\left|2\left[\ell_n(\psi,\pi,\xi)-\ell_n(\psi^*,\pi,\xi)\right]-A_n(\sqrt{n}t_{\eta.\lambda\varrho})-B_{\varrho n}(\sqrt{n}t_\lambda)\right|=o_p(1), \tag{76}$$

where

$$\begin{aligned}A_n(t_{\eta.\lambda\varrho})&=2t_{\eta.\lambda\varrho}'G_{\eta n}-t_{\eta.\lambda\varrho}'\mathcal{I}_\eta t_{\eta.\lambda\varrho},\\ B_{\varrho n}(t_\lambda)&=2t_\lambda' G_{\lambda.\eta\varrho n}-t_\lambda'\mathcal{I}_{\lambda.\eta\varrho}t_\lambda=Z_{\lambda\varrho n}'\mathcal{I}_{\lambda.\eta\varrho}Z_{\lambda\varrho n}-(t_\lambda-Z_{\lambda\varrho n})'\mathcal{I}_{\lambda.\eta\varrho}(t_\lambda-Z_{\lambda\varrho n}).\end{aligned} \tag{77}$$

Observe that $2[\ell_{0n}(\hat{\vartheta}_0)-\ell_{0n}(\vartheta_0^*)]=\max_{t_\eta}[2\sqrt{n}t_\eta' G_{\eta n}-nt_\eta'\mathcal{I}_\eta t_\eta]+o_p(1)=\max_{t_{\eta.\lambda\varrho}}A_n(\sqrt{n}t_{\eta.\lambda\varrho})+o_p(1)$ from applying Corollary 1 to $\ell_{0n}(\vartheta_0)$ and noting that the set of **possible** values of both $\sqrt{n}t_\eta$ and $\sqrt{n}t_{\eta.\lambda\varrho}$ approaches $\mathbb{R}^{\dim(\eta)}$. In conjunction with (76), we obtain, uniformly in $\pi\in\Theta_{\pi\epsilon}$,

$$2[\ell_n(\hat{\psi}_\pi,\pi,\xi)-\ell_{0n}(\hat{\vartheta}_0)]=B_{\varrho n}(\sqrt{n}\hat{t}_\lambda)+o_p(1). \tag{78}$$

Define $\tilde{t}_\lambda$ by $B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)=\max_{t_\lambda\in\alpha(1-\alpha)v(\Theta_\lambda)}B_{\varrho n}(\sqrt{n}t_\lambda)$. Then, we have

$$2[\ell_n(\hat{\psi}_\pi,\pi,\xi)-\ell_{0n}(\hat{\vartheta}_0)]=B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)+o_p(1)$$

uniformly in $\pi\in\Theta_{\pi\epsilon}$ because (i) $B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)\geq 2[\ell_n(\hat{\psi}_\pi,\pi,\xi)-\ell_{0n}(\hat{\vartheta}_0)]+o_p(1)$ from the definition of $\tilde{t}_\lambda$ and (78), and (ii) $2[\ell_n(\hat{\psi}_\pi,\pi,\xi)-\ell_{0n}(\hat{\vartheta}_0)]\geq B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)+o_p(1)$ from the definition of $\hat{\psi}$, (76), and $\tilde{t}_\lambda=O_p(n^{-1/2})$.

Finally, the asymptotic distribution of $\sup_\varrho B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)$ follows from applying Theorem 1(c) of Andrews (2001) to $B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)$. First, Assumption 2 of Andrews (2001) holds trivially for $B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)$. Second, Assumption 3 of Andrews (2001) is satisfied by (75) and Assumption 6. Assumption 4 of Andrews (2001) is satisfied by Proposition 8. Assumption 5* of Andrews (2001) holds with $B_T=n^{1/2}$ because $\alpha(1-\alpha)v(\Theta_\lambda)$ is locally equal to the cone $v(\mathbb{R}^q)$ given that $\alpha(1-\alpha)>0$ for all $\alpha\in\Theta_{\alpha\epsilon}$. Therefore, $\sup_{\varrho\in\Theta_{\varrho\epsilon}}B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda)\Rightarrow\sup_{\varrho\in\Theta_{\varrho\epsilon}}(\hat{t}_{\lambda\varrho}'\mathcal{I}_{\lambda.\eta\varrho}\hat{t}_{\lambda\varrho})$ follows from Theorem 1(c) of Andrews (2001). □

Proof of Proposition 10. The proof is similar to that of Proposition 8. Define $\Lambda^j_{k,m,x_{-m}}(\psi,\pi)$ and $\Lambda^j_{k,m}(\psi,\pi)$ as in the proof of Proposition 8. Expanding $l_{k\vartheta x_0}-1$ five times around ψ^* similarly to (78) while fixing π gives, with $\overline{\psi}\in[\psi,\psi^*]$,

$$l_{k\vartheta x_0}-1=\sum_{j=1}^4(\Delta^j\psi)'\Lambda^j_{k,0}(\psi^*,\pi)+(\Delta^5\psi)'\Lambda^5_{k,0}(\overline{\psi},\pi)+u_{kx_0}(\psi,\pi), \tag{79}$$

where $u_{kx_0}(\psi,\pi):=\sum_{j=1}^4(\Delta^j\psi)'[\Lambda^j_{k,0,x_0}(\psi^*,\pi)-\Lambda^j_{k,0}(\psi^*,\pi)]+(\Delta^5\psi)'[\Lambda^5_{k,0,x_0}(\overline{\psi},\pi)-\Lambda^5_{k,0}(\overline{\psi},\pi)]$.

Define $\overline{p}_{\psi\pi k,m} := \overline{p}_{\psi\pi}(Y_k|\overline{Y}_{-m}^{k-1})$, and define, for $m \geq 0$,

$$s_{k,m}(\pi) := \begin{pmatrix} \nabla_{\eta}\overline{p}_{\psi^*\pi k,m}/\overline{p}_{\psi^*\pi k,m} \\ \zeta_{k,m}(\varrho)/2 \\ \nabla_{\lambda_\mu\lambda_\sigma}\overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \\ \nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi k,m}/2\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \\ \nabla_{\lambda_\beta\lambda_\mu}\overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \\ \nabla_{\lambda_\beta\lambda_\sigma}\overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \\ \widetilde{\nabla}_{v(\lambda_\beta)}\overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \end{pmatrix},$$

where $\widetilde{\nabla}_{v(\lambda_\beta)} := (\nabla_{\lambda_{\beta 1}\lambda_{\beta 1}}/2, \ldots, \nabla_{\lambda_{\beta q}\lambda_{\beta q}}/2, \nabla_{\lambda_{\beta 1}\lambda_{\beta 2}}, \ldots, \nabla_{\lambda_{\beta,q-1}\lambda_{\beta q}})'$. Noting that $\nabla_{\lambda}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$ and $\nabla_{\lambda\eta'}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$ from (15) and (16), we may rewrite (79) as, with $t(\psi,\pi)$ defined in (32),

$$l_{\vartheta k x_0} - 1 = t(\psi,\pi)' s_{k,0}(\pi) + r_{k,0}(\pi) + u_{kx_0}(\psi,\pi), \tag{80}$$

where $r_{k,0}(\pi) := \tau(\psi)'\widetilde{\Lambda}_{k,0}(\pi) + (\Delta^5\psi)'\Lambda_{k,0}^5(\overline{\psi},\pi) + \lambda_\mu^4[\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi k,0} - b(\alpha)\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi k,0}]/4!\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0}$, $\tau(\psi)$ is the vector that collects the elements of $\{\Delta^j\psi\}_{j=2}^4$ that are not in $t(\psi,\pi)$, and $\widetilde{\Lambda}_{k,0}(\pi)$ denotes the vector of the elements of $\{\Lambda_{k,0}^j(\psi^*,\pi)\}_{j=2}^4$ that correspond to $\tau(\psi)$.

The stated result follows from Corollary 1 if the terms on the right hand side of (80) satisfy Assumption 3. Similarly to the proof of Proposition 9, define $v_{k,m}(\vartheta) := (s_{k,m}(\pi)', \widetilde{\Lambda}_{k,0}(\pi)', \Lambda_{k,0}^5(\psi,\pi)')'$. Because $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) = 0$ from (30), we can rewrite (31) using the mean value theorem as, for $\bar{\varrho} \in [0,\varrho]$,

$$\zeta_{k,0}(\varrho) = \frac{\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\alpha\varrho}(Y_k|\overline{Y}_0^{k-1}) - \nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})}{\varrho\alpha(1-\alpha)\overline{p}_{\psi^*\alpha\varrho}(Y_k|\overline{Y}_0^{k-1})} = \frac{\nabla_{\varrho}\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\alpha\bar{\varrho}}(Y_k|\overline{Y}_0^{k-1})}{\alpha(1-\alpha)\overline{p}_{\psi^*\alpha\bar{\varrho}}(Y_k|\overline{Y}_0^{k-1})}.$$

The right hand side is well-defined when $\varrho \to 0$ and satisfies Proposition 6. Therefore, $v_{k,\infty}(\vartheta) := \lim_{m\to\infty} v_{k,m}(\vartheta)$ is well-defined, and $v_{k,0}(\vartheta)$ and $v_{k,\infty}(\vartheta)$ satisfy (74)-(75) from repeating the argument in the proof of Proposition 9.

We proceed to show that the terms on the right hand side of (80) satisfy Assumption 3. Observe that $t(\psi,\pi) = 0$ if and only if $\psi = \psi^*$. $s_{k,0}(\pi)$ and $u_{kx_0}(\psi,\pi)$ satisfy Assumption 3 from the argument in the proof of Proposition 8 with replacing Assumption 6 with Assumption 7. We show that each component of $r_{k,0}(\pi)$ satisfies Assumption 3(c)(d). First, for $\tau(\psi)'\tilde{\Delta}_{k,0}(\pi)$, observe that $\nabla_{\lambda\eta^j}\overline{p}_{\psi^*\pi k,0} = 0$ for any $j \geq 1$ in view of (26)–(29). Therefore, $\tau(\psi)'\tilde{\Delta}_{k,0}(\pi)$ is written as, with $\dot{\eta} := \eta - \eta^*$,

$$\dot{\eta}'(\nabla_{\eta\eta'}\overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0})\dot{\eta} + R_{3k\vartheta} + R_{4k\vartheta}, \tag{81}$$

where, for $j = 3, 4$,

$$R_{jk\vartheta} := \frac{1}{\overline{p}_{\psi^*\pi k,0}} \sum_{(u_1,\ldots,u_d)\in D(j)} \frac{1}{u_1!\cdots u_d!} \frac{\partial^j \overline{p}_{\psi^*\pi k,0}}{\partial\psi_1^{u_1}\cdots\partial\psi_d^{u_d}} (\psi_1-\psi_1^*)^{u_1}\cdots(\psi_d-\psi_d^*)^{u_d}, \tag{82}$$

where $D(3)$ and $D(4)$ are defined as

$$D(3) := \{(u_1,\ldots,u_d) : u_1+\cdots+u_d = 3\},$$
$$D(4) := \{(u_1,\ldots,u_d) : u_1+\cdots+u_d = 4,\ (\psi_1-\psi_1^*)^{u_1}\cdots(\psi_d-\psi_d^*)^{u_d} \neq \lambda_\mu^4\}.$$

The first term in (81) clearly satisfies Assumption 3(c)(d). We proceed to show that $R_{3k\vartheta}$ and $R_{4k\vartheta}$ satisfy Assumption 3(c)(d). The terms in $R_{3k\vartheta}$ belongs to one of the following three groups; (i) the term associated with λ_σ^3; (ii) the term associated with λ_μ^3; (iii) the other terms. Then, Assumption 3(c)(d) is satisfied by the term (i) because $\lambda_\sigma^3 = \lambda_\sigma[\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12] - (\lambda_\mu^3 b(\alpha))\lambda_\mu\lambda_\sigma/12 = O(|\psi||t(\psi,\pi)|)$; the term (ii) from Lemma 6(a) because $t(\psi,\pi)$ includes $\varrho\lambda_\mu^2$; the terms in (iii) because they either contain $\dot\eta$ or a term of the form $\lambda_\mu^i\lambda_\sigma^j\lambda_\beta^k$ with $i+j+k=3$ and $i,j\neq 3$. Similarly, the terms in $R_{4k\vartheta}$ satisfy Assumption3(c)(d) because they either contain $\dot\eta$ or a term of the form $\lambda_\mu^i\lambda_\sigma^j\lambda_\beta^k$ with $i+j+k=4$ and $i \neq 4$. Second, $(\Delta^5\psi)'\overline{\Delta}_{k,0}^5(\overline{\psi},\pi)$ satisfies Assumption 3(c)(d) from Proposition 6, (75) and $\lambda_\mu^5 = (12\lambda_\mu/b(\alpha))[\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12] - 12(\lambda_\sigma/b(\alpha))\lambda_\mu\lambda_\sigma = O(|\psi||t(\psi,\pi)|)$. Third, $\lambda_\mu^4[\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi k,0} - b(\alpha)\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi k,0}]/4!\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0}$ satisfies Assumption 3(c)(d) from Lemma 6(b). This proves that $r_{k,0}(\pi)$ satisfies Assumption 3(c)(d), and the stated result is proven. □

Proof of Proposition 11. The proof is similar to the proof of Proposition 3(c) of Kasahara and Shimotsu (2015). Let $(\hat\psi_\alpha, \hat\varrho_\alpha) := \arg\max_{(\psi,\varrho)\in\Theta_\psi\times\Theta_\varrho} \ell_n(\psi,\alpha,\varrho,\xi)$ denote the MLE of (ψ,ϱ) for a given α. Consider the sets $\Theta_\lambda^1 := \{\lambda\in\Theta_\lambda : |\lambda_\mu| \geq n^{-1/8}(\log n)^{-1}\}$ and $\Theta_\lambda^2 := \{\lambda\in\Theta_\lambda : |\lambda_\mu| < n^{-1/8}(\log n)^{-1}\}$, so that $\Theta_\lambda = \Theta_\lambda^1 \cup \Theta_\lambda^2$. For $j=1,2$, define $(\hat\psi_\alpha^j, \hat\varrho_\alpha^j) := \arg\max_{(\psi,\varrho)\in\Theta_\psi\times\Theta_\varrho,\lambda\in\Theta_\lambda^j} \ell_n(\psi,\alpha,\varrho,\xi)$. Then, uniformly in α,

$$\ell_n(\hat\psi_\alpha,\alpha,\hat\varrho_\alpha,\xi) = \max\left\{\ell_n(\hat\psi_\alpha^1,\alpha,\hat\varrho_\alpha^1,\xi), \ell_n(\hat\psi_\alpha^2,\alpha,\hat\varrho_\alpha^2,\xi)\right\}.$$

Henceforth, we suppress the dependence of $\hat\psi_\alpha$, $\hat\varrho_\alpha$, etc. on α.

Define $B_{\varrho n}(t_\lambda(\lambda,\varrho,\alpha))$ as in (77) in the proof of Proposition 9 but using $t(\psi,\pi)$ and $s_{\varrho k}$ defined in (32) and (35) and replacing t_λ in (77) with $t_\lambda(\lambda,\varrho,\alpha)$. Observe that the proof of Proposition 9 goes through up to (78) with the current notation and that $G_{\varrho n}$ and $\mathcal{I}_\varrho$ are continuous in ϱ. Further, $\hat\varrho^1 = O_p(n^{-1/4}(\log n)^2)$ because $\hat\varrho^1(\hat\lambda_\mu^1)^2 = O_p(n^{-1/2})$ from Proposition 10(a) and $|\hat\lambda_\mu^1| \geq n^{-1/8}(\log n)^{-1}$. Consequently, $B_{\hat\varrho^1 n}(\sqrt{n}t_\lambda(\hat\lambda^1,\hat\varrho^1,\alpha)) = B_{0n}(\sqrt{n}t_\lambda(\hat\lambda^1,\hat\varrho^1,\alpha)) + o_p(1)$, and, uniformly in α,

$$2[\ell_n(\hat\psi,\alpha,\hat\varrho,\xi) - \ell_{0n}(\hat\vartheta_0)] = \max\{B_{0n}(\sqrt{n}t_\lambda(\hat\lambda^1,\hat\varrho^1,\alpha)), B_{\hat\varrho^2 n}(\sqrt{n}t_\lambda(\hat\lambda^2,\hat\varrho^2,\alpha))\} + o_p(1). \tag{83}$$

We proceed to construct parameter spaces $\tilde\Lambda_{\lambda\alpha}^1$ and $\tilde\Lambda_{\lambda\alpha\varrho}^2$ that are locally equal to the cones Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$ defined in (37). Define $c(\alpha) := \alpha(1-\alpha)$, and denote the elements of $t_\lambda(\hat\lambda^j,\hat\varrho^j,\alpha)$

corresponding to (33) by

$$t_\lambda(\hat\lambda^j,\hat\varrho^j,\alpha)=\begin{pmatrix}\hat t^j_{\varrho\mu^2}\\ \hat t^j_{\mu\sigma}\\ \hat t^j_{\sigma^2}\\ \hat t^j_{\beta\mu}\\ \hat t^j_{\beta\sigma}\\ \hat t^j_{v(\beta)}\end{pmatrix}:=\begin{pmatrix}c(\alpha)\hat\varrho^j(\hat\lambda^j_\mu)^2\\ c(\alpha)\hat\lambda_\mu\hat\lambda_\sigma\\ c(\alpha)[(\hat\lambda^j_\sigma)^2+b(\alpha)(\hat\lambda^j_\mu)^4/12]\\ c(\alpha)\hat\lambda^j_\beta\hat\lambda^j_\mu\\ c(\alpha)\hat\lambda^j_\beta\hat\lambda^j_\sigma\\ c(\alpha)v(\hat\lambda^j_\beta)\end{pmatrix}.$$

Note that $t_\lambda(\hat\lambda^1,\hat\varrho^1,\alpha)$ satisfies $\hat\lambda^1_\sigma=O_p(n^{-3/8}\log n)$ and $\hat\lambda^1_\beta=O_p(n^{-3/8}\log n)$ because $(\hat t^1_{\mu\sigma},\hat t^1_{\beta\mu})=O_p(n^{-1/2})$ from Proposition 10(a) and $|\hat\lambda^1_\mu|\geq n^{-1/8}(\log n)^{-1}$. Furthermore, $t_\lambda(\hat\lambda^2,\hat\varrho^2,\alpha)$ satisfies $\hat t^2_{\sigma^2}=c(\alpha)(\hat\lambda^2_\sigma)^2+o_p(n^{-1/2})$ because $|\hat\lambda^2_\mu|<n^{-1/8}(\log n)^{-1}$. Consequently,

$$\begin{aligned}&\hat t^1_{\beta\sigma}=o_p(n^{-1/2}),\quad \hat t^1_{v(\beta)}=o_p(n^{-1/2}),\quad \hat t^1_{\sigma^2}=c(\alpha)b(\alpha)(\hat\lambda^1_\mu)^4/12+o_p(n^{-1/2}),\\ &\hat t^2_{\sigma^2}=c(\alpha)(\hat\lambda^2_\sigma)^2+o_p(n^{-1/2}).\end{aligned}\tag{84}$$

In view of this, let $t_\lambda(\lambda,\varrho,\alpha):=(t_{\varrho\mu^2},t_{\mu\sigma},t_{\sigma^2},t'_{\beta\mu},t'_{\beta\sigma},t'_{v(\beta)})'\in\mathbb{R}^{q_\lambda}$, and consider the following sets:

$$\begin{aligned}\tilde\Lambda^1_{\lambda\alpha}:=\{&t_\lambda(\lambda,\varrho,\alpha):t_{\varrho\mu^2}=c(\alpha)\varrho\lambda^2_\mu,t_{\mu\sigma}=c(\alpha)\lambda_\mu\lambda_\sigma,t_{\sigma^2}=c(\alpha)b(\alpha)\lambda^4_\mu/12,\\ &t_{\beta\mu}=c(\alpha)\lambda_\beta\lambda_\mu,t_{\beta\sigma}=0,t_{v(\beta)}=0\text{ for some }(\lambda,\varrho)\in\Theta_\lambda\times\Theta_\varrho\},\\ \tilde\Lambda^2_{\lambda\alpha\varrho}:=\{&t_\lambda(\lambda,\varrho,\alpha):t_{\varrho\mu^2}=c(\alpha)\varrho\lambda^2_\mu,t_{\mu\sigma}=c(\alpha)\lambda_\mu\lambda_\sigma,t_{\sigma^2}=c(\alpha)\lambda^2_\sigma,\\ &t_{\beta\mu}=c(\alpha)\lambda_\beta\lambda_\mu,t_{\beta\sigma}=c(\alpha)\lambda_\beta\lambda_\sigma,t_{v(\beta)}=c(\alpha)v(\lambda_\beta)\text{ for some }\lambda\in\Theta_\lambda\}.\end{aligned}$$

$\tilde\Lambda^1_{\lambda\alpha}$ is indexed by α but does not depend on ϱ because $B_{0n}(\cdot)$ in (83) does not depend on ϱ, whereas $\tilde\Lambda^2_{\lambda\alpha\varrho}$ is indexed by both α and ϱ because $B_{\hat\varrho^2 n}(\cdot)$ in (83) depends on $\hat\varrho^2$. Define $(\tilde\lambda^1_\alpha,\tilde\varrho^1_\alpha)$ and $\tilde\lambda^2_{\alpha\varrho}$ by $B_{0n}(\sqrt{n}t_\lambda(\tilde\lambda^1_\alpha,\tilde\varrho^1_\alpha,\alpha))=\max_{t_\lambda(\lambda,\varrho,\alpha)\in\tilde\Lambda^1_{\lambda\alpha}}B_{0n}(\sqrt{n}t_\lambda(\lambda,\varrho,\alpha))$ and $B_{\varrho n}(\sqrt{n}t_\lambda(\tilde\lambda^2_{\alpha\varrho},\varrho,\alpha))=\max_{t_\lambda(\lambda,\varrho,\alpha)\in\tilde\Lambda^2_{\lambda\alpha\varrho}}B_{\varrho n}(\sqrt{n}t_\lambda(\lambda,\varrho,\alpha))$.

Define $W_n(\alpha):=\max\{B_{0n}(\sqrt{n}t_\lambda(\tilde\lambda^1_\alpha,\tilde\varrho^1_\alpha,\alpha)),\sup_{\varrho\in\Theta_{\varrho\epsilon}}B_{\varrho n}(\sqrt{n}t_\lambda(\tilde\lambda^2_{\alpha\varrho},\varrho,\alpha))\}$, then we have

$$2[\ell_n(\hat\psi,\alpha,\hat\varrho,\xi)-\ell_{0n}(\hat\vartheta_0)]=W_n(\alpha)+o_p(1),\tag{85}$$

uniformly in $\alpha\in\Theta_\alpha$ because (i) $W_n(\alpha)\geq 2[\ell_n(\hat\psi,\alpha,\hat\varrho,\xi)-\ell_{0n}(\hat\vartheta_0)]+o_p(1)$ in view of the definition of $(\tilde\lambda^1_\alpha,\tilde\varrho^1_\alpha,\tilde\lambda^2_{\alpha\varrho})$, (83), and (84), and (ii) with $\tilde\eta^1_\alpha:=\arg\max_\eta\ell_n(\eta,\tilde\lambda^1_\alpha,\alpha,\tilde\varrho^1_\alpha,\xi)$ and $\tilde\eta^2_{\alpha\varrho}:=\arg\max_\eta\ell_n(\eta,\tilde\lambda^2_{\alpha\varrho},\alpha,\varrho,\xi)$, we have $2[\ell_n(\hat\psi,\alpha,\hat\varrho,\xi)-\ell_{0n}(\hat\vartheta_0)]\geq$ $\max\{2[\ell_n(\tilde\eta^1_\alpha,\tilde\lambda^1_\alpha,\alpha,\tilde\varrho^1_\alpha,\xi),\sup_{\varrho\in\Theta_{\varrho\epsilon}}\ell_n(\tilde\eta^2_{\alpha\varrho},\tilde\lambda^2_{\alpha\varrho},\alpha,\varrho,\xi)\}-2\ell_{0n}(\hat\vartheta_0)+o_p(1)=W_n(\alpha)+o_p(1)$ from the definition of $(\hat\psi,\hat\varrho)$.

The asymptotic distribution of the LRTS follows from applying Theorem 1(c) of Andrews (2001) to $(B_{0n}(\sqrt{n}t_\lambda(\tilde\lambda^1_\alpha,\tilde\varrho^1_\alpha,\alpha)),B_{\varrho n}(\sqrt{n}t_\lambda(\tilde\lambda^2_{\alpha\varrho},\varrho,\alpha)))$. First, Assumption 2 of Andrews (2001) holds trivially for $B_{\varrho n}(\sqrt{n}t(\lambda,\varrho,\alpha))$. Second, Assumption 3 of Andrews (2001) is satisfied by (75) and Assumption 7. Assumption 4 of Andrews (2001) is satisfied by Proposition 10. Assumption 5* of Andrews (2001) holds with $B_T = n^{1/2}$ because $\tilde{\Lambda}^1_{\lambda\alpha}$ is locally (in a neighborhood of $\varrho = 0, \lambda = 0$) equal to the cone Λ^1_λ and $\tilde{\Lambda}^2_{\lambda\alpha\varrho}$ is locally equal to the cone $\Lambda^2_{\lambda\varrho}$ uniformly in $\varrho \in \Theta_{\varrho\epsilon}$. Consequently, $W_n(\alpha) \Rightarrow \sup_{\varrho\in\Theta_{\varrho\epsilon}} \max\{\mathbb{I}\{\varrho = 0\}(\hat{t}^1_\lambda)'\mathcal{I}_{\lambda.\eta 0}\hat{t}^1_\lambda, (\hat{t}^2_{\lambda\varrho})'\mathcal{I}_{\lambda.\eta\varrho}\hat{t}^2_{\lambda\varrho}\}$ uniformly in α from Theorem 1(c) of Andrews (2001), and the stated result follows from (85). □

Proof of Proposition 12. The proof is similar to that of Proposition 10. Expanding $l_{k\vartheta x_0} - 1$ five times around ψ^* and proceeding as in the proof of Proposition 10 gives

$$l_{\vartheta k x_0} - 1 = t(\psi, \pi)' s_{k,0}(\pi) + r_{k,0}(\pi) + u_{k x_0}(\psi, \pi), \tag{86}$$

where $t(\psi, \pi)$ is defined in (44),

$$s_{k,m}(\pi) := \begin{pmatrix} \nabla_\eta \overline{p}_{\psi^*\pi k,m}/\overline{p}_{\psi^*\pi k,m} \\ \zeta_{k,m}(\varrho)/2 \\ \nabla_{\mu^3} f^*_k/3!\alpha(1-\alpha) f^*_k \\ \nabla_{\mu^4} f^*_k/4!\alpha(1-\alpha) f^*_k \\ \nabla_{\lambda_\beta\lambda_\mu} \overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \\ \widetilde{\nabla}_{v(\lambda_\beta)} \overline{p}_{\psi^*\pi k,m}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,m} \end{pmatrix},$$

$r_{k,0}(\pi) := \tau(\psi)'\widetilde{\Lambda}_{k,0}(\pi) + (\Delta^5\psi)'\Lambda^5_{k,0}(\overline{\psi}, \pi)$, and $\{\tau(\psi), \widetilde{\Lambda}_{k,0}(\pi), \Delta^5\psi, \Lambda^5_{k,0}(\overline{\psi}, \pi), \overline{p}_{\psi\pi k,m}, \widetilde{\nabla}_{v(\lambda_\beta)}, u_{kx_0}(\psi, \pi)\}$ are defined similarly to those in the proof of Proposition 10.

The stated result is proven if the terms on the right hand side of (86) satisfy Assumption 3. $t(\psi, \pi) = 0$ if and only if $\psi = \psi^*$. $s_{k,0}(\pi)$ and $u_{kx_0}(\psi, \pi)$ satisfy Assumption 3 by the same argument as the proof of Proposition 10. For $r_{k,0}(\pi)$, first, $(\Delta^5\psi)'\overline{\Delta}^5_{k,0}(\overline{\psi}, \pi)$ satisfies Assumption 3(c)(d) from a similar argument to the proof of Proposition 10. Second, similar to the proof of Proposition 10, write $\tau(\psi)'\tilde{\Delta}_{k,0}(\pi)$ as $\dot{\eta}'(\nabla_{\eta\eta'}\overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0})\dot{\eta} + \tilde{R}_{3k\vartheta} + \tilde{R}_{4k\vartheta}$, where $\dot{\eta} := \eta - \eta^*$ and $\tilde{R}_{jk\vartheta}$ is defined as $R_{jk\vartheta}$ in (82) with $D(j)$ replaced with $\tilde{D}(j) := \{(u_1, \ldots, u_d) : u_1 + \cdots + u_d = j, (\psi_1 - \psi_1^*)^{u_1} \cdots (\psi_d - \psi_d^*)^{u_d} \neq \lambda^j_\mu\}$ for $j = 3, 4$. The term $\dot{\eta}'(\nabla_{\eta\eta'}\overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0})\dot{\eta}$ clearly satisfies Assumption 3(c)(d). The terms in $\tilde{R}_{3k\vartheta}$ satisfy Assumption 3(c)(d) because they contain either $\dot{\eta}$ or $\lambda^2_\mu\lambda_\beta$ or $\lambda_\mu\lambda^2_\beta$ or λ^3_β. The terms in $\tilde{R}_{4k\vartheta}$ satisfy assumptions (c)(d) because they either contain $\dot{\eta}$ or a term of the form $\lambda^i_\mu\lambda^{4-i}_\beta$ with $1 \leq i \leq 3$. Therefore, $r_{k,0}(\pi)$ satisfies Assumption 3(c)(d), and the stated result is proven. □

Proof of Proposition 13. The proof is similar to the proof of Proposition 11. Let $(\hat{\psi}, \hat{\pi}) := \arg\max_{(\psi,\pi)\in\Theta_\psi\times\Theta_{\pi\epsilon}} \ell_n($ denote the MLE of (ψ, π). Consider the sets $\Theta^1_\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\mu| \geq n^{-1/6}(\log n)^{-1}\}$ and $\Theta^2_\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\mu| < n^{-1/6}(\log n)^{-1}\}$, so that $\Theta_\lambda = \Theta^1_\lambda \cup \Theta^2_\lambda$. For $j = 1, 2$, define $(\hat{\psi}^j, \hat{\pi}^j) := \arg\max_{(\psi,\pi)\in\Theta_\psi\times\Theta_{\pi\epsilon},\lambda\in\Theta^j_\lambda} \ell_n(\psi, \pi, \xi)$, so that $\ell_n(\hat{\psi}, \hat{\pi}, \xi) = \max_{j\in\{1,2\}} \ell_n(\hat{\psi}^j, \hat{\pi}^j, \xi)$.

Define $B_{\varrho n}(t_\lambda(\lambda, \pi))$ as in (77) in the proof of Proposition 9 but using $t(\psi, \pi)$ and $s_{\varrho k}$ defined in (44) and (45) and replacing t_λ in (77) with $t_\lambda(\lambda, \pi)$. Observe that $\hat{\varrho}^1 = O_p(n^{-1/6}(\log n)^2)$ because

$\hat{\varrho}^1(\hat{\lambda}_\mu^1)^2 = O_p(n^{-1/2})$ from Proposition 12(a) and $|\hat{\lambda}_\mu^1| \geq n^{-1/6}(\log n)^{-1}$. Using the argument of the proof of Proposition 11 leading to (83), we obtain

$$2[\ell_n(\hat{\psi}, \hat{\pi}, \xi) - \ell_{0n}(\hat{\vartheta}_0)] = \max\{B_{0n}(\sqrt{n}t_\lambda(\hat{\lambda}^1, \hat{\pi}^1)), B_{\hat{\varrho}^2 n}(\sqrt{n}t_\lambda(\hat{\lambda}^2, \hat{\pi}^2))\} + o_p(1).$$

We proceed to construct parameter spaces that are locally equal to the cones Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$ defined in (46). Define $c(\alpha) := \alpha(1-\alpha)$, and denote the elements of $t_\lambda(\hat{\lambda}^j, \hat{\pi}^j)$ corresponding to (44) by

$$t_\lambda(\hat{\lambda}^j, \hat{\pi}^j) = \begin{pmatrix} \hat{t}^j_{\varrho\mu^2} \\ \hat{t}^j_{\mu^3} \\ \hat{t}^j_{\mu^4} \\ \hat{t}^j_{\beta\mu} \\ \hat{t}^j_{v(\beta)} \end{pmatrix} := \begin{pmatrix} c(\hat{\alpha}^j)\hat{\varrho}^j(\hat{\lambda}^j_\mu)^2 \\ c(\hat{\alpha}^j)(1-2\hat{\alpha}^j)(\hat{\lambda}^j_\mu)^3 \\ c(\hat{\alpha}^j)(1-6\hat{\alpha}^j+6(\hat{\alpha}^j)^2)(\hat{\lambda}^j_\mu)^4 \\ c(\hat{\alpha}^j)\hat{\lambda}^j_\beta\hat{\lambda}^j_\mu \\ c(\hat{\alpha}^j)v(\hat{\lambda}^j_\beta) \end{pmatrix}.$$

Note that $\hat{\lambda}_\beta^1 = O_p(n^{-1/3}\log n)$ because $\hat{t}^1_{\beta\mu} = O_p(n^{-1/2})$ from Proposition 12(a) and $|\hat{\lambda}^1_\mu| \geq n^{-1/6}(\log n)^{-1}$. Therefore, in view of $|\hat{\lambda}^2_\mu| < n^{-1/6}(\log n)^{-1}$,

$$\hat{t}^1_{v(\beta)} = o_p(n^{-1/2}), \quad \hat{t}^2_{\mu^3} = o_p(n^{-1/2}), \quad \hat{t}^2_{\mu^4} = o_p(n^{-1/2}).$$

In view of this, let $t_\lambda(\lambda,\pi) := (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t'_{\beta\mu}, t'_{v(\beta)})' \in \mathbb{R}^{q_\lambda}$, and consider the following sets:

$$\begin{aligned}
\tilde{\Lambda}^1_\lambda := \{t_\lambda(\lambda,\pi) : {} & t_{\varrho\mu^2} = c(\alpha)\varrho\lambda_\mu^2, t_{\mu^3} = c(\alpha)(1-2\alpha)\lambda_\mu^3, t_{\mu^4} = c(\alpha)(1-6\alpha+6\alpha^2)\lambda_\mu^4, \\
& t_{\beta\mu} = c(\alpha)\lambda_\beta\lambda_\mu, t_{v(\beta)} = 0 \text{ for some } (\lambda,\varrho,\alpha) \in \Theta_\lambda \times \Theta_{\varrho\epsilon} \times \Theta_{\alpha\epsilon}\}, \\
\tilde{\Lambda}^2_{\lambda\pi} := \{t_\lambda(\lambda,\pi) : {} & t_{\varrho\mu^2} = c(\alpha)\varrho\lambda_\mu^2, t_{\mu^3} = t_{\mu^4} = 0, \\
& t_{\beta\mu} = c(\alpha)\lambda_\beta\lambda_\mu, t_{v(\beta)} = c(\alpha)v(\lambda_\beta) \text{ for some } \lambda \in \Theta_\lambda\}.
\end{aligned}$$

Define $(\tilde{\lambda}^1, \tilde{\pi}^1)$ and $\tilde{\lambda}^2_\pi$ by $B_{0n}(\sqrt{n}t_\lambda(\tilde{\lambda}^1, \tilde{\pi}^1)) = \max_{t_\lambda(\lambda,\pi)\in\tilde{\Lambda}^1_\lambda} B_{0n}(\sqrt{n}t_\lambda(\lambda,\pi))$ and $B_{\varrho n}(\sqrt{n}t_\lambda(\tilde{\lambda}^2_\pi,\pi)) = \max_{t_\lambda(\lambda,\pi)\in\tilde{\Lambda}^2_{\lambda\pi}} B_{\varrho n}(\sqrt{n}t_\lambda(\lambda,\pi))$. Observe that $\tilde{\Lambda}^1_\lambda$ is locally (in a neighborhood of $\varrho = 0$, $\lambda = 0$) equal to the cone Λ^1_λ because, when $\delta > 0$ is sufficiently small, for any $u \in [-\delta, \delta]$ there exists $\alpha_* \in [0.4, 0.6]$ such that $u = (1-2\alpha_*)/(1-6\alpha_*+6\alpha_*^2)$. $\tilde{\Lambda}^2_{\lambda\pi}$ is locally equal to the cone $\Lambda^2_{\lambda\varrho}$ uniformly in $\varrho \in \Theta_{\varrho\epsilon}$.

Define $W_n := \max\{B_{0n}(\sqrt{n}t_\lambda(\tilde{\lambda}^1, \tilde{\pi}^1)), \sup_{\pi\in\Theta_{\pi\epsilon}} B_{\varrho n}(\sqrt{n}t_\lambda(\tilde{\lambda}^2_\pi, \pi))\}$. Proceeding as in the proof of Proposition 11 gives $2[\ell_n(\hat{\psi}, \hat{\pi}, \xi) - \ell_{0n}(\hat{\vartheta}_0)] = W_n + o_p(1)$, and the asymptotic distribution of the LRTS follows from applying Theorem 1(c) of Andrews (2001) to $(B_{0n}(\sqrt{n}t_\lambda(\tilde{\lambda}^1, \tilde{\pi}^1)), B_{\varrho n}(\sqrt{n}t_\lambda(\tilde{\lambda}^2_\pi, \pi))$.
□

Proof of Propositions 15, 16 and 17. Let $\mathcal{N}_h^*$ denote an arbitrary small neighborhood of Υ_h^*, and let $\hat{\psi}_h$ denote a local MLE that maximizes $\ell_n(\psi_h, \pi_h, \xi_{M_0+1})$ subject to $\psi_h \in \mathcal{N}_h^*$. Proposition 14 and $\Upsilon^* = \cup_{h=1}^{M_0}\Upsilon_h^*$ imply that $\ell_n(\hat{\vartheta}_{M_0+1}, \xi_{M_0+1}) = \max_{h=1,\ldots,M_0} \ell_n(\hat{\psi}_h, \pi_h, \xi_{M_0+1})$ with probability approaching 1. Because $\psi_\ell^* \notin \mathcal{N}_h^*$ for any $\ell \neq h$, it follows from Proposition 14 that $\hat{\psi}_h - \psi_h^* = o_p(1)$.

Next, $\ell_n(\psi_h, \pi_h, \xi_{M_0+1}) - \ell_n(\psi_h^*, \pi_h, \xi_{M_0+1})$ admits the same expansion as $\ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi)$ in (21) or (36). Therefore, the stated result follows from applying the proof of Propositions 9, 11, and 13 to $\ell_n(\hat{\psi}_h, \pi_h, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})$ for each h and combining the results to derive the joint asymptotic distribution of $\{\ell_n(\hat{\psi}_h, \pi_h, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})\}_{h=1}^{M_0}$. □

8.3 Auxiliary results

8.3.1 Missing information principle

The following lemma extends equations (3.1)-(3.2) in Louis (1982), expressing the higher order derivatives of the log-likelihood function in terms of the conditional expectation of the derivatives of the complete data log-likelihood function. For notational brevity, assume ϑ is scalar. Let $\nabla^j \ell(Y) := \nabla^j_\vartheta \log P(Y;\vartheta)$ and $\nabla^j \ell(Y, X) := \nabla^j_\vartheta \log P(Y, X;\vartheta)$. For random variables $V_1, \ldots, V_q$ and Y, define the central conditional moment of $(V_1^{r_1} \cdots V_q^{r_q})$ as $\mathbb{E}^c[V_1^{r_1} \cdots V_q^{r_q}|Y] := \mathbb{E}[(V_1 - \mathbb{E}[V_1|Y])^{r_1} \cdots (V_q - \mathbb{E}[V_q|Y])^{r_q}|Y]$.

Lemma 1. *For any random variables X and Y with density $P(Y, X;\theta)$ and $P(Y;\theta)$,*

$$
\begin{aligned}
&\nabla \ell(Y) = \mathbb{E}\left[\nabla \ell(Y,X)|Y\right], \quad \nabla^2 \ell(Y) = \mathbb{E}\left[\nabla^2 \ell(Y,X)\middle|Y\right] + \mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right], \\
&\nabla^3 \ell(Y) = \mathbb{E}\left[\nabla^3 \ell(Y,X)\middle|Y\right] + 3\mathbb{E}^c\left[\nabla^2 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right] + \mathbb{E}^c\left[(\nabla \ell(Y,X))^3\middle|Y\right], \\
&\nabla^4 \ell(Y) = \mathbb{E}\left[\nabla^4 \ell(Y,X)\middle|Y\right] + 4\mathbb{E}^c\left[\nabla^3 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right] + 3\mathbb{E}^c\left[(\nabla^2 \ell(Y,X))^2\middle|Y\right] \\
&\quad + 6\mathbb{E}^c\left[\nabla^2 \ell(Y,X)(\nabla \ell(Y,X))^2\middle|Y\right] + \mathbb{E}^c\left[(\nabla \ell(Y,X))^4\middle|Y\right] - 3\left\{\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right]\right\}^2, \\
&\nabla^5 \ell(Y) = \mathbb{E}\left[\nabla^5 \ell(Y,X)\middle|Y\right] + 5\mathbb{E}^c\left[\nabla^4 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right] + 10\mathbb{E}^c\left[\nabla^3 \ell(Y,X)\nabla^2 \ell(Y,X)\middle|Y\right] \\
&\quad + 10\mathbb{E}^c\left[\nabla^3 \ell(Y,X)(\nabla \ell(Y,X))^2\middle|Y\right] + 15\mathbb{E}^c\left[(\nabla^2 \ell(Y,X))^2\nabla \ell(Y,X)\middle|Y\right] \\
&\quad + 10\mathbb{E}^c\left[\nabla^2 \ell(Y,X)(\ell(Y,X))^3\middle|Y\right] - 30\mathbb{E}^c\left[\nabla^2 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right] \\
&\quad + \mathbb{E}^c\left[(\nabla \ell(Y,X))^5\middle|Y\right] - 10\mathbb{E}^c\left[(\nabla \ell(Y,X))^3\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right], \\
&\nabla^6 \ell(Y) = \mathbb{E}\left[\nabla^6 \ell(Y,X)\middle|Y\right] \\
&\quad + 6\mathbb{E}^c\left[\nabla^5 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right] + 15\mathbb{E}^c\left[\nabla^4 \ell(Y,X)\nabla^2 \ell(Y,X)\middle|Y\right] \\
&\quad + 15\mathbb{E}^c\left[\nabla^4 \ell(Y,X)(\nabla \ell(Y,X))^2\middle|Y\right] + 60\mathbb{E}^c\left[\nabla^3 \ell(Y,X)\nabla^2 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right] \\
&\quad + 10\mathbb{E}^c\left[(\nabla^3 \ell(Y,X))^2\middle|Y\right] + 15\mathbb{E}^c\left[(\nabla^2 \ell(Y,X))^3\middle|Y\right] \\
&\quad + 20\mathbb{E}^c\left[\nabla^3 \ell(Y,X)(\nabla \ell(Y,X))^3\middle|Y\right] - 60\mathbb{E}^c\left[\nabla^3 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right]\mathbb{E}\left[(\nabla \ell(Y,X))^2\middle|Y\right] \\
&\quad + 45\mathbb{E}^c\left[(\nabla^2 \ell(Y,X))^2(\nabla \ell(Y,X))^2\middle|Y\right] - 90\left\{\mathbb{E}^c\left[\nabla^2 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right]\right\}^2 \\
&\quad - 45\mathbb{E}^c\left[(\nabla^2 \ell(Y,X))^2\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right] \\
&\quad + 15\mathbb{E}^c\left[\nabla^2 \ell(Y,X)(\nabla \ell(Y,X))^4\middle|Y\right] - 90\mathbb{E}^c\left[\nabla^2 \ell(Y,X)(\nabla \ell(Y,X))^2\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right] \\
&\quad - 60\mathbb{E}^c\left[\nabla^2 \ell(Y,X)\nabla \ell(Y,X)\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^3\middle|Y\right] \\
&\quad + \mathbb{E}^c\left[(\nabla \ell(Y,X))^6\middle|Y\right] - 15\mathbb{E}^c\left[(\nabla \ell(Y,X))^4\middle|Y\right]\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right] \\
&\quad - 10\left\{\mathbb{E}^c\left[(\nabla \ell(Y,X))^3\middle|Y\right]\right\}^2 + 30\left\{\mathbb{E}^c\left[(\nabla \ell(Y,X))^2\middle|Y\right]\right\}^3.
\end{aligned}
$$

provided that the conditional expectation on the right hand side exists. When $P(Y;\theta)$ in the left hand side is replaced with $P(Y|Z;\theta)$, the stated result holds with $P(Y,X;\theta)$ and $\mathbb{E}[\cdot|Y]$ on the right hand side replaced with $P(Y,X|Z;\theta)$ and $\mathbb{E}[\cdot|Y,Z]$.

Proof of Lemma 1. The stated result follows from a direct calculation and relations such as $\nabla_\vartheta^j P(Y;\vartheta)/P(Y;\vartheta) = \mathbb{E}[\nabla_\vartheta^j P(Y,X;\vartheta)/P(Y,X;\vartheta)|Y]$ and

$$
\begin{aligned}
\nabla \log f &= \nabla f/f, \quad \nabla^2 \log f = \nabla^2 f/f - (\nabla \log f)^2,\\
\nabla^3 \log f &= \nabla^3 f/f - 3\nabla^2 f \nabla f/f^2 + 2(\nabla f/f)^3,\\
\nabla^4 \log f &= \nabla^4 f/f - 4\nabla^3 f\nabla f/f^2 - 3(\nabla^2 f/f)^2 + 12\nabla^2 f(\nabla f)^2/f^3 - 6(\nabla f/f)^4,\\
\nabla^5 \log f &= \nabla^5 f/f - 5\nabla^4 f\nabla f/f^2 - 10\nabla^3 f\nabla^2 f/f^2 + 20\nabla^3 f(\nabla f)^2/f^3\\
&\quad + 30(\nabla^2 f)^2\nabla f/f^3 - 60\nabla^2 f(\nabla f)^3/f^4 + 24(\nabla f/f)^5,\\
\nabla^6 \log f &= \nabla^6 f/f - 6\nabla^5 f\nabla f/f^2 - 15\nabla^4 f\nabla^2 f/f^2 + 30\nabla^4 f(\nabla f)^2/f^3 - 10(\nabla^3 f)^2/f^2\\
&\quad + 120\nabla^3 f\nabla^2 f\nabla f/f^3 - 120\nabla^3 f(\nabla f)^3/f^4 + 30(\nabla^2 f)^3/f^3\\
&\quad - 270(\nabla^2 f)^2(\nabla f)^2/f^4 + 360\nabla^2 f(\nabla f)^4/f^5 - 120(\nabla f)^6/f^6, \qquad (87)\\
\nabla^3 f/f &= \nabla^3 \log f + 3\nabla^2 \log f\nabla \log f + (\nabla \log f)^3,\\
\nabla^4 f/f &= \nabla^4 \log f + 4\nabla^3 \log f\nabla \log f + 3(\nabla^2 \log f)^2 + 6\nabla^2 \log f(\nabla \log f)^2 + (\nabla \log f)^4,\\
\nabla^5 f/f &= \nabla^5 \log f + 5\nabla^4 \log f\nabla \log f + 10\nabla^3 \log f\nabla^2 \log f + 10\nabla^3 \log f(\nabla \log f)^2\\
&\quad + 15(\nabla^2 \log f)^2\nabla \log f + 10\nabla^2 \log f(\nabla \log f)^3 + (\nabla \log f)^5,\\
\nabla^6 f/f &= \nabla^6 \log f + 6\nabla^5 \log f\nabla \log f + 15\nabla^4 \log f\nabla^2 \log f + 15\nabla^4 \log f(\nabla \log f)^2\\
&\qquad + 10(\nabla^3 \log f)^2 + 60\nabla^3 \log f\nabla^2 \log f\nabla \log f + 20\nabla^3 \log f(\nabla \log f)^3\\
&\qquad + 15(\nabla^2 \log f)^3 + 45(\nabla^2 \log f)^2(\nabla \log f)^2 + 15\nabla^2 \log f(\nabla \log f)^4 + (\nabla \log f)^6.
\end{aligned}
$$

For example, $\nabla^3 \ell(Y)$ is derived by writing $\nabla^3 \ell(Y)$ as, with suppressing ϑ,

$$
\begin{aligned}
&\nabla^3 \ell(Y)\\
&= \frac{\nabla^3 P(Y)}{P(Y)} - 3\frac{\nabla^2 P(Y)}{P(Y)}\frac{\nabla P(Y)}{P(Y)} + 2\left(\frac{\nabla P(Y)}{P(Y)}\right)^3\\
&= \mathbb{E}\left[\left.\frac{\nabla^3 P(Y,X)}{P(Y,X)}\right|Y\right] - 3\mathbb{E}\left[\left.\frac{\nabla^2 P(Y,X)}{P(Y,X)}\right|Y\right]\mathbb{E}\left[\left.\frac{\nabla P(Y,X)}{P(Y,X)}\right|Y\right] + 2\left\{\mathbb{E}\left[\left.\frac{\nabla P(Y,X)}{P(Y,X)}\right|Y\right]\right\}^3\\
&= \mathbb{E}\left[\nabla^3 \ell(Y,X) + 3\nabla^2 \ell(Y,X)\nabla \ell(Y,X) + (\nabla \ell(Y,X))^3 \middle| Y\right]\\
&\quad - 3\mathbb{E}\left[\nabla^2 \ell(Y,X) + (\nabla \ell(Y,X))^2 \middle| Y\right]\mathbb{E}\left[\nabla \ell(Y,X)|Y\right] + 2\left\{\mathbb{E}\left[\nabla \ell(Y,X)|Y\right]\right\}^3,
\end{aligned}
$$

and collecting terms. $\nabla^4 \ell(Y)$, $\nabla^5 \ell(Y)$, and $\nabla^6 \ell(Y)$ are derived similarly. □

8.3.2 Auxiliary Lemmas

Henceforth, the conditioning variable W_{-m+1}^n is suppressed from the conditioning sets and conditional densities unless confusions might arise.

The following Lemma provides bounds on $\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\mathcal{F}]$ defined in (7) and (60) and is used in the proof of Lemma 3. For $j = 2, \ldots, 6$, define $||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty := \sum_{(\ell_1,\ldots,\ell_j)\in\sigma(\mathcal{I}(j))} ||\phi_{t_1}^{\ell_1}||_\infty \cdots ||\phi_{t_j}^{\ell_j}||_\infty$, where $||\phi_t^i||_\infty = \sup_{\vartheta\in\mathcal{N}^*}\sup_{x,x'} |\phi^i(\vartheta, Y_t, x, \overline{Y}_{t-1}, x')|$ as defined in the proof of Lemma 3.

Lemma 2. *Under Assumptions 1, 2, and 4, there exists a finite nonstochastic constant C that does not depend on ρ such that, for all $m' \geq m \geq 0$, all $-m < t_1 \leq t_2 \leq \cdots \leq t_j \leq n$, all $\vartheta \in \mathcal{N}^*$ and all $x \in \mathcal{X}$, and $j = 2, \ldots, 6$,*

$$
\begin{aligned}
&(a)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n]| \leq C\rho^{(t_2-t_1-1)_+\vee(t_3-t_2-1)_+\vee\cdots\vee(t_j-t_{j-1}-1)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty,\\
&(b)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n, X_{-m}=x]| \leq C\rho^{(t_2-t_1-1)_+\vee(t_3-t_2-1)_+\vee\cdots\vee(t_j-t_{j-1}-1)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty,\\
&(c)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n, X_{-m}=x] - \Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n]| \leq C\rho^{(m+t_1-1)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty,\\
&(d)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n, X_{-m}=x] - \Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m'}^n, X_{-m'}=x]| \leq C\rho^{(m+t_1-1)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty,\\
&(e)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n] - \Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^{n-1}]| \leq C\rho^{(n-1-t_j)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty,\\
&(f)\quad |\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^n, X_{-m}=x] - \Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\overline{Y}_{-m}^{n-1}, X_{-m}=x]| \leq C\rho^{(n-1-t_j)_+}||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty.
\end{aligned}
$$

Proof of Lemma 2. Recall $\sup_{\vartheta\in\mathcal{N}^*}\sup_{x,x'} |\phi^i(\vartheta, Y_t, x, \overline{Y}_{t-1}, x') - \mathbb{E}_\vartheta[\phi^i(\vartheta, Y_t, x, \overline{Y}_{t-1}, x')|\mathcal{F}]|$ $\leq 2\sup_{\vartheta\in\mathcal{N}^*}\sup_{x,x'} |\phi^i(\vartheta, Y_t, x, \overline{Y}_{t-1}, x')|$ for the conditioning sets $\mathcal{F}$ that appear in the lemma. Define $\tilde{\phi}_{\vartheta t}^i := \phi^i(\vartheta, \overline{Z}_{t-1}^t) - \mathbb{E}_\vartheta[\phi^i(\vartheta, \overline{Z}_{t-1}^t)|\overline{Y}_{-m}^n]$, so that $\mathbb{E}_\vartheta^c[\phi_{\vartheta t_1}^{\ell_1}\cdots\phi_{\vartheta t_j}^{\ell_j}|\overline{Y}_{-m}^n] = \mathbb{E}_\vartheta[\tilde{\phi}_{\vartheta t_1}^{\ell_1}\cdots\tilde{\phi}_{\vartheta t_j}^{\ell_j}|\overline{Y}_{-m}^n]$. Henceforth, we suppress the subscript ϑ from $\phi_{\vartheta t}^i$ and $\tilde{\phi}_{\vartheta t}^i$.

Observe that $\phi^i(\vartheta, \overline{Z}_{t-1}^t)$ depends on X_t and X_{t-1}. For $j = 2, \ldots, 6$, parts (c) and (d) follow from Corollary 2(b), parts (e) and (f) for $t_j \leq n-1$ follow from Corollary 2(c), and parts (e) and (f) for $t_j = n$ follow from $|\Phi_{\vartheta\mathcal{T}(j)}^{\mathcal{I}(j)}[\cdot]| \leq 2^j||\phi_{\mathcal{T}(j)}^{\mathcal{I}(j)}||_\infty$.

We proceed to show parts (a) and (b) for $j = 2, \ldots, 6$. The results for $j = 2$ and $j = 3$ follow from Corollary 2 and

$$
\begin{aligned}
E(X_{t_1} - EX_{t_1})\cdots(X_{t_j} - EX_{t_j}) &= \mathrm{cov}[X_{t_1}, (X_{t_2} - EX_{t_2})\cdots(X_{t_j} - EX_{t_j})]\\
&= \mathrm{cov}[(X_{t_1} - EX_{t_1})\cdots(X_{t_{j-1}} - EX_{t_{j-1}}), X_{t_j}].
\end{aligned}
\tag{88}
$$

Before proving the results for $j \geq 4$, we collect some results. For a conditioning set $\mathcal{F} = \overline{Y}_{-m}^n$ or

$\{\overline{Y}^n_{-m}, X_m = x\}$, Corollary 2(a) and (88) imply that, with $\sigma_5, \sigma_{61}, \sigma_{62}, \sigma_{63}$ defined in (61),

$$|\mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_j-t_{j-1}-1)_+}||\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}||_\infty, \tag{89}$$

$$\max_{(\{a,b\},\{c,d\})\in D_1}|\mathbb{E}^c_\vartheta[\phi^{\ell_a}_{t_a}\phi^{\ell_b}_{t_b}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_c}_{t_c}\phi^{\ell_d}_{t_d}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_3-t_2-1)_+\vee(t_4-t_3-1)_+}||\phi^{\mathcal{I}(4)}_{\mathcal{T}(4)}||_\infty, \tag{90}$$

$$\begin{aligned}&|\mathbb{E}^c_\vartheta[\phi^{\ell_a}_{t_a}\phi^{\ell_b}_{t_b}\phi^{\ell_c}_{t_c}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_d}_{t_d}\phi^{\ell_e}_{t_e}|\mathcal{F}]| \leq \mathcal{C}||\phi^{\mathcal{I}(5)}_{\mathcal{T}(5)}||_\infty \\ &\times\begin{cases}\rho^{(t_2-t_1-1)_+\vee(t_3-t_2-1)_+\vee(t_5-t_4-1)_+} & \text{for } (\{a,b,c\},\{d,e\})\in\sigma_5\setminus(\{1,2,3\},\{4,5\}),\\ \rho^{(t_2-t_1-1)_+\vee(t_4-t_3-1)_+\vee(t_5-t_4-1)_+} & \text{for } (\{a,b,c\},\{d,e\})\in\sigma_5\setminus(\{3,4,5\},\{1,2\}),\end{cases}\end{aligned} \tag{91}$$

$$\max_{(\{a,b,c,d\},\{e,f\})\in D_2}|\mathbb{E}^c_\vartheta[\phi^{\ell_a}_{t_a}\phi^{\ell_b}_{t_b}\phi^{\ell_c}_{t_c}\phi^{\ell_d}_{t_d}|\overline{Y}^n_{-m}]\mathbb{E}^c_\vartheta[\phi^{\ell_e}_{t_e}\phi^{\ell_f}_{t_f}|\overline{Y}^n_{-m}]| \leq \mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty, \tag{92}$$

$$\max_{(\{a,b,c\},\{d,e,f\})\in D_3}|\mathbb{E}^c_\vartheta[\phi_{t_a}\phi_{t_b}\phi_{t_c}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_d}\phi_{t_e}\phi_{t_f}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_3-t_2-1)_+\vee\cdots\vee(t_6-t_5-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty, \tag{93}$$

$$\max_{(\{a,b\},\{c,d\},\{e,f\})\in D_4}|\mathbb{E}^c_\vartheta[\phi_{t_a}\phi_{t_b}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty, \tag{94}$$

$$\begin{aligned}&|\mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_k}_{t_k}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_{k+1}}_{t_{k+1}}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}]| \\ &\quad = \text{cov}_\vartheta[\tilde{\phi}^{\ell_1}_{t_1}\cdots\tilde{\phi}^{\ell_k}_{t_k}, \tilde{\phi}^{\ell_{k+1}}_{t_{k+1}}\cdots\tilde{\phi}^{\ell_j}_{t_j}|\mathcal{F}] \leq \mathcal{C}\rho^{(t_{k+1}-t_k-1)_+}||\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}||_\infty \quad \text{for any } 2\leq k\leq j-2,\end{aligned} \tag{95}$$

where $D_1 = (\{1,3\},\{2,4\}),(\{1,4\},\{2,3\})$ in (90), $D_2 = \sigma_{61}\setminus((\{3,4,5,6\},\{1,2\})\cup X_{61})$ in (92), $D_3 = \sigma_{62}\setminus(\{1,2,3\},\{4,5,6\})$ in (93), and $D_4 = \sigma_{63}\setminus X_{63}$ in (94), where X_{61} is the set of $\binom{4}{2} = 6$ partitions of $\{1,2,3,4,5,6\}$ of the form of $\{1,2,c,d\},\{e,f\}$, and $X_{63} := \{(\{1,2\},\{3,4\},\{5,6\}),(\{1,2\},\{3,5\},\{4,6\}),($

For $j = 4$, parts (a) and (b) follow from combining (89)–(90) and (95) with $k = 2$ because, for example, $\Phi^{1,1,1,1}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_4-t_3-1)_+}||\phi^{1,1,1,1}_{\mathcal{T}(4)}||_\infty$ from (89)–(90) and $\Phi^{1,1,1,1}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] = \text{cov}_\vartheta[\tilde{\phi}_{t_1}\tilde{\phi}_{t_2},\tilde{\phi}_{t_3}\tilde{\phi}_{t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_3}|\mathcal{F}]\,\mathbb{E}^c_\vartheta[\phi_{t_2}\phi_{t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_4}|\mathcal{F}]\,\mathbb{E}^c_\vartheta[\phi_{t_2}\phi_{t_3}|\mathcal{F}] \leq \mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{1,1,1,1}_{\mathcal{T}(4)}||_\infty$ from (90) and (95). For $j = 5$, parts (a) and (b) follow from combining (89), (91) and (95) with $k = 2,3$.

For $j = 6$, first, $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ is bounded by $\mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_6-t_5-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$ from (89). Second, let $A = \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_3}|\mathcal{F}]\,\mathbb{E}^c_\vartheta[\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}]$, then all the terms on the right hand side of $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ in (60) except for A are bounded by $\mathcal{C}\rho^{(t_4-t_3-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$ from (89) and (93), and A is bounded by $\mathcal{C}\rho^{(t_4-t_3-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$ from (95). Therefore, $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ is bounded by $\mathcal{C}\rho^{(t_4-t_3-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$. Third, let $B_1 = \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}|\mathcal{F}]$ and $B_2 = -\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\mathbb{E}^c_\vartheta[\phi_1\phi_2\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}] + 2\sum_{(\{a,b\},\{c,d\},\{e,f\})\in X_{63}}\mathbb{E}^c_\vartheta[\phi_{t_a}\phi_{t_b}|\mathcal{F}] \times\mathbb{E}^c_\vartheta[\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]$, then all the terms on the right hand side of $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ except for B_1+B_2 are bounded by $\mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$ from (93)–(94). Further, we can write B_2 as $\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\{-\mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}] + \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]\}$ $= -\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]\text{cov}_\vartheta[\tilde{\phi}_{\theta t_1}\tilde{\phi}_{\theta t_2},\tilde{\phi}_{\theta t_c}\tilde{\phi}_{\theta t_d}|\mathcal{F}] \leq \mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$ from (95). Therefore, $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ is bounded by $\mathcal{C}\rho^{(t_3-t_2-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$. From a similar argument, $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ is also bounded by $\mathcal{C}\rho^{(t_5-t_4-1)_+}||\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}||_\infty$, and parts (a) and (b) follow. □

We next present the result that extends Lemmas 13 and 17 of DMR. Let $r_{\mathcal{I}(1)} = q_{i_1}$; $r_{\mathcal{I}(2)} = q_{i_1}/2$ if $i_1 = i_2$ and $(q_{i_1} \wedge q_{i_2})/2$ if $i_1 \neq i_2$; $r_{\mathcal{I}(3)} = q_{i_1}/3$ if $i_1 = i_2 = i_3$, $(q_{i_1}/2 \wedge q_{i_2}/4)$ if $i_1 \neq i_2 = i_3$, $(q_{i_1} \wedge q_{i_2} \wedge q_{i_3})/3$ if i_1, i_2, i_3 are distinct; $r_{\mathcal{I}(4)} = q_{i_1}/4$ if $i_1 = i_2 = i_3 = i_4$, $(q_{i_1} \wedge q_{i_3})/4$ if $i_1 \neq i_2 = i_3 = i_4$ or $i_1 = i_2 \neq i_3 = i_4$; $r_{\mathcal{I}(5)} = q_{i_1}/5$ if $i_1 = i_2 = i_3 = i_4 = i_5$; $(q_{i_1}/4 \wedge q_{i_2}/6)$ if $i_1 \neq i_2 = i_3 = i_4 = i_5$; $r_{\mathcal{I}(6)} = q_1/6$.

Lemma 3. *Under Assumptions 1, 2, and 4, for $j = 1, \ldots, 6$, there exist random variables $K_{\mathcal{I}(j)}, M_{\mathcal{I}(j)} \in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ such that, for all $1 \leq k \leq n$ and $m' \geq m \geq 0$,*

$$
\begin{aligned}
&(a) \qquad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta) - \overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta)| \leq K_{\mathcal{I}(j)}(k+m)^7 \rho^{\lfloor (k+m-1)/24 \rfloor}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},\\
&(b) \qquad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta) - \Delta^{\mathcal{I}(j)}_{j,k,m',x}(\vartheta)| \leq K_{\mathcal{I}(j)}(k+m)^7 \rho^{\lfloor (k+m-1)/1340 \rfloor}, \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},
\end{aligned}
$$

(c) $\sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} \sup_{m \geq 0} |\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)| + \sup_{\vartheta \in \mathcal{N}^*} \sup_{m \geq 0} |\overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta)| \leq M_{\mathcal{I}(j)}$, *$\mathbb{P}_{\vartheta^*}$-a.s. (d) Uniformly in $\vartheta \in \mathcal{N}^*$, $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$ and $\overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta)$ converge $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ to $\Delta^{\mathcal{I}(j)}_{j,k,\infty}(\vartheta) \in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ as $m \to \infty$.*

Proof of Lemma 3. Let $||\phi^i_t||_\infty := \sup_{\vartheta \in \mathcal{N}^*} \sup_{x,x'} |\phi^i(\vartheta, Y_t, x, \overline{Y}_{t-1}, x')|$. For each j, Part (c) follows from parts (a) and (b), $\sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Delta^{\mathcal{I}(j)}_{j,0,0,x}(\vartheta)| \in L^{\mathcal{I}(j)}$, and the distributional equivalence between $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$ and $\Delta^{\mathcal{I}(j)}_{j,0,k+m,x}(\vartheta)$. Part (d) follows from parts (a)-(c) because parts (a)-(c) imply that $\{\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)\}_{m \geq 0}$ and $\{\Delta^{\mathcal{I}(j)}_{j,k,m}(\vartheta)\}_{m \geq 0}$ are uniform $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$-Cauchy sequences with respect to $\vartheta \in \mathcal{N}^*$ that converge to the same limit and $L^q(\mathbb{P}_{\vartheta^*})$ is complete.

We prove parts (a) and (b) by extending the argument in the proof of Lemmas 13 and 17 in DMR. Recall $\mathcal{T}(j) = (t_1, \ldots, t_j)$. For part (a), define, suppressing the dependence of $A_{\mathcal{T}(j)}$ on ϑ and $\mathcal{I}(j)$,

$$
A_{\mathcal{T}(j)} := \begin{cases} \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^k_{-m}, X_{-m} = x\right] - \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^{k-1}_{-m}, X_{-m} = x\right] & \\ \quad - \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^k_{-m}\right] + \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^{k-1}_{-m}\right], & \text{if } \max\{t_1, \ldots, t_j\} < k, \\ \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^k_{-m}, X_{-m} = x\right] - \Phi^{\mathcal{I}(j)}_{\vartheta \mathcal{T}(j)}\left[\overline{Y}^k_{-m}\right], & \text{otherwise}, \end{cases}
$$

$$
A_{\mathcal{T}(j,\ell,k)} := A_{t_1 t_2 \cdots t_{j-\ell} \underbrace{k \cdots k}_{\ell \text{ times}}}, \quad \text{where } \mathcal{T}(j,\ell,k) := (\mathcal{T}(j-\ell), \underbrace{k, \cdots, k}_{\ell \text{ times}}).
$$

Then, we can write $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta) - \overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta) = \sum_{\mathcal{T}(j) \in \{-m+1,\ldots,k\}^j} A_{\mathcal{T}(j)} = \Delta_a + \Delta_b$, where

$$
\Delta_a := \sum_{\mathcal{T}(j) \in \{-m+1,\ldots,k-1\}^j} A_{\mathcal{T}(j)}, \quad \Delta_b := \sum_{\ell=1}^{j-1} \binom{j}{\ell} \sum_{\mathcal{T}(j-\ell) \in \{-m+1,\ldots,k-1\}^{j-\ell}} A_{\mathcal{T}(j,\ell,k)} + A_{(k,\ldots,k)},
$$

and $\sum_{\ell=1}^{j-1}$ is defined as 0 when $j = 1$. From Lemma 2 and the symmetry of $A_{\mathcal{T}(j)}$, Δ_a is bounded by $\mathcal{C}B_{j,k,m}M^{\mathcal{I}(j)}_{j,k,m}$, where

$$B_{j,k,m} := \sum_{-m+1\le t_1\le t_2\le\cdots\le t_j\le k-1}\left(\rho^{(m+t_1-1)_+}\wedge\rho^{(t_2-t_1-1)_+}\wedge\cdots\wedge\rho^{(t_j-t_{j-1}-1)_+}\wedge\rho^{(k-1-t_j-1)_+}\right)$$
$$= \sum_{1\le t_1\le t_2\le\cdots\le t_j\le k+m-1}\left(\rho^{(t_1-1)_+}\wedge\rho^{(t_2-t_1-1)_+}\wedge\cdots\wedge\rho^{(t_j-t_{j-1}-1)_+}\wedge\rho^{(k+m-1-t_j-1)_+}\right),$$
$$M^{\mathcal{I}(j)}_{j,k,m} := \max_{-m+1\le t_1,\dots,t_j\le k-1}||\phi^{i_1}_{t_1}||_\infty||\phi^{i_2}_{t_2}||_\infty\cdots||\phi^{i_j}_{t_j}||_\infty.$$

From $(t-1)_+\ge\lfloor t/2\rfloor$ and Lemma 14, $B_{j,k,m}$ is bounded by $C_{j2}(\rho)\rho^{\lfloor(k+m-1)/4j\rfloor}$.

We proceed to derive a bound on $M^{\mathcal{I}(j)}_{j,k,m}$. Define $||\phi^i||^\ell_\infty := \sum_{t=-\infty}^{\infty}(|t|\vee 1)^{-2}||\phi^i_t||^\ell_\infty$. When $i_1=i_2=\cdots=i_j$, from Lemma 15 we have $M^{\mathcal{I}(j)}_{j,k,m}\le(k+m)^{j+1}||\phi^{i_1}||^j_\infty$. In the other cases, from Lemma 15 and the relation $2|xy|\le x^2+y^2$, $3|xyz|\le|x|^3+|y|^3+|z|^3$, $|xy_1y_2y_3|\le(|xy_1|^2+y_2^2y_3^2)/2\le(x^4+y_1^4+2y_2^2y_3^2)/4$, $|xy_1y_2y_3y_4|\le(|xy_1|^2+y_2^2y_3^2y_4^2)/2\le(x^4+y_1^4+2y_2^2y_3^2y_4^2)/4$, we can bound $M^{\mathcal{I}(j)}_{j,k,m}$ by

$$\begin{aligned}
j=2\text{ and }i_1\ne i_2: &\quad (k+m)^2(||\phi^{i_1}||^2_\infty+||\phi^{i_2}||^2_\infty),\\
j=3\text{ and }i_1\ne i_2=i_3: &\quad (k+m)^3(||\phi^{i_1}||^2_\infty+||\phi^{i_2}||^4_\infty),\\
j=3\text{ and }i_1,i_2,i_3\text{ are distinct}: &\quad (k+m)^2(||\phi^{i_1}||^3_\infty+||\phi^{i_2}||^3_\infty+||\phi^{i_3}||^3_\infty),\\
j=4\text{ and }i_1\ne i_2=i_3=i_4: &\quad (k+m)^3(||\phi^{i_1}||^4_\infty+||\phi^{i_2}||^4_\infty),\\
j=4\text{ and }i_1=i_2\ne i_3=i_4: &\quad (k+m)^3(||\phi^{i_1}||^4_\infty+||\phi^{i_3}||^4_\infty),\\
j=5\text{ and }i_1\ne i_2=i_3=i_4=i_5: &\quad (k+m)^3(||\phi^{i_1}||^4_\infty+||\phi^{i_2}||^6_\infty).
\end{aligned}$$

Therefore, Δ_a is bounded by the right hand side of part (a). From Lemmas 2 and 14, Δ_b is bounded by $\mathcal{C}[\sum_{-m+1\le t_1\le\cdots\le t_{j-1}\le k-1}(\rho^{(m+t_1-1)_+}\wedge\rho^{(t_2-t_1-1)_+}\wedge\cdots\wedge\rho^{(k-1-t_{j-1}-1)_+})+\rho^{(m+k-1)_+}]M^{\mathcal{I}(j)}_{j,k+1,m}\le\mathcal{C}\rho^{\lfloor(k+m-1)/4(j-1)\rfloor}M^{\mathcal{I}(j)}_{j,k+1,m}$, and part (a) of the lemma follows.

For part (b), define, for $-m'+1\le t_1,\dots,t_j\le k$,

$$D_{\mathcal{T}(j),m',x} := \begin{cases}\Phi^{\mathcal{I}(j)}_{\theta\mathcal{T}(j)}[\overline{Y}^k_{-m'},X_{-m'}=x]-\Phi^{\mathcal{I}(j)}_{\theta\mathcal{T}(j)}[\overline{Y}^{k-1}_{-m'},X_{-m'}=x], & \text{if }\max\{t_1,\dots,t_j\}<k,\\ \Phi^{\mathcal{I}(j)}_{\theta\mathcal{T}(j)}[\overline{Y}^k_{-m'},X_{-m'}=x], & \text{otherwise,}\end{cases}$$

and define $D_{\mathcal{T}(j),m,x}$ similarly. Then, we can write $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\theta)=\sum_{\mathcal{T}(j)\in\{-m+1,\dots,k\}^j}D_{\mathcal{T}(j),m,x}$ and $\Delta^{\mathcal{I}(j)}_{j,k,m',x}(\theta)=\sum_{\mathcal{T}(j)\in\{-m'+1,\dots,k\}^j}D_{\mathcal{T}(j),m',x}=\Delta_c+\Delta_d$, where $\Delta_c:=\sum_{\mathcal{T}(j)\in\{-m+1,\dots,k\}^j}D_{\mathcal{T}(j),m',x}$ and

$$\Delta_d := \sum_{\ell=1}^{j}\binom{j}{\ell}\sum_{t_1=-m'+1}^{-m}\cdots\sum_{t_\ell=-m'+1}^{-m}\sum_{t_{\ell+1}=-m+1}^{k}\cdots\sum_{t_j=-m+1}^{k}D_{\mathcal{T}(j),m',x}.$$

By the same argument as part (a), $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\theta)-\Delta_c$ is bounded by the right hand of part (a). For

Δ_d, observe that, with $M_j := \max_{1\le \ell\le j} \binom{j}{\ell}$,

$$
\begin{aligned}
|\Delta_d| &\le M_j \sum_{\ell=1}^{j} \sum_{t_1=-m'+1}^{-m} \sum_{t_2=-m'+1}^{-m} \cdots \sum_{t_\ell=-m'+1}^{-m} \sum_{t_{\ell+1}=-m+1}^{k} \cdots \sum_{t_j=-m+1}^{k} \left| D_{\mathcal{T}(j),m',x} \right| \\
&\le j M_j \sum_{t_1=-m'+1}^{-m} \sum_{t_2=-m'+1}^{k} \cdots \sum_{t_j=-m'+1}^{k} \left| D_{\mathcal{T}(j),m',x} \right| \\
&\le j M_j j! \sum_{t_1=-m'+1}^{-m} \sum_{t_1 \le t_2 \le \cdots \le t_j \le k} \left| D_{\mathcal{T}(j),m',x} \right| .
\end{aligned}
$$

From Lemma 2, **if** $t_1 \le \cdots \le t_j$**,** we have $|D_{\mathcal{T}(j),m',x}| \le \mathcal{C}[\mathbb{I}\{t_j < k\}(\rho^{(t_2-t_1-1)_+} \wedge \rho^{(t_j-t_{j-1}-1)_+} \wedge \cdots \wedge \rho^{(k-1-t_j-1)_+}) + \mathbb{I}\{t_j = k\}(\rho^{(t_2-t_1-1)_+} \wedge \cdots \wedge \rho^{(t_j-t_{j-1}-1)_+})]||\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}||_\infty$. Hence, part (b) follows from Lemma 16. $\square$

Lemma 4. *Under Assumptions 2 and 4, there exist a random variable* $K \in L^{(1+\varepsilon)\max\{2,\dim(\vartheta)\}/\varepsilon}(\mathbb{P}_{\vartheta^*})$ *and* $\rho \in (0,1)$ *such that, for all* $k \ge 1$ *and* $m' \ge m \ge 0$*,*

$$
\sup_{\vartheta\in\mathcal{N}^*} \left| \frac{\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})}{\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m})} \right| \le K, \quad \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} \left| \frac{p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x)}{p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x)} - \frac{\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})}{\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m})} \right| \le K\rho^{k+m-1}.
$$

Furthermore, these bounds hold uniformly in $x \in \mathcal{X}$ *when* $\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})$ *and* $\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m})$ *are replaced with* $p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m'}, X_{-m'}=x)$ *and* $p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m'}, X_{-m'}=x)$*.*

Proof of Lemma 4. The first result follows from noting that $\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}) = \sum_{(x_{k-1},x_k)\in\mathcal{X}^2} g_\vartheta(Y_k|\overline{Y}_{k-1}, x_k)$ $\times q_{\vartheta_x}(x_{k-1},x_k)\mathbb{P}_\vartheta(x_{k-1}|\overline{Y}^{k-1}_{-m}) \in [\sigma_- G_{\vartheta k}, \sigma_+ G_{\vartheta k}]$ and using Assumption 4(b). For the second result, observe that $|p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x) - \overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})| \le \sum_{(x_{k-1},x_k)\in\mathcal{X}^2} g_\vartheta(Y_k|\overline{Y}_{k-1}, x_k) q_{\vartheta_x}(x_{k-1},x_k)$ $\times |\mathbb{P}_\vartheta(x_{k-1}|\overline{Y}^{k-1}_{-m}, X_{-m}=x) - \mathbb{P}_\vartheta(x_{k-1}|\overline{Y}^{k-1}_{-m})| \le \rho^{k+m-1}\sigma_+ G_{\vartheta k}/\sigma_-$, where the second inequality follows from Lemma 9. The second result follows from writing the left hand side as

$$
\frac{p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x) - \overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})}{p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x)} + \frac{\overline{p}_\vartheta(Y_k|\overline{Y}^{k-1}_{-m})}{\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m})} \frac{\overline{p}_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}) - p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x)}{p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m}, X_{-m}=x)},
$$

and using the derived bounds. The results with $p_\vartheta(Y_k|\overline{Y}^{k-1}_{-m'}, X_{-m'}=x)$ and $p_{\vartheta^*}(Y_k|\overline{Y}^{k-1}_{-m'}, X_{-m'}=x)$ are proven similarly. $\square$

The following result originally appeared in equations (59)–(60) of Kasahara and Shimotsu (2015). We state this as a lemma for ease of reference.

Lemma 5. *Let $f(\mu,\sigma^2)$ denote the density of $N(\mu,\sigma^2)$. Then*

$$\nabla_{\lambda_\mu^k} f(c_1\lambda_\mu, c_2\lambda_\mu^2)\Big|_{\lambda_\mu=0} = \begin{cases} c_1\nabla_\mu f(0,0) & \text{if } k=1, \\ c_1^2\nabla_{\mu^2} f(0,0) + 2c_2\nabla_{\sigma^2} f(0,0) & \text{if } k=2, \\ c_1^3\nabla_{\mu^3} f(0,0) + 6c_1c_2\nabla_{\mu\sigma^2} f(0,0) & \text{if } k=3, \\ c_1^4\nabla_{\mu^4} f(0,0) + 12c_1^2c_2\nabla_{\mu^2} f(0,0)\nabla_{\sigma^2} f(0,0) + 12c_2^2\nabla_{\sigma^4} f(0,0) & \text{if } k=4. \end{cases}$$

Proof of Lemma 5. Observe that a composite function $f(\lambda_\mu, h(\lambda_\mu))$ satisfies $\nabla_{\lambda_\mu^k} f(\lambda_\mu, h(\lambda_\mu)) = (\nabla_{\lambda_\mu} + \nabla_u)^k f(\lambda_\mu, h(u))|_{u=\lambda_\mu} = \sum_{j=0}^k \binom{k}{j} \nabla_{\lambda_\mu^{k-j} u^j} f(\lambda_\mu, h(u))|_{u=\lambda_\mu}$. Further, because $\nabla_{u^j} u^2|_{u=0} = 0$ except for $j=2$, it follows from Faà di Bruno's formula that

$$\nabla_{u^j} f(c_1\lambda_\mu, c_2u^2)|_{\lambda_\mu=u=0} = \begin{cases} 0 & \text{if } j=1,3, \\ 2c_2\nabla_h f(0,h(0)) & \text{if } j=2, \\ 12c_2^2\nabla_{h^2} f(0,h(0)) & \text{if } j=4, \end{cases}$$

and hence the stated result follows. □

Lemma 6. *Suppose the assumptions of Proposition 10 hold. Then, there exist a random variable $K(\vartheta)$ such that $\mathbb{E}_{\vartheta^*}\sup_{\vartheta\in\mathcal{N}^*}|K(\vartheta)|^2<\infty$ and, for all $k\geq 1$,*

$$(a)\quad \frac{\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \varrho K(\vartheta),$$

$$(b)\quad \frac{\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = b(\alpha)\frac{\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} + \varrho K(\vartheta).$$

Proof of Lemma 6. Part (a) holds if

$$\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0, \tag{96}$$

because (i) $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha\varrho}(Y_k|\overline{Y}_0^{k-1}) - \nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) = \nabla_\varrho\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha\bar\varrho}(Y_k|\overline{Y}_0^{k-1})\varrho$ for $\bar\varrho\in[0,\varrho]$ from the mean value theorem, (ii) $\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})$ does not depend on the value of ϱ, and (iii) $\nabla_\varrho\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha\bar\varrho}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\alpha\bar\varrho}(Y_k|\overline{Y}_0^{k-1}) \in L^2(\mathbb{P}_{\vartheta^*})$ uniformly in $\vartheta\in\mathcal{N}^*$ from Proposition 6(d).

We proceed to show (96). Let $\nabla^i \ell_t^* := \nabla_{\lambda_\mu^i} \log g_t^*$ with $\nabla \ell_t^* = \nabla^1 \ell_t^*$. Observe that

$$
\begin{aligned}
&\nabla_{\lambda_\mu^3} \log \overline{p}_{\psi^*\alpha 0}(Y_0^k|\overline{Y}_0) \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^3 \ell_t^* \middle| \overline{Y}_0^k\right] + 3\sum_{t_1=1}^k \sum_{t_2=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^2 \ell_{t_1}^* \nabla \ell_{t_2}^* \middle| \overline{Y}_0^k\right] \\
&\quad + \sum_{t_1=1}^k \sum_{t_2=1}^k \sum_{t_3=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \middle| \overline{Y}_0^k\right] \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^3 \ell_t^* + 3\nabla^2 \ell_t^* \nabla \ell_t^* + \nabla \ell_t^* \nabla \ell_t^* \nabla \ell_t^* \middle| \overline{Y}_0^k\right] \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla_{\lambda_\mu^3} g_t^* / g_t^* \middle| \overline{Y}_0^k\right] = 0,
\end{aligned}
\tag{97}
$$

where the first equality follows from Lemma 1, the second equality holds because (i) X_t is serially independent when $\varrho = 0$, (ii) $\nabla \ell_t^* = h_{1t} \nabla_\mu f_t^* / f_t^*$ and $\nabla^2 \ell_t^* = h_{2t} \nabla_\mu^2 f_t^* / f_t^* - (h_{1t} \nabla_\mu f_t^* / f_t^*)^2$, and (iii) $\mathbb{E}_{\psi^*\alpha 0}[h_{1t}|\overline{Y}_0^k] = \mathbb{E}_{\psi^*\alpha 0}[h_{2t}|\overline{Y}_0^k] = 0$ from (29), the third equality follows from (87), and the last equality follows from (29). (96) follows from (97) because (i) $\nabla_{\lambda_\mu^3} \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) / \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \nabla_{\lambda_\mu^3} \log \overline{p}_{\psi^*\pi}(Y_0^k|\overline{Y}_0) - \nabla_{\lambda_\mu^3} \log \overline{p}_{\psi^*\pi}(Y_0^{k-1}|\overline{Y}_0)$ from (87) and $\nabla_\lambda \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$. Therefore, part (a) is proven.

For part (b), from a similar argument to part (a), the stated result holds if

$$
\nabla_{\lambda_\mu^4} \overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) / \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = b(\alpha) \nabla_{\lambda_\sigma^2} \overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) / \overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}). \tag{98}
$$

Using Lemma 1, noting that X_t is serially independent when $\varrho = 0$, collecting the terms, and using (87), we obtain

$$
\begin{aligned}
&\nabla_{\lambda_\mu^4} \log \overline{p}_{\psi^*\alpha 0}(Y_0^k|\overline{Y}_0) \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^4 \ell_t^* \middle| \overline{Y}_0^k\right] + 4\sum_{t_1=1}^k \sum_{t_2=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^3 \ell_{t_1}^* \nabla \ell_{t_2}^* \middle| \overline{Y}_0^k\right] + 3\sum_{t_1=1}^k \sum_{t_2=1}^k \mathrm{cov}_{\psi^*\alpha 0}\left[\nabla^2 \ell_{t_1}^*, \nabla^2 \ell_{t_2}^* \middle| \overline{Y}_0^k\right] \\
&\quad + 6\sum_{t_1=1}^k \sum_{t_2=1}^k \sum_{t_3=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[(\nabla^2 \ell_{t_1}^* - E[\nabla^2 \ell_{t_1}^* |\overline{Y}_0^k]) \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \middle| \overline{Y}_0^k\right] \\
&\quad + \sum_{t_1=1}^k \sum_{t_2=1}^k \sum_{t_3=1}^k \sum_{t_4=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \nabla \ell_{t_4}^* \middle| \overline{Y}_0^k\right] - 3\left\{\sum_{t_1=1}^k \sum_{t_2=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \middle| \overline{Y}_0^k\right]\right\}^2 \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla^4 \ell_t^* + 4\nabla^3 \ell_t^* \nabla \ell_t^* + 3\left(\nabla^2 \ell_t^*\right)^2 + 6\nabla^2 \ell_t^* \nabla \ell_t^* \nabla \ell_t^* + (\nabla \ell_t^*)^4 \middle| \overline{Y}_0^k\right] \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*\alpha 0}\left[\nabla_{\lambda_\mu^4} g_t^* / g_t^* \middle| \overline{Y}_0^k\right].
\end{aligned}
\tag{99}
$$

(98) follows from (99) because (i) $\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) = \nabla_{\lambda_\mu^4}\log\overline{p}_{\psi^*\alpha 0}(Y_0^k|\overline{Y}_0) - \nabla_{\lambda_\mu^4}\log\overline{p}_{\psi^*\alpha 0}(Y_0^{k-1}|\overline{Y}_0)$ from (87), $\nabla_\lambda\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0$, and $\nabla_{\lambda_\mu^2}\log\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) = 0$, (ii) $\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1}) = \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2}g_k^*|\overline{Y}_0^k]$ from a similar argument to (17), and (iii) $\mathbb{E}_{\psi^*\alpha 0}[\nabla_{\lambda_\mu^4}g_t^*/g_t^*|\overline{Y}_0^k] = b(\alpha)\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2}g_k^*|\overline{Y}_0^k]$ from (29). Therefore, part (b) is proven.

□

Lemma 7. *Suppose the assumptions of Proposition 12 hold. Then, there exist a random variable $K(\vartheta)$ such that $\mathbb{E}_{\vartheta^*}\sup_{\vartheta\in\mathcal{N}^*}|K(\vartheta)|^2<\infty$ and, for all $k\geq 1$,*

$$(a)\quad \frac{\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \alpha(1-\alpha)(1-2\alpha)\frac{\nabla_{\mu^3}f_k^*}{f_k^*} + \varrho K(\vartheta),$$

$$(b)\quad \frac{\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})} = \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\frac{\nabla_{\mu^4}f_k^*}{f_k^*} + \varrho K(\vartheta).$$

Proof of Lemma 7. The proof is similar to the proof of Lemma 6(a). From an argument similar to the proof of Lemma 6, the stated results hold if

$$(A)\quad \nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \alpha(1-\alpha)(1-2\alpha)\nabla_{\mu^3}f_k^*/f_k^*,$$

$$(B)\quad \nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\alpha 0}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\nabla_{\mu^4}f_k^*/f_k^*.$$

Observe that $\nabla_{\lambda_\mu^3}\log\overline{p}_{\psi^*\alpha 0}(Y_0^k|\overline{Y}_0) = \sum_{t=1}^k\mathbb{E}_{\psi^*\alpha 0}[\nabla_{\lambda_\mu^3}g_t^*/g_t^*|\overline{Y}_0^k]$ in (97) and the equality (99) in the proof of Lemma 6 still hold under the assumptions of Proposition 12 if we use (43) in place of (29). Consequently, (A) and (B) follow from (42), (43), and the argument of the proof of Lemma 6, giving the stated result. □

8.3.3 Bounds on difference in state probabilities and conditional moments

Lemma 8. *Suppose $X_1,\ldots,X_n$ are random variables with $\max_{1\leq i\leq n}\mathbb{E}|X_i|^q<C$ for some $q>0$ and $C\in(0,\infty)$. Then, $\max_{1\leq i\leq n}|X_i| = o_p(n^{1/q})$.*

Proof of Lemma 8. For any $\varepsilon>0$, we have $\mathbb{P}(\max_{1\leq i\leq n}|X_i|>\varepsilon n^{1/q})\leq\sum_{1\leq i\leq n}\mathbb{P}(|X_i|>\varepsilon n^{1/q})$ $\leq\varepsilon^{-q}n^{-1}\sum_{1\leq i\leq n}\mathbb{E}(|X_i|^q\mathbb{I}\{|X_i|>\varepsilon n^{1/q}\})$ by a version of Markov inequality. As $n\to\infty$, the right hand side tends to 0 by the Dominated Convergence Theorem. □

The following Lemmas extend Corollary 1 and (39) of DMR and an equation on p. 2298 of DMR; DMR derive these results when $t_1=t_2$ and $t_3=t_4$ and W_{-m}^n is not present. For two probability measures μ_1 and μ_2, the total variation distance between μ_1 and μ_2 is defined as $||\mu_1-\mu_2||_{TV} := \sup_A|\mu_1(A)-\mu_2(A)|$. $||\cdot||_{TV}$ satisfies $\sup_{f(x):0\leq f(x)\leq 1}|\int f(x)d\mu_1(x)-\int f(x)d\mu_2(x)| = ||\mu_1-\mu_2||_{TV}$. In the following, x_m denotes "$X_{-m}=x_{-m}$."

Lemma 9. *Suppose Assumptions 1-2 hold and $\vartheta_x \in \Theta_x$. For all $-m \le t_1 \le t_2$ with $-m < n$ and all probability measures μ_1 and μ_2 on $\mathcal{B}(\mathcal{X})$,*

$$\left\| \sum_{x_{-m}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_1(x_{-m}) - \sum_{x\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_2(x_{-m}) \right\|_{TV}$$
$$\le \rho^{t_1+m}.$$

Proof of Lemma 9. We assume $t_1 > -m$ because the stated result holds trivially when $t_1 = -m$. We prove the lemma by showing that the stated bound holds for $X_{t_1}^{t_2} \in A$ for any $A \subseteq \mathcal{X}^{t_2-t_1}$. Observe that Lemma 1 of DMR still holds when W_{-m+1}^n is added to the conditioning variable because Assumption 1 implies that $\{(X_k, \overline{Y}_k)\}_{k=0}^\infty$ is a Markov chain given $\{W_k\}_{k=0}^\infty$. Therefore, $\{X_t\}_{t\ge -m}$ is a Markov chain when conditioned on $\{\overline{Y}_{-m}^n, W_{-m+1}^n\}$. Choose $B \in \mathcal{B}(\mathcal{X})$ so that $\mathbb{P}_{\vartheta_x}(X_{t_1} \in B | \overline{Y}_{-m}^n, W_{-m+1}^n) \ne \{0, 1\}$, then it follows from the Markov property of $\{X_t\}_{t\ge -m}$ that

$$\sum_{x_{-m}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_1(x_{-m}) - \sum_{x_{-m}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_2(x_{-m})$$
$$= \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | X_{t_1} \in B, \overline{Y}_{-m}^n, W_{-m+1}^n) \times \left[\sum_{x_{-m}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1} \in B | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_1(x_{-m}) \right.$$
$$\left. - \sum_{x_{-m}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1} \in B | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mu_2(x_{-m}) \right].$$

The stated result follows because $\sup_A \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A|\cdot) \le 1$ and the supremum of the term inside the brackets over $B \in \mathcal{B}(\mathcal{X})$ is bounded by ρ^{t_1+m} from Corollary 1 of DMR. □

Lemma 10. *Suppose Assumptions 1-2 hold and $\vartheta_x \in \Theta_x$. For all $-m \le t_1 \le t_2 \le n-1$,*

$$\left\| \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n) - \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot | \overline{Y}_{-m}^{n-1}, x_{-m}, W_{-m+1}^n) \right\|_{TV} \le \rho^{n-1-t_2}.$$

The same bound holds when x_{-m} is dropped from the conditioning variables.

Proof of Lemma 10. We show that the stated bound holds for any $A \subseteq \mathcal{X}^{t_2-t_1-1}$. Observe that the time-reversed process $\{Z_{n-k}\}_{0\le k\le n+m}$ is Markov when conditioned on W_{-m+1}^n. Consequently, for $k = n, n-1$, $\mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | X_{t_2}, \overline{Y}_{-m}^k, x_{-m}, W_{-m+1}^n) = \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | X_{t_2}, \overline{Y}_{-m}^{t_2}, x_{-m}, W_{-m+1}^n)$ holds. Choose $B \in \mathcal{B}(\mathcal{X})$ so that $\mathbb{P}_{\vartheta_x}(X_{t_2} \in B | \overline{Y}_{-m}^{t_2}, W_{-m+1}^n) \ne \{0, 1\}$, then it follows that

$$\begin{aligned}
&\mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n) - \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | \overline{Y}_{-m}^{n-1}, x_{-m}, W_{-m+1}^n) \\
&= \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A | X_{t_2} \in B, \overline{Y}_{-m}^{t_2}, x_{-m}, W_{-m+1}^n) \\
&\qquad \times \left[\mathbb{P}_{\vartheta_x}(X_{t_2} \in B | \overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n) - \mathbb{P}_{\vartheta_x}(X_{t_2} \in B | \overline{Y}_{-m}^{n-1}, x_{-m}, W_{-m+1}^n)) \right].
\end{aligned}$$

The supremum of the term inside the brackets over $B \in \mathcal{B}(\mathcal{X})$ is bounded by ρ^{n-1-t_2} because equation (39) of DMR p. 2294 holds when W_{-m+1}^n is added to the conditioning variables. Therefore, the stated bound holds. When x_{-m} is dropped from the conditioning variables, the stated result follow from a similar argument with using Lemma 9 and an analogue of Corollary 1 of DMR in place of equation (39) of DMR. □

Lemma 11. *Suppose Assumptions 1-2 hold and $\vartheta_x \in \Theta_x$. For all $-m \le t_1 \le t_2 < t_3 \le t_4$ with $-m < n$,*

$$\begin{aligned}&\big\|\mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot, X_{t_3}^{t_4} \in \cdot|\overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\\ &\quad -\mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in \cdot|\overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\mathbb{P}_{\vartheta_x}(X_{t_3}^{t_4} \in \cdot|\overline{Y}_{-m}^n, x_{-m}, W_{-m+1}^n)\big\|_{TV} \le \rho^{t_3-t_2}.\end{aligned}$$

The same bound holds when x_{-m} is dropped from the conditioning variables.

Proof of Lemma 11. We suppress the conditioning variable W_{-m+1}^n. We show that the stated bound holds for any $(A, B) \subseteq \mathcal{X}^{t_4-t_1+1}$. Observe that

$$\begin{aligned}&\mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A, X_{t_3}^{t_4} \in B|\overline{Y}_{-m}^n, x_{-m}) - \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A|\overline{Y}_{-m}^n, x_{-m})\mathbb{P}_{\vartheta_x}(X_{t_3}^{t_4} \in B|\overline{Y}_{-m}^n, x_{-m})\\ &= \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in A|\overline{Y}_{-m}^n, x_{-m})\left[\mathbb{P}_{\vartheta_x}(X_{t_3}^{t_4} \in B|\overline{Y}_{-m}^n, X_{t_1}^{t_2} \in A) - \mathbb{P}_{\vartheta_x}(X_{t_3}^{t_4} \in B|\overline{Y}_{-m}^n, x_{-m})\right].\end{aligned}$$

The term inside the brackets is bounded by $\rho^{t_3-t_2}$ from Lemma 9. The stated bound then follows immediately. □

The following corollaries are used for proving Lemma 2. Their proofs are omitted because they follow straightforwardly from the Lemma 9-11 and the fact that, for any two probability measures μ_1 and μ_2, $\sup_{f(x):\max_x |f(x)|\le 1} |\int f(x)d\mu_1(x) - \int f(x)d\mu_2(x)| = 2||\mu_1 - \mu_2||_{TV}$ (see, e.g., Levin et al. (2009, Proposition 4.5)).

Corollary 2. *Suppose that Assumptions 1-2 hold, $f(\overline{Y}_{-m}^n, X_{t_1}^{t_2}, W_{-m+1}^n; \vartheta) \in L^2(\mathbb{P}_{\vartheta^*})$, and $g(\overline{Y}_{-m}^n, X_{t_1}^{t_2}, W_{-m+1}^n; \vartheta) \in L^2(\mathbb{P}_{\vartheta^*})$. Define $||f||_\infty := \sup_{\vartheta\in\mathcal{N}^*} \max_{x_{t_1}^{t_2}} f_\vartheta(\overline{Y}_{-m}^n, x_{t_1}^{t_2}, W_{-m+1}^n)$, and define $||g||_\infty$ similarly. Let $f_{-m}^n(X_{t_1}^{t_2}; \vartheta) := f(\overline{Y}_{-m}^n, X_{t_1}^{t_2}, W_{-m+1}^n; \vartheta)$ and $g_{-m}^n(X_{t_1}^{t_2}; \vartheta) := g(\overline{Y}_{-m}^n, X_{t_1}^{t_2}, W_{-m+1}^n; \vartheta)$. Then, suppressing the conditioning variables W_{-m+1}^n and $W_{-m'+1}^n$ from the conditioning sets,*
(a) For all $-m \le t_1 \le t_2 < t_3 \le t_4$

$$\begin{aligned}&\left|\text{cov}_{\vartheta^*}(f_{-m}^n(X_{t_1}^{t_2}; \vartheta), g_{-m}^n(X_{t_3}^{t_4}; \vartheta)|\overline{Y}_{-m}^n, x_{-m})\right| \le 2\rho^{t_3-t_2}||f_{-m}^n||_\infty ||g_{-m}^n||_\infty,\\ &\left|\text{cov}_{\vartheta^*}(f_{-m}^n(X_{t_1}^{t_2}; \vartheta), g_{-m}^n(X_{t_3}^{t_4}; \vartheta)|\overline{Y}_{-m}^n)\right| \le 2\rho^{t_3-t_2}||f_{-m}^n||_\infty ||g_{-m}^n||_\infty.\end{aligned}$$

(b) For all $-m' \le -m \le t_1 \le t_2$,

$$\begin{aligned}&\left|\mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2}; \vartheta)|\overline{Y}_{-m}^n, x_{-m}] - \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2}; \vartheta)|\overline{Y}_{-m}^n]\right| \le 2\rho^{m+t_1}||f_{-m}^n||_\infty,\\ &\left|\mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2}; \vartheta)|\overline{Y}_{-m}^n, x_{-m}] - \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2}; \vartheta)|\overline{Y}_{-m'}^n, x_{-m'}]\right| \le 2\rho^{m+t_1}||f_{-m}^n||_\infty.\end{aligned}$$

(c) For all $-m \le t_1 \le t_2 \le n-1$,

$$\left| \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2};\vartheta)|\overline{Y}_{-m}^n, x_{-m}] - \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2};\vartheta)|\overline{Y}_{-m}^{n-1}, x_{-m}] \right| \le 2\rho^{n-1-t_2}||f_{-m}^n||_\infty,$$
$$\left| \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2};\vartheta)|\overline{Y}_{-m}^n] - \mathbb{E}_{\vartheta^*}[f_{-m}^n(X_{t_1}^{t_2};\vartheta)|\overline{Y}_{-m}^{n-1}] \right| \le 2\rho^{n-1-t_2}||f_{-m}^n||_\infty.$$

Lemma 12. *Suppose* X_k *is a stationary Markov process with the state space* $\mathcal{X} = \{1,2\}$ *and the transition matrix*

$$P = \begin{pmatrix} p_{11} & 1-p_{11} \\ 1-p_{22} & p_{22} \end{pmatrix}.$$

Let $q_k := \mathbb{I}\{X_k = 1\}$, $\alpha := \mathbb{E}[q_k]$, *and* $\varrho := p_{11} + p_{22} - 1$. *Then, (a)* $\mathbb{E}(q_k - \alpha)^2 = \alpha(1-\alpha)$, $\mathbb{E}(q_k-\alpha)^3 = \alpha(1-\alpha)(1-2\alpha)$, $\mathbb{E}(q_k-\alpha)^4 = \alpha(1-\alpha)(3\alpha^2 - 3\alpha + 1)$, *and* $corr(X_k, X_{k+\ell}) = \varrho^{|\ell|}$. *(b) For all* $k > m$ *and any probability measures* μ_1 *and* μ_2 *on* $\mathcal{B}(\mathcal{X})$,

$$\left\| \sum_{j\in\mathcal{X}} \mathbb{P}(X_k \in \cdot|X_m = j)\mu_1(j) - \sum_{j\in\mathcal{X}} \mathbb{P}(X_k \in \cdot|X_m = j)\mu_2(j) \right\|_{TV} \le \varrho^{k-m}.$$

(c) For all $t_1 \le t_2 \le \cdots \le t_k < t_{k+1} \le \cdots \le t_\ell$, $||\mathbb{P}(X_{t_1}^{t_k} \in \cdot, X_{t_{k+1}}^{t_\ell} \in \cdot) - \mathbb{P}(X_{t_1}^{t_k} \in \cdot)\mathbb{P}(X_{t_{k+1}}^{t_\ell} \in \cdot)||_{TV} \le \mathcal{C}\varrho^{t_{k+1}-t_k}$.

Proof of Lemma 12. For part (a), the first three results follow from the property of a Bernoulli random variable, and the last result holds because $X_k = 2 - q_k$ and Hamilton (1994, p. 684) shows that q_k follows an $AR(1)$ process with the autoregressive coefficient $p_{11} + p_{22} - 1$. For part (b), decompose P as

$$P = U\Omega U^{-1}, \quad U = (\iota\ R) = \begin{pmatrix} 1 & \frac{1-p_{11}}{2-p_{11}-p_{22}} \\ 1 & -\frac{1-p_{22}}{2-p_{11}-p_{22}} \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \pi \\ u' \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & \varrho \end{pmatrix},$$

where $\pi := (\mathbb{P}(X_k = 1), \mathbb{P}(X_k = 2)) = (1-p_{22}, 1-p_{11})/(2-p_{11}-p_{22})$ is a vector of the stationary distribution of X_k, and $u := (1,-1)'$ is an eigenvector of P associated with the eigenvalue ϱ. Let $v_1 = (\mathbb{P}(X_m = 1|\mu_1), \mathbb{P}(X_m = 2|\mu_1))'$ denote the probability mass function of X_m under μ_1, and define v_2 similarly. Let $e = (1,0)'$, then, in view of $v_1'\iota = v_2'\iota = 1$,

$$\begin{aligned}
&\sum_{j\in\mathcal{X}} \mathbb{P}(X_k = 1|X_m = j)\mu_1(j) - \sum_{j\in\mathcal{X}} \mathbb{P}(X_k = 1|X_m = j)\mu_2(j) \\
&= v_1'P^{k-m}e - v_2'P^{k-m}e = (v_1 - v_2)'(\iota\pi + R\varrho^{k-m}u')e = (v_1-v_2)'R\varrho^{k-m}.
\end{aligned}$$

Similarly, we obtain $\sum_{j\in\mathcal{X}} \mathbb{P}(X_k = 2|X_m = j)\mu_1(j) - \sum_{j\in\mathcal{X}} \mathbb{P}(X_k = 2|X_m = j)\mu_2(j) = -(v_1 - v_2)'R\varrho^{k-m}$, and the stated result follows because $\max_{v_1,v_2} |(v_1-v_2)'R| = 1$. Part (c) follows from repeating the proof of Lemmas 9 and 11 using part (b). □

8.3.4 The sums of powers of ρ

Lemma 13. *For all $\rho \in (0,1)$, $c \geq 1$, $q \geq 1$, and $b > a$,*

$$\sum_{t=-\infty}^{\infty} \left(\rho^{\lfloor (t-a)/cq \rfloor} \wedge \rho^{\lfloor (b-t)/q \rfloor}\right) \leq \frac{q(c+1)\rho^{\lfloor (b-a)/(c+1)q \rfloor}}{1-\rho},$$
$$\sum_{t=-\infty}^{\infty} \left(\rho^{\lfloor (t-a)/q \rfloor} \wedge \rho^{\lfloor (b-t)/cq \rfloor}\right) \leq \frac{q(c+1)\rho^{\lfloor (b-a)/(c+1)q \rfloor}}{1-\rho}.$$

Proof of Lemma 13. The first result holds because the left hand side is bounded by

$$\begin{aligned}
&\sum_{t=-\infty}^{\lfloor (a+bc)/(c+1) \rfloor} \rho^{\lfloor (b-t)/q \rfloor} + \sum_{t=\lfloor (a+bc)/(c+1) \rfloor + 1}^{\infty} \rho^{\lfloor (t-a)/cq \rfloor} \\
&\leq q\rho^{\lfloor \{b - \lfloor (a+bc)/(c+1) \rfloor\}/q \rfloor}/(1-\rho) + cq\rho^{\lfloor \{\lfloor (a+bc)/(c+1) \rfloor + 1 - a\}/cq \rfloor}/(1-\rho) \\
&\leq q(1+c)\rho^{\lfloor (b-a)/(c+1)q \rfloor}/(1-\rho).
\end{aligned}$$

The second result is proven by bounding the left hand side by $\sum_{t=-\infty}^{\lfloor (ac+b)/(c+1) \rfloor} \rho^{\lfloor (b-t)/cq \rfloor}$ $+ \sum_{t=\lfloor (ac+b)/(c+1) \rfloor + 1}^{\infty} \rho^{\lfloor (t-a)/q \rfloor}$ and proceeding similarly.

□

The following lemma generalizes the result in the last inequality on p. 2299 of DMR.

Lemma 14. *For all $\rho \in (0,1)$, $k \geq 1$, $q \geq 1$, and $n \geq 0$,*

$$\sum_{0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq n} \left(\rho^{\lfloor t_1/q \rfloor} \wedge \rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \cdots \wedge \rho^{\lfloor (t_k - t_{k-1})/q \rfloor} \wedge \rho^{\lfloor (n-t_k)/q \rfloor}\right) \leq C_{kq}(\rho)\rho^{\lfloor n/2kq \rfloor},$$

where $C_{kq}(\rho) := q^k k(k+1)!(1-\rho)^{-k}$.

Proof of Lemma 14. When $k = 1$, the stated result follows from Lemma 13 with $c = 1$. We first show that the following holds for $k \geq 2$:

$$\sum_{t_1 \leq t_2 \leq \cdots \leq t_k \leq n} \left(\rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \cdots \wedge \rho^{\lfloor (t_k - t_{k-1})/q \rfloor} \wedge \rho^{\lfloor (n-t_k)/q \rfloor}\right) \leq \frac{q^{k-1}(k+1)!\rho^{\lfloor (n-t_1)/kq \rfloor}}{(1-\rho)^{k-1}}. \quad (100)$$

We prove (100) by induction. When $k = 2$, it follows from Lemma 13 with $c = 1$ that $\sum_{t_2=t_1}^{n}(\rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \rho^{\lfloor (n-t_2)/q \rfloor}) \leq 2q\rho^{\lfloor (n-t_1)/2q \rfloor}/(1-\rho)$, giving (100). Suppose (100) holds when

$k=\ell$. Then (100) holds when $k=\ell+1$ because, from Lemma 13,

$$\begin{aligned}
&\sum_{t_1\le t_2\le\cdots\le t_\ell\le t_{\ell+1}\le n}\left(\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\rho^{\lfloor (t_3-t_2)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_{\ell+1}-t_\ell)/q\rfloor}\wedge\rho^{\lfloor (n-t_{\ell+1})/q\rfloor}\right)\\
&\le\sum_{t_2=t_1}^{n}\left(\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\sum_{t_2\le\cdots\le t_{\ell+1}\le n}\left(\rho^{\lfloor (t_3-t_2)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_{\ell+1}-t_\ell)/q\rfloor}\wedge\rho^{\lfloor (n-t_{\ell+1})/q\rfloor}\right)\right)\\
&\le\frac{q^{\ell-1}\ell!}{(1-\rho)^{\ell-1}}\sum_{t_2=t_1}^{n}\left(\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\rho^{\lfloor (n-t_2)/\ell q\rfloor}\right)\\
&\le\frac{q^{\ell}(\ell+1)!}{(1-\rho)^{\ell}}\rho^{\lfloor (n-t_1)/(\ell+1)q\rfloor},
\end{aligned}$$

and hence (100) holds for all $k\ge 2$. We proceed to show the stated result. Observe that

$$\begin{aligned}
&\sum_{0\le t_1\le t_2\le\cdots\le t_k\le n}\left(\rho^{\lfloor t_1/q\rfloor}\wedge\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right)\\
&\le\sum_{t_1=0}^{n/2}\sum_{t_1\le t_2\le\cdots\le t_{k-1}\le t_k}\sum_{t_k=t_1}^{n-t_1}\left(\rho^{\lfloor t_1/q\rfloor}\wedge\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right)\\
&\quad+\sum_{t_k=n/2}^{n}\sum_{t_1\le t_2\le\cdots\le t_{k-1}\le t_k}\sum_{t_1=n-t_k}^{t_k}\left(\rho^{\lfloor t_1/q\rfloor}\wedge\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right)\\
&=2\sum_{t_1=0}^{n/2}\sum_{t_1\le t_2\le\cdots\le t_{k-1}\le t_k}\sum_{t_k=t_1}^{n-t_1}\left(\rho^{\lfloor t_1/q\rfloor}\wedge\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right)\\
&=2\sum_{t_1=0}^{n/2}\sum_{t_1\le t_2\le\cdots\le t_{k-1}\le t_k}\sum_{t_k=t_1}^{n-t_1}\left(\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right)\\
&\le 2\sum_{t_1=0}^{n/2}\sum_{t_1\le t_2\le\cdots\le t_{k-1}\le t_k\le n}\left(\rho^{\lfloor (t_2-t_1)/q\rfloor}\wedge\cdots\wedge\rho^{\lfloor (t_k-t_{k-1})/q\rfloor}\wedge\rho^{\lfloor (n-t_k)/q\rfloor}\right),
\end{aligned}$$

where the first equality holds by symmetry, and the second equality follows from $n-t_k\ge t_1$. From (100), the right hand side is no larger than $q^{k-1}(k+1)!(1-\rho)^{(1-k)}\sum_{t_1=0}^{n/2}\rho^{\lfloor (n-t_1)/kq\rfloor}\le q^k k(k+1)!(1-\rho)^{-k}\rho^{\lfloor n/2kq\rfloor}$, giving the stated result. □

The next lemma generalizes equation (46) and p. 2294 of DMR, who derive a similar bound when $\ell=1,2$.

Lemma 15. *Let $a_j>0$ for all j. For all positive integer $\ell\ge 1$ and all $k\ge 1$ and $m\ge 0$, we have* $\max_{-m+1\le t_1,\ldots,t_\ell\le k}a_{t_1}\cdots a_{t_\ell}\le (k+m)^{\ell+1}A_\ell$, *where $A_\ell:=\sum_{t=-\infty}^{\infty}(|t|\vee 1)^{-2}a_t^\ell$.*

Proof of Lemma 15. When $\ell=1$, the stated result follows from $\max_{-m+1\le t\le k}a_t\le\sum_{t=-m+1}^{k}a_t=\sum_{t=-m+1}^{k}(|t|\vee 1)^2(|t|\vee 1)^{-2}a_t\le (k+m)^2\sum_{t=-\infty}^{\infty}(|t|\vee 1)^{-2}a_t$. When $\ell\ge 2$, from the Hölder's inequality, we have $\max_{-m+1\le t_1\le\ldots\le t_\ell\le k}a_{t_1}a_{t_2}\cdots a_{t_\ell}\le(\sum_{t=-m+1}^{k}a_t)^\ell=[\sum_{t=-m+1}^{k}(|t|\vee 1)^{2/\ell}(|t|\vee$

$1)^{-2/\ell}a_t]^\ell \leq [\sum_{t=-m+1}^k (|t| \vee 1)^{2/(\ell-1)}]^{(\ell-1)} \sum_{t=-m+1}^k (|t| \vee 1)^{-2} a_t^\ell \leq [(k+m)^{1+2/(\ell-1)}]^{\ell-1} A_\ell = (k+m)^{\ell+1} A_\ell$. □

The following lemma generalizes the bound derived on p. 2301 of DMR.

Lemma 16. *For $\alpha > 0$, $q > 0$, and $c_{jt} \geq 0$, define $c_{jq}^\infty(\rho^\alpha) := \sum_{t=-\infty}^{\infty} \rho^{\lfloor \alpha |t|/q \rfloor} c_{jt}$. For all $\rho \in (0,1)$, $k \geq 1$, and $0 \leq m \leq m'$,*

$$\sum_{t_1=-m'+1}^{-m} \sum_{t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5 \leq t_6 \leq k} \Big(\rho^{\lfloor (k-1-t_6)/q \rfloor} \wedge \rho^{\lfloor (t_6-t_5)/q \rfloor} \wedge \rho^{\lfloor (t_5-t_4)/q \rfloor} \wedge \rho^{\lfloor (t_4-t_3)/q \rfloor} \wedge \rho^{\lfloor (t_3-t_2)/q \rfloor} \wedge \rho^{\lfloor (t_2-t_1)/q \rfloor} \Big) \prod_{j=1}^{6} c_{jt_j} \leq \rho^{\lfloor (k-1+m)/2qa_7 \rfloor} c_{1q}^\infty \left(\rho^{1/2a_7} \right) \prod_{j=2}^{6} c_{jq}^\infty \left(\rho^{1/4a_j} \right), \tag{101}$$

where (a_j, b_j) are defined recursively with $(a_2, b_2) = (1,1)$ and, for $j \geq 3$,

$$a_{j+1} = \frac{4a_j(a_j + b_j)}{2a_j - 1}, \quad b_{j+1} = \frac{a_j(4b_j - 1)}{2a_j - 1}.$$

$a_j, b_j \geq 3/2$ for all j. Direct calculations using Matlab produce $a_7 \doteq 334.5406$.

Proof of Lemma 16. First, observe that the following result holds for $a, b > 1/4$, $t_1 \leq 0$, and $t_j, t_{j+1} \geq t_1$:

$$\begin{aligned} &(a) \text{ if } t_j \leq \frac{at_{j+1} + t_1}{a+b}, \quad \text{then } \frac{|t_j|}{4a} \leq \frac{a(4a+1)t_{j+1} + (2a-1)t_1}{4a(a+b)} - t_j, \\ &(b) \text{ if } t_j \geq \frac{at_{j+1} + t_1}{a+b}, \quad \text{then } \frac{|t_j|}{4a} \leq \frac{b}{a} t_j - \frac{a(4b-1)t_{j+1} + (2a+4b+1)t_1}{4a(a+b)}. \end{aligned} \tag{102}$$

(a) holds because (i) when $t_j \leq 0$, we have $t_j \leq (at_{j+1}+t_1)/(a+b) \Rightarrow (4a-1)t_j/4a \leq [a(4a-1)t_{j+1} + (4a-1)t_1]/4a(a+b) \Rightarrow -t_j/4a \leq [a(4a-1)t_{j+1} + (4a-1)t_1]/4a(a+b) - t_j$ and $a(4a-1)t_{j+1} + (4a-1)t_1 \leq a(4a-1)t_{j+1} + (4a-1)t_1 + 2a(t_{j+1} - t_1) = a(4a+1)t_{j+1} + (2a-1)t_1$; (ii) when $t_j \geq 0$, we have $t_j \leq (at_{j+1}+t_1)/(a+b) \Rightarrow (4a+1)t_j/4a \leq [a(4a+1)t_{j+1} + (4a+1)t_1]/4a(a+b) \Rightarrow t_j/4a \leq [a(4a+1)t_{j+1} + (4a+1)t_1]/4a(a+b) - t_j$ and $(4a+1)t_1 \leq (2a-1)t_1$.

(b) holds because (i) when $t_j \leq 0$, we have $t_j \geq (at_{j+1}+t_1)/(a+b) \Rightarrow (4b+1)t_j/4a \geq [a(4b+1)t_{j+1} + (4b+1)t_1]/4a(a+b) \Rightarrow -t_j/4a \leq bt_j/a - [a(4b+1)t_{j+1} + (4b+1)t_1]/4a(a+b)$ and $a(4b+1)t_{j+1} + (4b+1)t_1 \geq a(4b+1)t_{j+1} + (4b+1)t_1 - 2a(t_{j+1} - t_1) = a(4b-1)t_{j+1} + (2a+4b+1)t_1$; (ii) when $t_j \geq 0$, we have $t_j \geq (at_{j+1}+t_1)/(a+b) \Rightarrow (4b-1)t_j/4a \geq [a(4b-1)t_{j+1} + (4b-1)t_1]/4a(a+b) \Rightarrow t_j/4a \leq bt_j/a - [a(4b-1)t_{j+1} + (4b-1)t_1]/4a(a+b)$ and $a(4b-1)t_{j+1} + (4b-1)t_1 \geq a(4b-1)t_{j+1} + (2a+4b+1)t_1$.

We proceed to derive the stated bound. It follows from (a) and (b) and $\lfloor x+y \rfloor \geq \lfloor x \rfloor + \lfloor y \rfloor$

that, with $\bar{t}_j = (a_j t_{j+1} + t_1)/(a_j + b_j)$,

$$
\begin{aligned}
&\sum_{t_j=-m'+1}^{k} \left(\rho^{\lfloor (t_{j+1}-t_j)/q \rfloor} \wedge \rho^{\lfloor (b_j t_j - t_1)/a_j q \rfloor} \right) c_{jt_j} \\
&\leq \rho^{\lfloor \frac{a_j(4b_j-1)t_{j+1}-(2a_j-1)t_1}{4a_j(a_j+b_j)q} \rfloor} \left(\sum_{t_j \leq \bar{t}_j} \rho^{\lfloor \frac{a_j(4a_j+1)t_{j+1}+(2a_j-1)t_1}{4a_j(a_j+b_j)q} - \frac{t_j}{q} \rfloor} + \sum_{t_j \geq \bar{t}_j} \rho^{\lfloor \frac{b_j}{a_j q} t_j - \frac{a_j(4b_j-1)t_{j+1}+(2a_j+4b_j+1)t_1}{4a_j(a_j+b_j)q} \rfloor} \right) c_{jt_j} \\
&\leq \rho^{\lfloor \frac{a_j(4b_j-1)t_{j+1}-(2a_j-1)t_1}{4a_j(a_j+b_j)q} \rfloor} c_{jq}^{\infty} \left(\rho^{1/4a_j} \right) \\
&= \rho^{\lfloor \frac{b_{j+1}t_{j+1}-t_1}{a_{j+1}} \rfloor} c_{jq}^{\infty} \left(\rho^{1/4a_j} \right).
\end{aligned} \tag{103}
$$

Observe that $a_{j+1} \geq 2a_j \geq 2$ and $b_{j+1} \geq 2b_j - (1/2) \geq 3/2$ for all $j \geq 2$. Therefore, we can apply (102) and (103) to the left hand side of (101) sequentially for $j = 2, 3, \ldots, 6$. Consequently, the left hand side of (101) is no larger than

$$
\sum_{t_1=-m'+1}^{-m} \rho^{\lfloor \frac{b_7(k-1)-t_1}{a_7 q} \rfloor} c_{1t_1} \prod_{j=2}^{6} c_{jq}^{\infty} \left(\rho^{1/4a_j} \right).
$$

Observe that $|t_1| \leq k - 1 - 2t_1 - m$ because $t_1 \leq -m \Rightarrow -t_1 \leq -2t_1 - m \leq k - 1 - 2t_1 - m$. From $b_7(k-1) \geq k-1$ and $|t_1| \leq k - 1 - 2t_1 - m$, the sum is bounded by

$$
\sum_{t_1=-m'+1}^{-m} \rho^{\lfloor \frac{k-1-t_1}{a_7 q} \rfloor} c_{1t_1} = \rho^{\lfloor \frac{k-1+m}{2a_7 q} \rfloor} \sum_{t_1=-m'+1}^{-m} \rho^{\lfloor \frac{k-1-2t_1-m}{2a_7 q} \rfloor} c_{1t_1} \leq \rho^{\lfloor \frac{k-1+m}{2a_7 q} \rfloor} c_{1q}^{\infty} \left(\rho^{1/2a_7} \right),
$$

and the stated result follows. □

Lemma 17. *When the regime h and $h+1$ are combined into one regime in an M_0+1-regime model with $\vartheta_{M_0+1,x}$, the transition probability of X_k equals the transition probability of X_k under $\vartheta^*_{M_0,x}$ if and only if $\vartheta_{M_0+1,x} \in \Theta^*_{ph} \times \Pi^*_h$.*

Proof of Lemma 17. The "only if" part is trivial. To prove the "if" part, let $h = M_0$ without loss of generality. Let $X_k \sim \mathbb{P}_{\vartheta_{M_0+1}}$ and $\tilde{X}_k \sim \mathbb{P}_{\vartheta^*_{M_0}}$. The stated result follows if we show that, for any $\vartheta_{M_0+1} \in \Upsilon^*_{M_0}$, (a) $\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j | X_k \in J_{M_0}) = \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_{k+1} = j | \tilde{X}_k = M_0)$ for $j \leq M_0 - 1$ and (b) $\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} \in J_{M_0} | X_k \in J_{M_0}) = \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_{k+1} = M_0 | \tilde{X}_k = M_0)$. For (a), observe that

$$
\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j | X_k \in J_{M_0}) = \frac{\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j) - \mathbb{P}_{\vartheta_{M_0+1}}(\{X_{k+1} = j\} \cap \{X_k \in \bar{J}_{M_0}\})}{1 - \mathbb{P}_{\vartheta_{M_0+1}}(X_k \in \bar{J}_{M_0})}. \tag{104}
$$

If $p_{ij} = p^*_{ij}$ and $p_j = p^*_j$ for all $i, j \leq M_0 - 1$, then we have, for any $j \leq M_0 - 1$, $\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} =$

$j) = \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_{k+1} = j)$, $\mathbb{P}_{\vartheta_{M_0+1}}(X_k \in \bar{J}_{M_0}) = \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_k \leq M_0 - 1)$, and

$$\begin{aligned}
\mathbb{P}_{\vartheta_{M_0+1}}(\{X_{k+1} = j\} \cap \{X_k \in \bar{J}_{M_0}\}) &= \sum_{i \leq M_0 - 1} \mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j | X_k = i)\mathbb{P}_{\vartheta_{M_0+1}}(X_k = i) \\
&= \sum_{i \leq M_0 - 1} \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_{k+1} = j | \tilde{X}_k = i)\mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_k = i) \\
&= \mathbb{P}_{\vartheta^*_{M_0}}(\{\tilde{X}_{k+1} = j\} \cap \{\tilde{X}_k \leq M_0 - 1\}).
\end{aligned}$$

Substituting them to (104) gives $\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j | X_k \in J_{M_0}) = \mathbb{P}_{\vartheta^*_{M_0}}(\tilde{X}_{k+1} = j | \tilde{X}_k = M_0)$, and (a) follows. Because $\mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} \in \bar{J}_{M_0} | X_k \in J_{M_0}) = 1 - \sum_{j \leq M_0 - 1} \mathbb{P}_{\vartheta_{M_0+1}}(X_{k+1} = j | X_k \in J_{M_0})$, (b) follows from (a), and the stated result follows. □

References

Andrews, D. W. K. (1999), "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

— (2001), "Testing when a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

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

Carrasco, M., Hu, L., and Ploberger, W. (2014), "Optimal Test for Markov Switching Parameters," *Econometrica*, 82, 765–784.

Chen, J. and Li, P. (2009), "Hypothesis Test for Normal Mixture Models: The EM Approach," *Annals of Statistics*, 37, 2523–2542.

Chen, J., Li, P., and Fu, Y. (2012), "Inference on the Order of a Normal Mixture," *Journal of the American Statistical Association*, 107, 1096–1105.

Cho, J. S. and White, H. (2007), "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

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.

Dannemann, J. and Holtzmann, H. (2008), "Testing for two states in a hidden Markov model," *Canadian Journal of Statistics*, 36, 505–520.

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.

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.

Garcia, R. (1998), "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models," *International Economic Review*, 39.

Gassiat, E. and Keribin, C. (2000), "The Likelihood Ratio Test for the Number of Components in a Mixture with Markov Regime," *ESAIM: Probability and Statistics*, 4, 25–52.

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. (1994), *Time Series Analysis*, Princeton University Press.

Hansen, B. (1992), "The Likelihood Ratio Test Under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, 61–82.

— (1996), "Stochastic Equicontinuity for Unbounded Dependent Heterogeneous Arrays," *Econometric Theory*, 12, 347–359.

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.

Ho, N. and Nguyen, X. (2016), "Convergence rates of parameter estimation for some weakly identifiable finite mixtures," *Annals of Statistics*, forthcoming.

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.

Kasahara, H. and Shimotsu, K. (2015), "Testing the Number of Components in Normal Mixture Regression Models," *Journal of the American Statistical Association*, 110, 1632–1645.

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

Liu, X. and Shao, Y. (2003), "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *Annals of Statistics*, 31, 807–832.

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.

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.

Pollard, D. (1990), *Empirical Processes: Theory and Applications*, vol. 2 of *CBMS Conference Series in Probability and Statistics*, Hayward, CA: Institute of Mathematical Statistics.

Qu, Z. and Fan, Z. (2015), "Likelihood Ratio Based Tests for Markov Regime Switching," Preprint, Boston Unversity.

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

Woodbury, M. A. (1971), "Discussion of paper by Hartley and Hocking," *Biometrics*, 27, 808–817.