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

October 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 (see, e.g., Hamilton, 1989; 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 true parameter being on the boundary of the parameter space, and the degeneracy of Fisher information matrix. Testing the number of regimes for Markov regime switching models with normal density, which are popular in empirical applications, poses a further difficulty 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.

*This research is support by the Natural Science and Engineering Research Council of Canada, JSPS Grant-in-Aid for Scientific Research (C) No. 17K03653, and the Institute of Statistical Mathematics Cooperative Use Registration (2017 ISM CUR-171). The authors thank Marine Carrasco, Liang Hu, and Werner Ploberger for making their code available.

The issue of non-identifiability under the null hypothesis and the degeneracy in Fisher information matrix has been well recognized in the existing literature. In testing the null hypothesis of no regime switching, 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 constancy in general dynamic models 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 a closely related problem of testing the number of components in finite mixture normal regression models, Kasahara and Shimotsu (2015) show 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). Chesher (1984) and Lee and Chesher (1986) investigate the related problem of testing for neglected heterogeneity under iid setting.

Cho and White (2007) derive the asymptotic distribution of the quasi-likelihood ratio test statistic (Q-LRTS) for testing single regime against two regimes by rewriting the model as a two-component mixture models, thereby ignoring the temporal dependence of the regimes. Qu and Zhuo (2017) 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 under some restriction on transition probabilities of latent regimes. Both Cho and White (2007) and Qu and Zhuo (2017) 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 M_0 regimes with $M_0 \geq 2$ 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 Le Cam's differentiable in quadratic mean (DQM) expansion that expands the 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 and heterogeneous 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.

We consider the conditional likelihood given an arbitrary distribution of the initial unobserved regime and show that the asymptotic distribution of the LRTS does not depend on the initial distribution. This approach follows Douc et al. (2004) [DMR, hereafter], who derive the asymptotic distribution of the MLE of regime-switching models 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. We then show that these derivatives of period density-ratios can be approximated by a stationary, ergodic and square integrable martingale difference sequence by conditioning on the infinite past, and this approximation is shown to satisfy the regularity conditions for our DQM expansion.

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 degree of deficiency of the Fisher information matrix and 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 is characterized by the maximum of M_0 random variables, each of which represents the LRTS for testing each of M_0 subsets.

We also derive the asymptotic distribution of our LRTS under local alternatives. We show that the value of the unidentified parameter affects the convergence rate of contiguous local alternatives. When the regime-specific density is normal, some contiguous local alternatives are of the order $n^{-1/8}$, and our LRT is shown to have nontrivial power against them. The test of Carrasco et al. (2014) do not have power against such alternatives, whereas the test of Qu and Zhuo (2017) rules out such alternatives. Simulations show that our bootstrap LRT has good finite sample properties.

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 under H_0. Section 8 derives the asymptotic distribution under local alternatives. Section 9 establishes the consistency of parametric bootstrap. Section 10 reports the results from simulations and an empirical application using the U.S. GDP per capita quarterly growth rate data. Section 11 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 a $k \times 1$ vector $a = (a_1, \ldots, a_k)'$ and a function $f(a)$, let $\nabla_a f(a) := (\partial f(a)/\partial a_1, \ldots, \partial f(a)/\partial a_k)'$, and let $\nabla_a^j f(a)$ denote a collection of derivatives of the form $(\partial^j/\partial a_{i_1}\partial a_{i_2} \ldots \partial a_{i_j})f(a)$. 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 $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Let $\lfloor x \rfloor$ denote the largest integer less than or equal to x, and define $(x)_+ := \max\{x, 0\}$. Given a sequence $\{f_k\}_{k=1}^n$, let $\nu_n(f_k) := n^{-1/2}\sum_{k=1}^n [f_k - \mathbb{E}_{\vartheta^*}(f_k)]$. For a sequence $X_{n\varepsilon}$ that is indexed by $n = 1, 2, \ldots$ and ε, we write $X_{n\varepsilon} = O_{p\varepsilon}(a_n)$ if, for any $\delta > 0$, there exist $\varepsilon > 0$ and $M, n_0 < \infty$ such that $\mathbb{P}(|X_{n\varepsilon}/a_n| \leq M) \geq 1 - \delta$ for all $n > n_0$, and we write $X_{n\varepsilon} = o_{p\varepsilon}(a_n)$ if, for any $\delta_1, \delta_2 > 0$, there exist $\varepsilon > 0$ and n_0 such that $\mathbb{P}(|X_{n\varepsilon}/a_n| \leq \delta_1) \geq 1 - \delta_2$ for all $n > n_0$. Loosely speaking, $X_{n\varepsilon} = O_{p\varepsilon}(a_n)$ and $X_{n\varepsilon} = o_{p\varepsilon}(a_n)$ mean that $X_{n\varepsilon} = O_p(a_n)$ and $X_{n\varepsilon} = o_p(a_n)$ when ε is sufficiently small, respectively. All limits are taken as $n \to \infty$ unless stated otherwise. 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 $\mathcal{Y} \subset \mathbb{R}^{q_y}$ and $\mathcal{W} \subset \mathbb{R}^{q_w}$, and let $\mathcal{B}(\mathcal{X}_M \times \mathcal{Y} \times \mathcal{W})$ denote the associated Borel σ-field. For a stochastic process $\{Z_k\}$ and $a < b$, define $\mathbf{Z}_a^b := (Z_a, Z_{a+1}, \ldots, Z_b)$. Denote $\overline{\mathbf{Y}}_{k-1} := (Y_{k-1}, \ldots, Y_{k-s})$ for a fixed integer s and $\overline{\mathbf{Y}}_a^b := (\overline{\mathbf{Y}}_a, \overline{\mathbf{Y}}_{a+1}, \ldots, \overline{\mathbf{Y}}_b)$.

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* $(\mathbf{X}_0^{k-2}, \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^\infty)$ *given* X_{k-1}*. (c) For each* $k \geq 1$*,* Y_k *is conditionally independent of* $(\mathbf{Y}_{-s+1}^{k-s-1}, \mathbf{X}_0^{k-1}, \mathbf{W}_0^{k-1}, \mathbf{W}_{k+1}^\infty)$ *given* $(\overline{\mathbf{Y}}_{k-1}, X_k, W_k)$*. (d)* $\mathbf{W}_1^\infty$ *is conditionally independent of* $(\overline{\mathbf{Y}}_0, X_0)$ *given* W_0*.*[1] *(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 $\vartheta_M = (\vartheta'_{M,y}, \vartheta'_{M,x})'$, 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 $(\overline{\mathbf{Y}}_{k-1}, X_k, W_k)$ by $g_{\vartheta_{M,y}}(y_k|\overline{\mathbf{y}}_{k-1}, x_k, w_k) = \sum_{j \in \mathcal{X}_M} \mathbb{I}\{x_k = j\} f(y_k|\overline{\mathbf{y}}_{k-1}, w_k; \gamma, \theta_j)$ so that $f(y_k|\overline{\mathbf{y}}_{k-1}, w_k; \gamma, \theta_j)$ is the conditional density of y_k given $(\overline{\mathbf{y}}_{k-1}, w_k)$ 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 varies

[1] Assumption 1(a)–(d) imply that W_k is conditionally independent of $(\mathbf{X}_0^{k-1}, \overline{\mathbf{Y}}_0^{k-1})$ given $\mathbf{W}_0^{k-1}$.

across regimes. Let

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

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

We make the following assumptions that correspond to (A1)–(A3) in 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* $w \in \mathcal{W}$, $0 < \inf_{\vartheta_y \in \Theta_{M,y}} \sum_{x \in \mathcal{X}_M} g_{\vartheta_{M,y}}(y'|\overline{y}, x, w)$ *and* $\sup_{\vartheta_{M,y} \in \Theta_{M,y}} \sum_{x \in \mathcal{X}_M} g_{\vartheta_{M,y}}(y'|\overline{y}, x, w) < \infty$. *(c)* $b_+ := \sup_{\vartheta_{M,y} \in \Theta_y} \sup_{\overline{\mathbf{y}}_0, y_1, w, x} g_{\vartheta_{M,y}}(y_1|\overline{\mathbf{y}}_0, x, w) < \infty$ *and* $\mathbb{E}_{\vartheta^*}(|\log b_-(\overline{\mathbf{Y}}_0^1, W_1)|) < \infty$, *where* $b_-(\overline{\mathbf{y}}_0^1, w_1) := \inf_{\vartheta_{M,y} \in \Theta_{M,y}} \sum_{x \in \mathcal{X}_M} g_{\vartheta_{M,y}}(y_1|\overline{\mathbf{y}}_0, x, w_1)$.

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 $\theta_{M,x} \in \Theta_{M,x}$.[2] 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 $\{Z_k\}_{k=0}^{\infty} := \{(X_k, \overline{\mathbf{Y}}_k)\}_{k=0}^{\infty}$ 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 $(\mathbf{Z}_0^{k-2}, \mathbf{W}_0^{k-1}, \mathbf{W}_{k+1}^{\infty})$ given (Z_{k-1}, W_k). Consequently, Lemma 1, Corollary 1, and Lemma 9 of DMR 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. We denote the probability measure and the associated expectation of $\{(Z_k, W_k)\}_{k=\infty}^{\infty}$ under stationarity by $\mathbb{P}_{\vartheta}$ and $\mathbb{E}_{\vartheta}$, respectively.[3]

Under Assumption 1(a)-(d), the density function of $\mathbf{Y}_1^n$ given $X_0 = x_0$, $\overline{\mathbf{Y}}_0$ and $\mathbf{W}_0^n$ for the model with M regimes is given by

$$p_{\vartheta_M}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) = \sum_{\mathbf{x}_1^n \in \mathcal{X}_M^n} \prod_{k=1}^{n} p_{\vartheta_M}(Y_k, x_k|\overline{\mathbf{Y}}_{k-1}, x_{k-1}, W_k). \tag{1}$$

[2] Assumptions 1(c) and 2(a) are also employed in DMR. As discussed in Kasahara and Shimotsu (2017), these assumptions together rule out models in which the conditional density Y_k depends on both current and lagged regimes. For example, if we specify $X_k = (\widetilde{X}_k, \widetilde{X}_{k-1})$ with $\widetilde{X}_k$ being a first-order Markov process, then the transition density of X_k inevitably has zeros. Kasahara and Shimotsu (2017) show asymptotic normality of the MLE while relaxing Assumption 2(a) to allow for $\inf_{\vartheta_{M,x} \in \Theta_{M,x}} \min_{x,x' \in \mathcal{X}_M} q_{\vartheta_{M,x}}(x, x') = 0$. It is possible to derive the asymptotic distribution of the LRT under similar assumptions to Kasahara and Shimotsu (2017), albeit with a tedious derivation.

[3] 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 $\{(Z_k, W_k)\}_{k=\infty}^{\infty}$ is stationary, we use notations such as $\mathbb{P}_{\vartheta}$ and $\mathbb{E}_{\vartheta}$ without an overline for simplicity.

Define the conditional log-likelihood function and stationary log-likelihood function as

$$\ell_n(\vartheta, x_0) := \log p_\vartheta(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) = \sum_{k=1}^n \log p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0),$$

$$\ell_n(\vartheta) := \log p_\vartheta(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, \mathbf{W}_0^n) = \sum_{k=1}^n \log p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k),$$

where we use the fact that $p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^n, x_0) = p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k, x_0)$ and $p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^n) = p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)$, which follows from Assumptions 1. Note that

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

and $\mathbb{P}_\vartheta(x_{k-1} | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1}) = \sum_{x_0 \in \mathcal{X}} \mathbb{P}_\vartheta(x_{k-1} | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1}, x_0) \mathbb{P}_\vartheta(x_0 | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1})$. Let $\rho := 1 - \sigma_- / \sigma_+ \in [0, 1)$. Lemma 10(a) in the appendix shows that, for all probability measures μ_1 and μ_2 on $\mathcal{B}(\mathcal{X})$ and all $(\overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1})$,

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

Consequently, $p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1}, x_0) - p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^{k-1})$ 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 $x_0 \in \mathcal{X}$,*

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

As discussed on p. 2263 of DMR, the stationary density $p_\vartheta(Y_k | \overline{\mathbf{Y}}_0^{k-1}, \mathbf{W}_0^k)$ 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

$$\ell_n(\vartheta_M, \xi_M) := \log \left(\sum_{x_0=1}^M p_{\vartheta_M}(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0) \xi_M(x_0) \right), \tag{3}$$

where $p_{\vartheta_M}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)$ 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. The null hypothesis can be written as $H_0 : \theta_1^* = \theta_2^*$.[4] When $\theta_1 = \theta_2$, the parameter $\vartheta_{2,x}$ is not identified 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 heteroscedastic 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 heteroscedastic 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 $\mathcal{N}_\varepsilon$.

[4] 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).

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

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

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 recall $\nu_n(f_k):=n^{-1/2}\sum_{k=1}^n[f_k-\mathbb{E}_{\vartheta^*}(f_k)]$.

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 t_ϑ **satisfies** $\psi\to\psi^*$ **if** $t_\vartheta\to 0$~~$t_\vartheta=0$ if and only if $\psi=\psi^*$~~ and $(s_{\pi k},r_{\vartheta k},u_{\vartheta k x_0})$ satisfy, for some $C\in(0,\infty)$, $\delta>0$, $\varepsilon>0$, and $\rho\in(0,1)$, (a) $\mathbb{E}_{\vartheta^*}\sup_{\pi\in\Theta_\pi}|s_{\pi k}|^{2+\delta}<C$, (b) $\sup_{\pi\in\Theta_\pi}|P_n(s_{\pi k}s_{\pi k}')-\mathcal{I}_\pi|=o_p(1)$ with $0<\inf_{\pi\in\Theta_\pi}\lambda_{\min}(\mathcal{I}_\pi)\le\sup_{\pi\in\Theta_\pi}\lambda_{\max}(\mathcal{I}_\pi)<\infty$, (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) $\sup_{x_0\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}_\varepsilon}P_n(|u_{\vartheta k x_0}|/|\psi-\psi^*|)^j=O_p(n^{-1})$ for $j=1,2,3$, (f) $\sup_{x_0\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}_\varepsilon}P_n(|s_{\pi k}||u_{\vartheta k x_0}|/|\psi-\psi^*|)=O_p(n^{-1})$, (g) $\sup_{\vartheta\in\mathcal{N}_\varepsilon}|\nu_n(s_{\pi k})|=O_p(1)$.*

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 of $\ell_n(\psi,\pi,x_0)$ in a neighborhood $\mathcal{N}_{c/\sqrt{n}}$ for any $c>0$.

Proposition 2. *Suppose that Assumption 3(a)–(f) holds. Then, for all $c>0$,*

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

The next proposition expands $\ell_n(\psi,\pi,x_0)$ in $A_{n\varepsilon}(x_0):=\{\vartheta\in\mathcal{N}_\varepsilon:\ell_n(\psi,\pi,x_0)-\ell_n(\psi^*,\pi,x_0)\ge 0\}$. This proposition is useful for deriving the asymptotic distribution of the LRTS because a consistent MLE is in $A_{n\varepsilon}(x_0)$ by definition, and it is difficult to find a uniform approximation of $\ell_n(\psi,\pi,x_0)$ in $\mathcal{N}_\varepsilon$. Let $A_{n\varepsilon c}(x_0):=A_{n\varepsilon}(x_0)\cup\mathcal{N}_{c/\sqrt{n}}$.

Proposition 3. *Suppose that Assumption 3 holds. Then, (a) $\sup_{x_0\in\mathcal{X}}\sup_{\vartheta\in A_{n\varepsilon}(x_0)}|t_\vartheta|=$*

$O_{p\varepsilon}(n^{-1/2})$, *and (b) for all* $c > 0$,

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

The following corollary of Proposition 2 and 3 shows 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 $A_{n\varepsilon}(\xi) := \{\vartheta \in \mathcal{N}_\varepsilon : \ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi) \geq 0\}$ and $A_{n\varepsilon_n c}(\xi) := A_{n\varepsilon_n}(\xi) \cup \mathcal{N}_{c/\sqrt{n}}$, 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,* $\sup_{\xi\in\Xi}\sup_{\vartheta\in A_{n\varepsilon}(\xi)} |t_\vartheta| = O_{p\varepsilon}(n^{-1/2})$ *and, for all* $c > 0$, $\sup_{\xi\in\Xi}\sup_{\vartheta\in A_{n\varepsilon c}(\xi)} |\ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi) - \sqrt{n}t_\vartheta'\nu_n(s_{\pi k}) + n t_\vartheta' \mathcal{I}_\pi t_\vartheta/2| = o_{p\varepsilon}(1)$.

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) of the density ratio $l_{\vartheta k x_0}$ 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 $\overline{\mathbf{Y}}_{-\infty}^{k-1}$ in place of $\overline{\mathbf{Y}}_0^{k-1}$. 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 satisfies the assumptions on the remainder terms $r_{\vartheta k}$ and $u_{\vartheta k x_0}$.

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

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

We use a short-handed notation $\phi_{\vartheta k}^i := \phi^i(\vartheta, \overline{\mathbf{Z}}_{k-1}^k)$. 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

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

For example, $\mathbb{E}^c_\vartheta \left[\phi_{\vartheta k_1}\phi_{\vartheta k_2}\phi_{\vartheta k_3}|\mathcal{F}\right] := \mathbb{E}_\vartheta \left[(\phi_{\vartheta k_1} - \mathbb{E}_\vartheta[\phi_{\vartheta k_1}|\mathcal{F}]) (\phi_{\vartheta k_2} - \mathbb{E}_\vartheta[\phi_{\vartheta k_2}|\mathcal{F}]) (\phi_{\vartheta k_3} - \mathbb{E}_\vartheta[\phi_{\vartheta k_3}|\mathcal{F}])|\mathcal{F}\right]$.

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

$$\begin{aligned}
&\Phi^{\mathcal{I}(1)}_{\vartheta\mathcal{T}(1)}[\mathcal{F}] := \mathbb{E}_\vartheta \left[\phi^{i_1}_{\vartheta t_1}\middle|\mathcal{F}\right], \quad \Phi^{\mathcal{I}(2)}_{\vartheta\mathcal{T}(2)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(2))|} \sum_{(\ell_1,\ell_2)\in\sigma(\mathcal{I}(2))} \mathbb{E}^c_\vartheta \left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\middle|\mathcal{F}\right], \\
&\Phi^{\mathcal{I}(3)}_{\vartheta\mathcal{T}(3)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(3))|} \sum_{(\ell_1,\ell_2,\ell_3)\in\sigma(\mathcal{I}(3))} \mathbb{E}^c_\vartheta \left[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}\middle|\mathcal{F}\right], \\
&\Phi^{\mathcal{I}(4)}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] := \frac{1}{|\sigma(\mathcal{I}(4))|} \sum_{(\ell_1,\ldots,\ell_4)\in\sigma(\mathcal{I}(4))} \tilde{\Phi}^{\ell_1\ell_2\ell_3\ell_4}_{\vartheta\mathcal{T}(4)},
\end{aligned} \tag{7}$$

where $\tilde{\Phi}^{\ell_1\ell_2\ell_3\ell_4}_{\vartheta\mathcal{T}(4)} := \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}\phi^{\ell_4}_{\vartheta t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_2}_{\vartheta t_2}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_3}_{\vartheta t_3}\phi^{\ell_4}_{\vartheta t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_3}_{\vartheta t_3}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_4}_{\vartheta t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{\vartheta t_1}\phi^{\ell_4}_{\vartheta t_4}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_2}_{\vartheta t_2}\phi^{\ell_3}_{\vartheta t_3}|\mathcal{F}]$, and $\Phi^{\mathcal{I}(5)}_{\vartheta\mathcal{T}(5)}[\mathcal{F}]$ and $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ are defined in Section 12.2.1 in appendix. 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 $x \in \mathcal{X}$,

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

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 $\sum_{\mathcal{T}(j)\in\{-m+1,\ldots,k-1\}^j}$ 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 dropping $X_{-m} = x$ from the conditioning variable.

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

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

denote the stationary density of $\mathbf{Y}^k_{-m+1}$ associated with ϑ conditional on $\{\overline{\mathbf{Y}}_{-m}, \mathbf{W}^k_{-m}\}$, where X_{-m} is drawn from its true conditional stationary distribution $\mathbb{P}_{\vartheta^*}(X_{-m}|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m})$. Let $\overline{p}_\vartheta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m}) := \overline{p}_\vartheta(\mathbf{Y}^k_{-m+1}|\overline{\mathbf{Y}}_{-m}, \mathbf{W}^k_{-m})/\overline{p}_\vartheta(\mathbf{Y}^{k-1}_{-m+1}|\overline{\mathbf{Y}}_{-m}, \mathbf{W}^{k-1}_{-m})$ denote the associated conditional density of Y_k given $(\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k)$.[5]

Define the density ratio as $l_{k,m,x}(\vartheta) := p_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m}=x)/p_{\vartheta^*}(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, \mathbf{W}_{-m}^k, X_{-m}=x)$. 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

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

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

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 $x\in\mathcal{X}$,*

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

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

Assumption 4. *There exists a positive real δ such that on $\mathcal{N}^* := \{\vartheta\in\Theta : |\vartheta_y - \vartheta_y^*| < \delta\}$ the following conditions hold: (a) For all $(\overline{\mathbf{y}}, y', x, w)\in\mathcal{Y}^s\times\mathcal{Y}\times\mathcal{X}\times\mathcal{W}$, $g_{\vartheta_y}(y'|\overline{\mathbf{y}}, x, w)$ is six times continuously differentiable on $\mathcal{N}^*$. (b) $\mathbb{E}_{\vartheta^*}[\sup_{\vartheta\in\mathcal{N}^*}\sup_{x\in\mathcal{X}} |\nabla^j \log g_{\vartheta_y}(Y_1|\overline{\mathbf{Y}}_0, x, W)|^{2q_j}] < \infty$ 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*

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

$q_0 = (1+\varepsilon)q_\vartheta$ and $q_g = (1+\varepsilon)q_\vartheta/\varepsilon$ for some $\varepsilon > 0$ and $q_\vartheta > \max\{3, dim(\vartheta)\}$. (c) For almost all $(\overline{\mathbf{y}}, y', w) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}$, there exists a function $f_{\overline{\mathbf{y}},y',w} : \mathcal{X} \to \mathbb{R}^+$ such that $\sup_{\vartheta\in\mathcal{N}^*} g_{\vartheta_y}(y'|\overline{\mathbf{y}}, x, w) \leq f_{\overline{\mathbf{y}},y',w}(x) < \infty$ and, for almost all $(\overline{\mathbf{y}}, x, w) \in \mathcal{Y}^s \times \mathcal{X} \times \mathcal{W}$, for $j = 1, 2, \ldots, 6$, there exist functions $f^j_{\overline{\mathbf{y}},w,x} : \mathcal{Y} \to \mathbb{R}^+$ in L^1 such that $|\nabla^j g_{\vartheta_y}(y'|\overline{\mathbf{y}}, x, w)| \leq f^j_{\overline{\mathbf{y}},x,w}(y')$ 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 $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$-Cauchy sequence 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. From this result and Proposition 4, $\{\nabla^j \ell_{k,m,x}(\vartheta)\}_{m\geq 0}$ and $\{\nabla^j \overline{\ell}_{k,m}(\vartheta)\}_{m\geq 0}$ converge to $\nabla^j \ell_{k,\infty}(\vartheta)$ uniformly in $\vartheta \in \mathcal{N}^*$ $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{r_j}(\mathbb{P}_{\vartheta^*})$ as the following proposition shows.

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

$$(a) \quad \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta) - \nabla^j \overline{\ell}_{k,m}(\vartheta)| \leq K_j (k+m)^7 \rho_*^{k+m-1} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(b) \quad \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta) - \nabla^j \ell_{k,m',x}(\vartheta)| \leq K_j (k+m)^7 \rho_*^{k+m-1} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(c) \quad \sup_{m\geq 0} \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \ell_{k,m,x}(\vartheta)| + \sup_{m\geq 0} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \overline{\ell}_{k,m}(\vartheta)| \leq M_{j,k} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

where $r_1 = 6q_0$, $r_2 = 3q_0$, $r_3 = 2q_0$, $r_4 = 3q_0/2$, $r_5 = 6q_0/5$, and $r_6 = q_0$. (d) Uniformly in $\vartheta \in \mathcal{N}^*$ and $x \in \mathcal{X}$, $\nabla^j \ell_{k,m,x}(\vartheta)$ and $\nabla^j \overline{\ell}_{k,m}(\vartheta)$ converge $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{r_j}(\mathbb{P}_{\vartheta^*})$ to $\nabla^j \ell_{k,\infty}(\vartheta) \in L^{r_j}(\mathbb{P}_{\vartheta^*})$ as $m \to \infty$.

Finally, we prove the uniform convergence of the derivatives of density-ratios by expressing them as polynomials of the derivatives of log-density and applying Proposition 5 and the Hölder's inequality. As shown in the following Proposition 6, $\{\nabla^j l_{k,m,x}(\vartheta)\}_{m\geq 0}$ and $\{\nabla^j \overline{l}_{k,m}(\vartheta)\}_{m\geq 0}$ 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 $L^{q_\vartheta}(\mathbb{P}_{\vartheta^*})$.

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

$$(a) \quad \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta) - \nabla^j \overline{l}_{k,m}(\vartheta)| \leq K_{j,k} (k+m)^7 \rho_*^{k+m-1} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(b) \quad \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta) - \nabla^j l_{k,m',x}(\vartheta)| \leq K_{j,k} (k+m)^7 \rho_*^{k+m-1} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.},$$

$$(c) \quad \sup_{m\geq 0} \sup_{x\in\mathcal{X}} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j l_{k,m,x}(\vartheta)| + \sup_{m\geq 0} \sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \overline{l}_{k,m}(\vartheta)| \leq K_{j,k} \quad \mathbb{P}_{\vartheta^*}\text{-a.s.}$$

(d) Uniformly in $\vartheta \in \mathcal{N}^*$ and $x \in \mathcal{X}$, $\nabla^j l_{k,m,x}(\vartheta)$ and $\nabla^j \overline{l}_{k,m}(\vartheta)$ converge $\mathbb{P}_{\vartheta^*}$-a.s. and in $L^{q_\vartheta}(\mathbb{P}_{\vartheta^*})$ to $\nabla^j l_{k,\infty}(\vartheta) \in L^{q_\vartheta}(\mathbb{P}_{\vartheta^*})$ as $m \to \infty$. (e) $\sup_{\vartheta\in\mathcal{N}^*} |\nabla^j \overline{l}_{k,0}(\vartheta) - \nabla^j l_{k,\infty}(\vartheta)| \leq K_{j,k} k^7 \rho_*^{k-1}$ $\mathbb{P}_{\vartheta^*}$-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 $s_{\pi k}$ in (5) is a function of $\nabla^j \overline{l}_{k,0}(\vartheta)$'s. Proposition

6 and the dominated convergence theorem for conditional expectations (Durrett, 2010, Theorem 5.5.9) imply that $\mathbb{E}_{\vartheta^*}[\nabla^j l_{k,\infty}(\vartheta)|\overline{\mathbf{Y}}_{-\infty}^{k-1}] = 0$ for all $\vartheta \in \mathcal{N}^*$. Therefore, $\{\nabla^j l_{k,\infty}(\vartheta)\}_{k=-\infty}^{\infty}$ is a stationary, ergodic, and square integrable martingale difference sequence, and $\{\nabla^j l_{k,\infty}(\vartheta)\}_{j=1}^{5}$ satisfies Assumption 3(a)(b)(g).

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. Assumption 2(a) restricts p_{11} and p_{22} away from 0 and 1. We assume that the parameter space for $\vartheta_{2,x} = (p_{11}, p_{22})'$ takes the form $[\epsilon, 1-\epsilon]^2$ for a small $\epsilon \in (0, 1/2)$. This assumption is also 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 $p_{\vartheta_1^*}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, x_0)$ if the parameter $\vartheta_2 = (\theta_1, \theta_2, \gamma, p_{11}, p_{22})'$ lies in a subset of the parameter space

$$\Gamma^* := \{(\theta_1, \theta_2, \gamma, p_{11}, p_{22}) \in \Theta_2 : \theta_1 = \theta_2 = \theta^* \text{ and } \gamma = \gamma^*\}.$$

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

Let $\ell_n(\vartheta_2, \xi_2) := \log\left(\sum_{x_0=1}^{2} p_{\vartheta_2}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)\xi_2(x_0)\right)$ denote the two-regime log-likelihood for a given initial distribution $\xi_2(x_0) \in \Xi_2$, and let $\hat{\vartheta}_2 := \arg\max_{\vartheta_2 \in \Theta_2} \ell_n(\vartheta_2, \xi_2)$ denote the maximum likelihood estimator (MLE) of ϑ_2 given ξ_2. Because ξ_2 does not matter asymptotically, we treat ξ_2 fixed 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 $\ell_{0,n}(\vartheta_1) := \sum_{k=1}^{n} \log f(Y_k|\overline{\mathbf{Y}}_{k-1}, W_k; \gamma, \theta)$ 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 $p_{\vartheta_1^*}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^0)$ and $p_{\vartheta_2}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^0)$ is 0 if and only if $\vartheta_2 \in \Gamma^*$.

Assumption 5. *(a) Θ_1 and Θ_2 are compact, and ϑ_1^* is in the interior of Θ_1. (b) For all $(x, x') \in \mathcal{X}$ and all $(\overline{\mathbf{y}}, y', w) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}$, $f(y'|\overline{\mathbf{y}}_0, w; \gamma, \theta)$ is continuous in (γ, θ). (c) If $\vartheta_1 \neq \vartheta_1^*$, then $\mathbb{P}_{\vartheta_1^*}\left(f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma, \theta) \neq f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma^*, \theta^*)\right) > 0$. (d) $\mathbb{E}_{\vartheta_1^*}[\log p_{\vartheta_2}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)] = \mathbb{E}_{\vartheta_1^*}[\log p_{\vartheta_1^*}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)]$ for all $m \geq 0$ if and only if $\vartheta_2 \in \Gamma^*$.*

The following proposition shows that the MLE of ϑ_1^* and $\vartheta_{2,y}^*$ are consistent under this condition.

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

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 $\alpha := \mathbb{P}_{\vartheta_{2,x}}(X_k = 1) = (1 - p_{22})/(2 - p_{11} - p_{22})$. The parameter spaces for ϱ and α under restriction $p_{11}, p_{22} \in [\epsilon, 1-\epsilon]$ are given by $\Theta_\varrho := [-1+2\epsilon, 1-2\epsilon]$ and $\Theta_\alpha := [\epsilon, 1-\epsilon]$, respectively. Because the mapping from (p_{11}, p_{22}) to (ϱ, α) is one-to-one, we reparameterize π as $\pi := (\varrho, \alpha)' \in \Theta_\pi := \Theta_\varrho \times \Theta_\alpha$, and let $p_{\psi\pi}(\cdot|\cdot) := p_{\vartheta_2}(\cdot|\cdot)$. Henceforth, we suppress $\mathbf{W}_0^n$ for notational brevity and write, for example, $p_{\psi\pi}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, \mathbf{W}_0^n, x_0)$ as $p_{\psi\pi}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, x_0)$ and $p_{\psi\pi}(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1}, w_k)$ as $p_{\psi\pi}(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1})$ when doing so does not cause confusion. 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 $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ with $\overline{p}_{\psi\pi}(Y_1^k|\overline{\mathbf{Y}}_0)$ defined in (9). This approximation is valid because Proposition 6 implies that $\nabla^j p_{\psi\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}, x_0)/p_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}, x_0) - \nabla^j \overline{p}_{\psi\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ 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 $\overline{\mathbf{Y}}_{-\infty}^{\infty}$ and $\mathbf{X}_{-\infty}^{\infty}$ are independent when $\psi = \psi^*$, we have

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

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 $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ 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 $p_{\vartheta_2}(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1}) = g_{\vartheta_{2,y}}(y_k|\overline{\mathbf{y}}_{k-1}, x_k) q_{\vartheta_{2,x}}(x_{k-1}, x_k) = \sum_{j=1}^2 \mathbb{I}\{x_k = j\} f(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \theta_j) q_{\vartheta_{2,x}}(x_{k-1}, x_k)$.

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 $p_{\psi\pi}(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1}) = g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k)q_\pi(x_{k-1}, x_k)$ and

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

Henceforth, let f_k^*, ∇f_k^*, g_k^*, and ∇g_k^* denote $f(Y_k|\overline{\mathbf{Y}}_{k-1}; \gamma^*, \theta^*)$, $\nabla f(Y_k|\overline{\mathbf{Y}}_{k-1}; \gamma^*, \theta^*)$, $g_{\psi^*}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$, and $\nabla g_{\psi^*}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$, respectively, and similarly for $\log f_k^*$,$\nabla \log f_k^*$, $\log g_k^*$, and $\nabla \log g_k^*$. Expanding $g_\psi(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$ twice with respect to $\psi = (\gamma', \nu', \lambda')'$ and evaluating at ψ^* gives

$$\begin{aligned} \nabla_\eta g_k^* &= \nabla_{(\gamma',\theta')'} f_k^*, \quad \nabla_\lambda g_k^* = (q_k - \alpha)\nabla_\theta f_k^*, \\ \nabla_{\lambda\eta'} g_k^* &= (q_k - \alpha)\nabla_{\theta(\gamma',\theta')} f_k^*, \quad \nabla_{\lambda\lambda'} g_k^* = (q_k - \alpha)^2 \nabla_{\theta\theta'} f_k^*. \end{aligned} \tag{13}$$

Recall $\varrho := \text{corr}_{\vartheta_2^*}(q_k, q_{k+1})$. Observe that q_k satisfies

$$\begin{aligned} &\mathbb{E}_{\vartheta_2^*}(q_k - \alpha)^2 = \alpha(1-\alpha), \quad \mathbb{E}_{\vartheta_2^*}(q_k - \alpha)^3 = \alpha(1-\alpha)(1-2\alpha), \\ &\mathbb{E}_{\vartheta_2^*}(q_k - \alpha)^4 = \alpha(1-\alpha)(3\alpha^2 - 3\alpha + 1), \quad \text{corr}_{\vartheta_2^*}(q_k, q_{k+\ell}) = \varrho^{|\ell|}, \end{aligned} \tag{14}$$

where the first three results follow from the property of a Bernoulli random variable, and the last result holds because q_k follows an $AR(1)$ process with the autoregressive coefficient ϱ (Hamilton, 1994, p. 684). Then, it follows from (10) and (14) that

$$\mathbb{E}_{\vartheta^*}[q_k - \alpha|\overline{\mathbf{Y}}_{-\infty}^n] = 0, \quad \mathbb{E}_{\vartheta^*}[(q_{t_1} - \alpha)(q_{t_2} - \alpha)|\overline{\mathbf{Y}}_{-\infty}^n] = \alpha(1-\alpha)\varrho^{t_2 - t_1}, \quad t_2 \geq t_1. \tag{15}$$

From Louis Information Principle (Lemma 1), $\log p_{\psi\pi}(y_k, x_k|\overline{\mathbf{y}}_{k-1}, x_{k-1}) = \log g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k) + \log q_\pi(x_{k-1}, x_k)$, and the definition of $\overline{p}_{\psi\pi}(Y_1^k|\overline{\mathbf{Y}}_0)$ in (9), we obtain

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

Applying (13), (15), and $g_k^* = f_k^*$ to the right hand side gives

$$\frac{\nabla_\eta \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \nabla_{(\gamma',\theta')'} \log f_k^*, \quad \frac{\nabla_\lambda \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = 0. \tag{16}$$

Similarly, it follows from Lemma 1, (13), (15), (16), and $g_k^* = f_k^*$ that

$$
\begin{aligned}
&\nabla_{\lambda\eta'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0, && (17)\\
&\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) \\
&= \nabla_{\lambda\lambda'}\log \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) \\
&= \sum_{t=1}^{k}\mathbb{E}_{\vartheta^*}\left[\nabla_{\lambda\lambda'}\log g_t^*\middle|\overline{\mathbf{Y}}_0^{k}\right] - \sum_{t=1}^{k-1}\mathbb{E}_{\vartheta^*}\left[\nabla_{\lambda\lambda'}\log g_t^*\middle|\overline{\mathbf{Y}}_0^{k-1}\right] \\
&\quad + \sum_{t_1=1}^{k}\sum_{t_2=1}^{k}\mathbb{E}_{\vartheta^*}\left[\frac{\nabla_{\lambda}g_{t_1}^*}{g_{t_1}^*}\frac{\nabla_{\lambda'}g_{t_2}^*}{g_{t_2}^*}\middle|\overline{\mathbf{Y}}_0^{k}\right] - \sum_{t_1=1}^{k-1}\sum_{t_2=1}^{k-1}\mathbb{E}_{\vartheta^*}\left[\frac{\nabla_{\lambda}g_{t_1}^*}{g_{t_1}^*}\frac{\nabla_{\lambda'}g_{t_2}^*}{g_{t_2}^*}\middle|\overline{\mathbf{Y}}_0^{k-1}\right] \\
&= \alpha(1-\alpha)\left[\frac{\nabla_{\theta\theta'}f_k^*}{f_k^*} + \sum_{t=1}^{k-1}\varrho^{k-t}\left(\frac{\nabla_{\theta}f_t^*}{f_t^*}\frac{\nabla_{\theta'}f_k^*}{f_k^*} + \frac{\nabla_{\theta}f_k^*}{f_k^*}\frac{\nabla_{\theta'}f_t^*}{f_t^*}\right)\right]. && (18)
\end{aligned}
$$

Note that $-1+2\epsilon \le \varrho \le 1-2\epsilon$ in Θ_ϱ. Because the first-order derivative with respect to λ is identically equal to zero in (16), the unique elements of $\nabla_{\eta}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ and $\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ 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 3 from Lemma 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

$$
\begin{aligned}
v(\lambda) &:= (\lambda_1^2, \dots, \lambda_q^2, \lambda_1\lambda_2, \dots, \lambda_1\lambda_q, \lambda_2\lambda_3, \dots, \lambda_2\lambda_q, \dots, \lambda_{q-1}\lambda_q)', \\
V(s) &:= (s_{11}/2, \dots, s_{qq}/2, s_{12}, \dots, s_{1q}, s_{23}, \dots, s_{2q}, \dots, s_{q,q-1})'.
\end{aligned} \tag{19}
$$

Noting that $\alpha(1-\alpha) > 0$ for $\alpha \in \Theta_\alpha$, define, with $t_\lambda(\lambda, \pi) := \alpha(1-\alpha)v(\lambda)$,

$$
t(\psi,\pi) := \begin{pmatrix} \eta - \eta^* \\ t_\lambda(\lambda,\pi) \end{pmatrix}, \ s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\varrho k} \end{pmatrix}, \text{ where } s_{\eta k} := \frac{\nabla_{\eta}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix}, \tag{20}
$$

and $s_{\lambda\varrho k} := V(s_{\lambda\lambda\varrho k})$ with

$$
s_{\lambda\lambda\varrho k} := \frac{\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\alpha(1-\alpha)\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \frac{\nabla_{\theta\theta'}f_k^*}{f_k^*} + \sum_{t=1}^{k-1}\varrho^{k-t}\left(\frac{\nabla_{\theta}f_t^*}{f_t^*}\frac{\nabla_{\theta'}f_k^*}{f_k^*} + \frac{\nabla_{\theta}f_k^*}{f_k^*}\frac{\nabla_{\theta'}f_t^*}{f_t^*}\right). \tag{21}
$$

Here, $s_{\varrho k}$ in (20) 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 $\mathcal{N}_\varepsilon := \{\vartheta_2 \in \Theta_2 : |t(\psi,\pi)| < \varepsilon\}$. Let $A_{n\varepsilon}(\xi) := \{\vartheta \in \mathcal{N}_\varepsilon : \ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi) \ge 0\}$ and $A_{n\varepsilon c}(\xi) := A_{n\varepsilon}(\xi) \cup \mathcal{N}_{c/\sqrt{n}}$, where we suppress the subscript 2 from ξ_2. We use this definition of $A_{n\varepsilon c}(\xi)$ through Sections 6.1-6.3. As shown in Sections 6.2 and 6.3, Assumption 6 does not hold for regime switching models with normal distribution.

Assumption 6. *$0 < \inf_{\varrho \in \Theta_\varrho} \lambda_{\min}(\mathcal{I}_\varrho) \le \sup_{\varrho \in \Theta_\varrho} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ for $\mathcal{I}_\varrho = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho k} s_{\varrho k}')$, where $s_{\varrho k}$ is given in (20).*

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

$$\sup_{\xi \in \Xi} \sup_{\vartheta \in A_{n\varepsilon c}(\xi)} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n} t(\psi, \pi)' \nu_n(s_{\varrho k}) + n t(\psi, \pi)' \mathcal{I}_\varrho t(\psi, \pi)/2 \right| = o_{p\varepsilon}(1). \quad (22)$$

We proceed to derive the asymptotic distribution of the LRTS. With $s_{\varrho k}$ defined in (20), define

$$\begin{aligned} &\mathcal{I}_\eta := \mathbb{E}_{\vartheta^*}(s_{\eta k} s_{\eta k}'), \quad \mathcal{I}_{\lambda \varrho_1 \varrho_2} := \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\lambda \varrho_1 k} s_{\lambda \varrho_2 k}'), \quad \mathcal{I}_{\lambda \eta \varrho} := \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\lambda \varrho k} s_{\eta k}'), \\ &\mathcal{I}_{\eta \lambda \varrho} := \mathcal{I}_{\lambda \eta \varrho}', \quad \mathcal{I}_{\lambda.\eta \varrho_1 \varrho_2} := \mathcal{I}_{\lambda \varrho_1 \varrho_2} - \mathcal{I}_{\lambda \eta \varrho_1} \mathcal{I}_\eta^{-1} \mathcal{I}_{\eta \lambda \varrho_2}, \quad \mathcal{I}_{\lambda.\eta\varrho} := \mathcal{I}_{\lambda.\eta\varrho\varrho}, \quad Z_{\lambda\varrho} := (\mathcal{I}_{\lambda.\eta\varrho})^{-1} G_{\lambda.\eta\varrho}, \end{aligned} \quad (23)$$

where $G_{\lambda.\eta\varrho}$ is a q_λ-vector mean zero Gaussian process indexed by ϱ with $\mathrm{cov}(G_{\lambda.\eta\varrho_1}, G_{\lambda.\eta\varrho_2}) = \mathcal{I}_{\lambda.\eta\varrho_1\varrho_2}$. Define the set of admissible values of $\sqrt{n}\alpha(1-\alpha)v(\lambda)$ 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 $\tilde{t}_{\lambda\varrho}$ by

$$r_{\lambda\varrho}(\tilde{t}_{\lambda\varrho}) = \inf_{t_\lambda \in v(\mathbb{R}^q)} r_{\lambda\varrho}(t_\lambda), \quad r_{\lambda\varrho}(t_\lambda) := (t_\lambda - Z_{\lambda\varrho})' \mathcal{I}_{\lambda.\eta\varrho}(t_\lambda - Z_{\lambda\varrho}). \quad (24)$$

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 $M = 1$, $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho \in \Theta_{\varrho\epsilon}} \left(\tilde{t}_{\lambda\varrho}' \mathcal{I}_{\lambda.\eta\varrho} \tilde{t}_{\lambda\varrho} \right)$.*

In proposition 9, the LRTS and its asymptotic distribution depends on the choice of ϵ because $\Theta_\varrho = [-1+2\epsilon, 1-2\epsilon]$. 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

$$f(y_k | \overline{\mathbf{y}}_{k-1}; \gamma, \theta_j) = f(y_k | \overline{\mathbf{y}}_{k-1}; \gamma, \zeta_j, \sigma_j^2) = \frac{1}{\sigma_j} \phi \left(\frac{y_k - \mu_j - \varpi(\overline{\mathbf{y}}_{k-1}; \gamma, \beta_j)}{\sigma_j} \right), \quad (25)$$

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

Kasahara and Shimotsu (2015) (θ in Kasahara and Shimotsu corresponds to ζ here):

$$\begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \nu_\zeta + (1-\alpha)\lambda_\zeta \\ \nu_\zeta - \alpha\lambda_\zeta \\ \nu_\sigma + (1-\alpha)(2\lambda_\sigma + C_1\lambda_\mu^2) \\ \nu_\sigma - \alpha(2\lambda_\sigma + C_2\lambda_\mu^2) \end{pmatrix}, \tag{26}$$

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

$$g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k) = f\left(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \nu_\zeta + (q_k-\alpha)\lambda_\zeta, \nu_\sigma + (q_k-\alpha)(2\lambda_\sigma + (C_2-q_k)\lambda_\mu^2)\right). \tag{27}$$

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. We proceed to derive the derivatives of $g_\psi(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$ evaluated at ψ^*. $\nabla_\psi g_k^*$, $\nabla_{\lambda\eta'} g_k^*$, and $\nabla_{\lambda\lambda'} g_k^*$ are the same as those given in (13) except for $\nabla_{\lambda_\mu^2} g_k^*$ and that those with respect to λ_σ^j are multiplied by 2^j. Higher-order derivatives of $g_\psi(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$ 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

$$\begin{aligned} &\nabla_{\mu^2} f(\mu,\sigma^2) = 2\nabla_{\sigma^2} f(\mu,\sigma^2), \quad \nabla_{\mu^3} f(\mu,\sigma^2) = 2\nabla_{\mu\sigma^2} f(\mu,\sigma^2), \quad \text{and} \\ &\nabla_{\mu^4} f(\mu,\sigma^2) = 2\nabla_{\mu^2\sigma^2} f(\mu,\sigma^2) = 4\nabla_{\sigma^2\sigma^2} f(\mu,\sigma^2), \end{aligned} \tag{28}$$

we have

$$\nabla_{\lambda_\mu^i} g_k^* = d_{ik} \nabla_{\mu^i} f_k^*, \quad i = 1, \dots, 4, \tag{29}$$

where

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

It follows from $\mathbb{E}_{\vartheta^*}[q_k|\overline{\mathbf{Y}}_{-\infty}^n] = \alpha$, (14), and elementary calculation that

$$\begin{aligned} \mathbb{E}_{\vartheta^*}[d_{ik}|\overline{\mathbf{Y}}_{-\infty}^n] &= 0, \quad \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^i} g_k^*|\overline{\mathbf{Y}}_{-\infty}^k] = 0, \quad i = 1,2,3, \\ \mathbb{E}_{\vartheta^*}[d_{4k}|\overline{\mathbf{Y}}_{-\infty}^n] &= \alpha(1-\alpha)b(\alpha), \\ \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^4} g_k^*|\overline{\mathbf{Y}}_{-\infty}^k] &= \alpha(1-\alpha)b(\alpha)\nabla_{\mu^4} f_k^* = \alpha(1-\alpha)b(\alpha)4\nabla_{\sigma^2\sigma^2} f_k^* = b(\alpha)\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2} g_k^*|\overline{\mathbf{Y}}_{-\infty}^k], \end{aligned} \tag{30}$$

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

We proceed to derive $\nabla^j \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$. Repeating the calculation leading to (16)–(18) and using (30) gives the following: first, (16) and (17) still hold; second, the elements of $\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ except for the $(1,1)$th element are given by (18) after adjusting that the derivative with respect to λ_σ must be multiplied by 2 (e.g., $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma} g_k^*|\overline{\mathbf{Y}}_{-\infty}^n] = 2\nabla_{\sigma^2} f_k^*$ and $\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma\lambda_\mu} g_k^*|\overline{\mathbf{Y}}_{-\infty}^n] = 2\nabla_{\sigma^2\mu} f_k^*$); third,

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

When $\varrho \neq 0$, $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ is a non-degenerate random variable as in the non-normal case. When $\varrho = 0$, however, $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ 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 (31) as

$$\frac{\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\alpha(1-\alpha)\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \sum_{t=1}^{k-1}\varrho^{k-t}\left(2\frac{\nabla_\mu f_t^*}{f_t^*}\frac{\nabla_\mu f_k^*}{f_k^*}\right) = \varrho\zeta_{k,0}(\varrho). \tag{32}$$

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

Define $v(\lambda_\beta)$ as $v(\lambda)$ in (19) but replacing λ with λ_β. Collect the relevant parameters as

$$t(\psi,\pi) := \begin{pmatrix}\eta - \eta^* \\ t_\lambda(\lambda,\pi)\end{pmatrix}, \tag{33}$$

where

$$t_\lambda(\lambda,\pi) := \alpha(1-\alpha)\begin{pmatrix}\varrho\lambda_\mu^2 \\ \lambda_\mu\lambda_\sigma \\ \lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12 \\ \lambda_\beta\lambda_\mu \\ \lambda_\beta\lambda_\sigma \\ v(\lambda_\beta)\end{pmatrix}, \tag{34}$$

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 (21), define the elements of the generalized score by

$$\begin{pmatrix} * & s_{\lambda_{\mu\beta}\varrho k} & s_{\lambda_{\mu\sigma}\varrho k} \\ s_{\lambda_{\beta\mu}\varrho k} & s_{\lambda_{\beta\beta}\varrho k} & s_{\lambda_{\beta\sigma}\varrho k} \\ s_{\lambda_{\sigma\mu}\varrho k} & s_{\lambda_{\beta\sigma}\varrho k} & s_{\lambda_{\sigma\sigma}\varrho k} \end{pmatrix} = \frac{\nabla_{\theta\theta'} f_k^*}{f_k^*} + \sum_{t=1}^{k-1} \varrho^{k-t} \left(\frac{\nabla_\theta f_t^*}{f_t^*} \frac{\nabla_{\theta'} f_k^*}{f_k^*} + \frac{\nabla_\theta f_k^*}{f_k^*} \frac{\nabla_{\theta'} f_t^*}{f_t^*} \right). \tag{35}$$

Define the generalized score as

$$s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\varrho k} \end{pmatrix}, \quad \text{where} \quad s_{\eta k} := \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix} \quad \text{and} \quad s_{\lambda\varrho k} := \begin{pmatrix} \zeta_{k,0}(\varrho)/2 \\ 2s_{\lambda_{\mu\sigma}\varrho k} \\ 2s_{\lambda_{\sigma\sigma}\varrho k} \\ s_{\lambda_{\beta\mu}\varrho k} \\ 2s_{\lambda_{\beta\sigma}\varrho k} \\ V(s_{\lambda_{\beta\beta}\varrho k}) \end{pmatrix}. \tag{36}$$

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

Assumption 7. *(a)* $0 < \inf_{\varrho\in\Theta_\varrho} \lambda_{\min}(\mathcal{I}_\varrho) \leq \sup_{\varrho\in\Theta_\varrho} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ *for* $\mathcal{I}_\varrho = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho k} s_{\varrho k}')$, *where* $s_{\varrho k}$ *is given in (36). (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 (25). Then, under the null hypothesis of* $M = 1$, *(a)* $\sup_{\vartheta\in A_{n\varepsilon}(\xi)} |t(\psi,\pi)| = O_{p\varepsilon}(n^{-1/2})$*; and (b) for all* $c > 0$,

$$\sup_{\xi\in\Xi} \sup_{\vartheta\in A_{n\varepsilon c}(\xi)} \left| \ell_n(\psi,\pi,\xi) - \ell_n(\psi^*,\pi,\xi) - \sqrt{n} t(\psi,\pi)'\nu_n(s_{\varrho k}) + nt(\psi,\pi)'\mathcal{I}_\varrho t(\psi,\pi)/2 \right| = o_{p\varepsilon}(1). \tag{37}$$

Let $\Lambda_{\lambda\varrho n}$ be the set of possible values of $\sqrt{n} t_\lambda(\lambda,\pi)$ defined in (34). 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 (23) but with $s_{\varrho k}$ defined in (36), and the supremum is taken with respect to t_λ and $\varrho \in \Theta_\varrho$ 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 $q_\beta := \dim(\beta)$, $q_\lambda := 3+2q_\beta+q_\beta(q_\beta+1)/2$, and

$$\begin{aligned} \Lambda_\lambda^1 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}', t_{\beta\sigma}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : \\ &\qquad (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}')' \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_- \times \mathbb{R}^{q_\beta}, t_{\beta\sigma} = 0, t_{v(\beta)} = 0\}, \\ \Lambda_{\lambda\varrho}^2 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu\sigma}, t_{\sigma^2}, t_{\beta\mu}', t_{\beta\sigma}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : t_{\varrho\mu^2} = \varrho\lambda_\mu^2, t_{\mu\sigma} = \lambda_\mu\lambda_\sigma, \\ &\qquad t_{\sigma^2} = \lambda_\sigma^2, t_{\beta\mu} = \lambda_\beta\lambda_\mu, t_{\beta\sigma} = \lambda_\beta\lambda_\sigma, t_{v(\beta)} = v_\beta(\lambda_\beta) \text{ for some } \lambda \in \mathbb{R}^{2+q_\beta}\}. \end{aligned} \tag{38}$$

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 limits of the set of possible values 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 (23) but with $s_{\pi k}$ defined in (36). 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 $\tilde{t}^1_\lambda$ and $\tilde{t}^2_{\lambda\varrho}$ by

$$\begin{aligned} r_\lambda(\tilde{t}^1_\lambda) &= \inf_{t_\lambda \in \Lambda^1_\lambda} r_\lambda(t_\lambda), \quad r_\lambda(t_\lambda) := (t_\lambda - Z_{\lambda 0})' \mathcal{I}_{\lambda.\eta 0}(t_\lambda - Z_{\lambda 0}) \\ r_{\lambda\varrho}(\tilde{t}^2_{\lambda\varrho}) &= \inf_{t_\lambda \in \Lambda^2_{\lambda\varrho}} r_{\lambda\varrho}(t_\lambda), \quad r_{\lambda\varrho}(t_\lambda) := (t_\lambda - Z_{\lambda\varrho})' \mathcal{I}_{\lambda.\eta\varrho}(t_\lambda - Z_{\lambda\varrho}). \end{aligned} \tag{39}$$

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

Proposition 11. *Suppose that assumptions in Proposition 10 hold. Then, under the null hypothesis of* $M = 1$, $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho \in \Theta_\varrho} \max\{\mathbb{I}\{\varrho = 0\}(\tilde{t}^1_\lambda)' \mathcal{I}_{\lambda.\eta 0} \tilde{t}^1_\lambda, (\tilde{t}^2_{\lambda\varrho})' \mathcal{I}_{\lambda.\eta\varrho} \tilde{t}^2_{\lambda\varrho}\}$.

Remark 1. *Qu and Zhuo (2017) derived the asymptotic distribution of the LRTS under the restriction that* $\varrho \geq \epsilon > 0$.

Remark 2. *It is possible to extend our analysis to exponential-LR type tests studied by Andrews and Ploberger (1994) and Carrasco et al. (2014).*

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

$$f(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \theta_j) = f(y_k|\overline{\mathbf{y}}_{k-1}; \tilde{\gamma}, \theta_j, \sigma^2) = \frac{1}{\sigma}\phi\left(\frac{y_k - \mu_j - \varpi(\overline{\mathbf{y}}_{k-1}; \tilde{\gamma}, \beta_j)}{\sigma}\right). \tag{40}$$

for some function ϖ. Consider the following reparameterization:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \nu_\theta + (1-\alpha)\lambda \\ \nu_\theta - \alpha\lambda \\ \nu_\sigma - \alpha(1-\alpha)\lambda_\mu^2 \end{pmatrix}, \tag{41}$$

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

$$g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k) = f\left(y_k|\overline{\mathbf{y}}_{k-1}; \tilde{\gamma}, \nu_\theta + (q_k - \alpha)\lambda, \nu_\sigma - \alpha(1-\alpha)\lambda_\mu^2\right). \tag{42}$$

Let $\eta = (\tilde{\gamma}', \nu_\theta', \nu_\sigma)'$, then the first and second derivatives of $g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k)$ with respect to η and λ are the same as those given in (13) except for $\nabla_{\lambda_\mu^2} g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k)$. We derive higher-order derivatives of $g_\psi(y_k|\overline{\mathbf{y}}_{k-1}, x_k)$ with respect to λ_μ. From Lemma 5 and (28), we obtain

$$\begin{aligned} \nabla_{\lambda\eta^i} g_k^* &= d_{1k} \nabla_{\theta\eta^i} f_k^* \quad \text{for } i = 0, 1, \ldots, \\ \nabla_{\lambda_\mu^i} g_k^* &= d_{ik} \nabla_{\mu^i} f_k^* \quad \text{for } i = 0, 1, \ldots, 4, \end{aligned} \tag{43}$$

where $d_{0k} := 1$, $d_{1k} := q_k - \alpha$, $d_{2k} := (q_k-\alpha)^2 - \alpha(1-\alpha)$, $d_{3k} := (q_k-\alpha)^3 - 3(q_k-\alpha)\alpha(1-\alpha)$, and $d_{4k} := (q_k-\alpha)^4 - 6(q_k-\alpha)^2\alpha(1-\alpha) + 3\alpha^2(1-\alpha)^2$. It follows from $\mathbb{E}_{\vartheta^*}[q_k|\overline{\mathbf{Y}}_{-\infty}^n] = \alpha$, (14), and elementary calculation that

$$
\begin{aligned}
\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\mu^i} g_k^* | \overline{\mathbf{Y}}_0^k] = 0, \quad \mathbb{E}_{\vartheta^*}[d_{ik}|\overline{\mathbf{Y}}_0^k] = 0, \quad i = 1, 2,\\
\mathbb{E}_{\vartheta^*}[d_{3k}|\overline{\mathbf{Y}}_0^k] = \alpha(1-\alpha)(1-2\alpha), \quad \mathbb{E}_{\vartheta^*}[d_{4k}|\overline{\mathbf{Y}}_0^k] = \alpha(1-\alpha)(1-6\alpha+6\alpha^2).
\end{aligned} \tag{44}
$$

Repeating the calculation leading to (16)–(18) and using (44) gives the following: first, (16) and (17) still hold; second, the elements of $\nabla_{\lambda\lambda'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ are given by (18) except for the $(1,1)$th element; third, $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ is given by (31). Further, Lemma 7 in the Appendix shows that, when $\varrho = 0$, $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \alpha(1-\alpha)(1-2\alpha)\nabla_{\mu^3} f_k^*/f_k^*$ and $\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\nabla_{\mu^4} f_k^*/f_k^*$. Because $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$ 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

$$
t(\psi,\pi) := \begin{pmatrix} \eta - \eta^* \\ t_\lambda(\lambda,\pi) \end{pmatrix} \quad \text{and} \quad t_\lambda(\lambda,\pi) := \alpha(1-\alpha)\begin{pmatrix} \varrho\lambda_\mu^2 \\ (1-2\alpha)\lambda_\mu^3 \\ (1-6\alpha+6\alpha^2)\lambda_\mu^4 \\ \lambda_\beta\lambda_\mu \\ v(\lambda_\beta) \end{pmatrix}. \tag{45}
$$

Define the generalized score as

$$
s_{\varrho k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\varrho k} \end{pmatrix}, \quad \text{where} \quad s_{\eta k} := \begin{pmatrix} \nabla_\gamma f_k^*/f_k^* \\ \nabla_\theta f_k^*/f_k^* \end{pmatrix} \quad \text{and} \quad s_{\lambda\varrho k} := \begin{pmatrix} \zeta_{k,0}(\varrho)/2 \\ s_{\lambda_\mu^3 k}/3! \\ s_{\lambda_\mu^4 k}/4! \\ s_{\lambda_{\beta\mu}\varrho k} \\ V(s_{\lambda_{\beta\beta}\varrho k}) \end{pmatrix}, \tag{46}
$$

where $\zeta_{k,m}(\varrho)$ is defined as in (32), $s_{\lambda_\mu^i k} := \nabla_{\mu^i} f_k^*/f_k^*$ for $i = 3, 4$, and $s_{\lambda_{\beta\mu}\varrho k}$ and $s_{\lambda_{\beta\beta}\varrho k}$ are defined as in (35) but using the density function (40) in place of (25). Define, with $q_\beta := \dim(\beta)$ and $q_\lambda := 3 + q_\beta + q_\beta(q_\beta+1)/2$,

$$
\begin{aligned}
\Lambda_\lambda^1 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}')' \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}_-\times\mathbb{R}^{q_\beta}, t_{v(\beta)} = 0\},\\
\Lambda_{\lambda\varrho}^2 &:= \{t_\lambda = (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda} : t_{\varrho\mu^2} = \varrho\lambda_\mu^2, t_{\mu^3} = t_{\mu^4} = 0, t_{\beta\mu} = \lambda_\beta\lambda_\mu,\\
&\qquad t_{v(\beta)} = v_\beta(\lambda_\beta) \text{ for some } \lambda \in \mathbb{R}^{1+q_\beta}\}.
\end{aligned} \tag{47}
$$

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 LRTS.

Assumption 8. $0 < \inf_{\varrho \in \Theta_\varrho} \lambda_{\min}(\mathcal{I}_\varrho) \le \sup_{\varrho \in \Theta_\varrho} \lambda_{\max}(\mathcal{I}_\varrho) < \infty$ *for* $\mathcal{I}_\varrho = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho k} s_{\varrho k}')$, *where* $s_{\varrho k}$ *is given in (46).*

Proposition 12. *Suppose Assumptions 1, 2, 4, 5 and 8 hold, and the density for the* j*-th regime is given by (40). Then, statements (a) and (b) of Proposition 10 hold.*

Proposition 13. *Suppose that assumptions in Proposition 12 hold. Then, under the null hypothesis of* $M = 1$, $2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho \in \Theta_\varrho} \max\{\mathbb{I}\{\varrho = 0\}(\tilde{t}_\lambda^1)'\mathcal{I}_{\lambda.\eta 0}\tilde{t}_\lambda^1, (\tilde{t}_{\lambda\varrho}^2)'\mathcal{I}_{\lambda.\eta\varrho}\tilde{t}_{\lambda\varrho}^2\}$*, where* $\tilde{t}_\lambda^1$ *and* $\tilde{t}_{\lambda\varrho}^2$ *are defined as in (39) 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 (46) and* Λ_λ^1 *and* $\Lambda_{\lambda\varrho}^2$ *defined in (47).*

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

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 $M_0 \ge 2$. We suppress the covariate $\mathbf{W}_a^b$ unless confusion might arise.

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 $\mathbf{Y}_1^n$ given $\overline{\mathbf{Y}}_0$ and x_0 is

$$p_{\vartheta_{M_0}^*}(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, x_0) = \sum_{\mathbf{x}_1^n \in \mathcal{X}_{M_0}^n} \prod_{k=1}^n p_{\vartheta_{M_0}^*}(Y_k, x_k | \overline{\mathbf{Y}}_{k-1}, x_{k-1}), \tag{48}$$

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

Let the conditional density of $\mathbf{Y}_1^n$ of an $(M_0 + 1)$-regime model be

$$p_{\vartheta_{M_0+1}}(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, x_0) := \sum_{\mathbf{x}_1^n \in \mathcal{X}_{M_0+1}^n} \prod_{k=1}^n p_{\vartheta_{M_0+1}}(Y_k, x_k | \overline{\mathbf{Y}}_{k-1}, x_{k-1}), \tag{49}$$

where $p_{\vartheta_{M_0+1}}(y_k, x_k | \overline{\mathbf{y}}_{k-1}, x_{k-1})$ is defined similarly to $p_{\vartheta_{M_0}^*}(y_k, x_k | \overline{\mathbf{y}}_{k-1}, x_{k-1})$ 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 $\min_{i,j} p_{ij} \ge \epsilon$ for some $\epsilon \in (0, 1/2)$.

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

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

Define the set of values of ϑ_{M_0+1} that yields the true density (48) under $\mathbb{P}_{\vartheta_{M_0}^*}$ as $\Upsilon^* := \{\vartheta_{M_0+1} \in \Theta_{M_0+1,\epsilon} : p_{\vartheta_{M_0+1}}(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, x_0) = p_{\vartheta_{M_0}^*}(\mathbf{Y}_1^n | \overline{\mathbf{Y}}_0, x_0)\ \mathbb{P}_{\vartheta_{M_0}^*}\text{-a.s.}\}$. Under H_{0m}, the $(M_0 + 1)$-regime model (49) generates the true M_0-regime density (48) if $\theta_m = \theta_{m+1} = \theta_m^*$ and the transition matrix of X_k reduces to that of the true M_0-regime model.

We reparameterize the transition probability of X_k by writing $\vartheta_{M_0+1,x}$ as $\vartheta_{M_0+1,x} = (\vartheta_{xm}', \pi_{xm}')'$, where ϑ_{xm} is identified under H_{0m} while π_{xm} is not point identified under H_{0m}. The transition probability of X_k under $\vartheta_{M_0+1,x}$ equals the transition probability of X_k under $\vartheta_{M_0,x}^*$ if and only if $\vartheta_{xm} = \vartheta_{xm}^*$. The detailed derivation including the definition of ϑ_{xm}^* is provided in Section 12.2.6 in the appendix. Define the subset of Υ^* that corresponds to H_{0m} as

$$\begin{aligned}\Upsilon_m^* := \big\{\vartheta_{M_0+1} \in \Theta_{M_0+1} : \theta_j = \theta_j^* \text{ for } 1 \leq j < m;\ \theta_m = \theta_{m+1} = \theta_m^*;\\ \theta_j = \theta_{j-1}^* \text{ for } h+1 < j \leq M_0+1;\ \gamma = \gamma^*;\ \vartheta_{xm} = \vartheta_{xm}^*\big\},\end{aligned}$$

then $\Upsilon^* = \Upsilon_1^* \cup \cdots \cup \Upsilon_{M_0}^*$ holds.

For $M = M_0, M_0+1$, let $\ell_n(\vartheta_M, \xi_M) := \log\left(\sum_{x_0=1}^M p_{\vartheta_M}(\mathbf{Y}_1^n|\overline{\mathbf{Y}}_0, x_0)\xi_M(x_0)\right)$ 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 $\hat{\vartheta}_{M_0+1} := \arg\max_{\vartheta_{M_0+1}\in\Theta_{M_0+1}} \ell_n(\vartheta_{M_0+1}, \xi_{M_0+1})$. 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, and $\vartheta_{M_0}^*$ is in the interior of Θ_{M_0}. (b) For all $(x, x') \in \mathcal{X}$ and all $(\overline{\mathbf{y}}, y', w) \in \mathcal{Y}^s \times \mathcal{Y} \times \mathcal{W}$, $f(y'|\overline{\mathbf{y}}_0, w; \gamma, \theta)$ is continuous in (γ, θ). (c) $\mathbb{E}_{\vartheta_{M_0}^*}[\log(p_{\vartheta_{M_0}}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)] = \mathbb{E}_{\vartheta_{M_0}^*}[\log p_{\vartheta_{M_0}^*}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)]$ for all $m \geq 0$ if and only if $\vartheta_{M_0} = \vartheta_{M_0}^*$. (d) $\mathbb{E}_{\vartheta_{M_0}^*}[\log(p_{\vartheta_{M_0+1}}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^0)] = \mathbb{E}_{\vartheta_{M_0}^*}[\log p_{\vartheta_{M_0}^*}(Y_1|\overline{\mathbf{Y}}_{-m}^0, \mathbf{W}_{-m}^1)]$ for all $m \geq 0$ if and only if $\vartheta_{M_0+1} \in \Upsilon^*$.*

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

We proceed to derive the asymptotic distribution of the LRTS by analyzing the behavior of LRTS when $\vartheta_{M_0+1} \in \Upsilon_m^*$ for each m. Define $J_m := \{m, m+1\}$. Observe that, if $\mathbf{X}_1^k \in J_m^k$, then $\mathbf{X}_1^k$ follows a two-state Markov chain on J_m whose transition probability is characterized by $\alpha_m := \mathbb{P}_{\vartheta_{M_0+1}}(X_k = m|X_k \in J_m)$ and $\varrho_m := \text{corr}_{\vartheta_{M_0+1}}(X_{k-1}, X_k|(X_{k-1}, X_k) \in J_m^2)$. **See** ~~The detailed derivation is provided in~~ Section 12.2.6 in the appendix **for the detailed derivation.** Collect reparameterized π_{xm} into $\pi_{xm} := (\varrho_m, \alpha_m, \phi_m')'$, where ϕ_m does not affect the transition probability of $\mathbf{X}_1^k$ when $\mathbf{X}_1^k \in J_m^k$.

Define $q_{kj} := \mathbb{I}\{X_k = j\}$, then we can write α_m and ϱ_m as $\alpha_m = \mathbb{E}_{\vartheta_{M_0+1}}(q_{km}|X_k \in J_m)$ and $\varrho_m = \text{corr}_{\vartheta_{M_0+1}}(q_{k-1,m}, q_{km}|(X_{k-1}, X_k) \in J_m^2)$. Because $\overline{\mathbf{Y}}_{-\infty}^{\infty}$ provides no information for distinguishing between $X_k = m$ and $X_k = m+1$ if $\theta_m = \theta_{m+1}$, we can write α_m and ϱ_m as

$$\alpha_m = \mathbb{E}_{\vartheta_{M_0+1}}(q_{km}|X_k \in J_m, \overline{\mathbf{Y}}_{-\infty}^{\infty}) \quad \text{and} \quad \varrho_m = \text{corr}_{\vartheta_{M_0+1}}(q_{k-1,m}, q_{km}|(X_{k-1}, X_k) \in J_m^2, \overline{\mathbf{Y}}_{-\infty}^{\infty}). \tag{50}$$

7.1 Non-normal distribution

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

$$\begin{pmatrix} \theta_m \\ \theta_{m+1} \end{pmatrix} = \begin{pmatrix} \nu_m + (1-\alpha_m)\lambda_m \\ \nu_m - \alpha_m \lambda_m \end{pmatrix}.$$

Collect the reparameterized identified parameters into one vector $\psi_m := (\eta_m', \lambda_m')'$, where $\eta_m = (\gamma', \{\theta_j'\}_{j=1}^{m-1}, \nu_m', \{\theta_j'\}_{j=m+2}^{M_0+1}, \vartheta_{xm}')'$, so that the reparameterized (M_0+1)-regime log-likelihood function is $\ell_n(\psi_m, \pi_{xm}, \xi_{M_0+1})$. Let $\psi_m^* = (\eta_m^*, \lambda_m^*) = ((\vartheta_{M_0}^*)', 0')'$ denote the value of ψ_m under H_{0m}. Define the reparameterized conditional density of y_k as

$$g_{\psi_m}(y_k|\overline{\mathbf{y}}_{k-1}, x_k) := \mathbb{I}\{x_k \in J_m\} f(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \nu_m + (q_{km} - \alpha_m)\lambda_m) + \sum_{j \in \overline{J}_m} q_{kj} f(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \theta_j),$$

where $\overline{J}_m := \{1, \ldots, M_0+1\} \setminus J_m$. Let f_{mk}^* denote $f(Y_k|\overline{\mathbf{Y}}_{k-1}; \gamma^*, \theta_m^*)$. It follows from (50) and the law of iterated expectations that

$$\begin{aligned} &\mathbb{E}_{\vartheta_{M_0}^*}\left[\frac{\mathbb{I}\{X_k \in J_m\}(q_{km} - \alpha_m)}{g_{\psi_m^*}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)} \middle| \overline{\mathbf{Y}}_{-\infty}^n \right] \\ &= \mathbb{E}_{\vartheta_{M_0}^*}\left[\mathbb{E}_{\vartheta_{M_0}^*}\left[\frac{q_{km} - \alpha_m}{f_{mk}^*} \middle| X_k \in J_m, \overline{\mathbf{Y}}_{-\infty}^n \right] \mathbb{I}\{X_k \in J_m\} \middle| \overline{\mathbf{Y}}_{-\infty}^n \right] = 0, \\ &\mathbb{E}_{\vartheta_{M_0}^*}\left[\frac{\mathbb{I}\{X_{t_1} \in J_m\}\mathbb{I}\{X_{t_2} \in J_m\}(q_{t_1 h} - \alpha_m)(q_{t_2 h} - \alpha_m)}{g_{\psi_m^*}(Y_{t_1}|\overline{\mathbf{Y}}_{t_1-1}, X_{t_1}) g_{\psi_m^*}(Y_{t_2}|\overline{\mathbf{Y}}_{t_2-1}, X_{t_2})} \middle| \overline{\mathbf{Y}}_{-\infty}^n \right] \\ &= \mathbb{E}_{\vartheta_{M_0}^*}\left[\mathbb{E}_{\vartheta_{M_0}^*}\left[\frac{(q_{t_1 h} - \alpha_m)(q_{t_2 h} - \alpha_m)}{f_{mt_1}^* f_{mt_2}^*} \middle| \mathbf{X}_{t_1}^{t_2} \in J_m^{t_2-t_1+1}, \overline{\mathbf{Y}}_{-\infty}^n \right] \mathbb{I}\{(X_{t_1}, X_{t_2}) \in J_m^2\} \middle| \overline{\mathbf{Y}}_{-\infty}^n \right] \\ &= \frac{\alpha_m(1-\alpha_m)\varrho_m^{t_2-t_1}}{f_{mt_1}^* f_{mt_2}^*} \mathbb{P}_{\vartheta_{M_0}^*}((X_{t_1}, X_{t_2}) \in J_m^2 | \overline{\mathbf{Y}}_{-\infty}^n), \quad t_2 \geq t_1, \end{aligned} \tag{51}$$

where the second equality holds because $g_{\psi_m^*}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k) = f_{mk}^*$ if $X_k \in J_m$, and last equality holds because, conditional on $\{\mathbf{X}_{t_1}^{t_2} \in J_m^{t_2-t_1+1}, \overline{\mathbf{Y}}_{-\infty}^n\}$, $\mathbf{X}_{t_1}^{t_2}$ is a two-state stationary Markov process with parameter (α_m, ϱ_m).

Let g_{0k}^*, q_{0k}^*, and $\overline{p}_{0k}^*$ denote $g_{\vartheta_{M_0,y}^*}(Y_k, X_k|\overline{\mathbf{Y}}_{k-1}, X_{k-1})$, $q_{\vartheta_{M_0,x}^*}(X_{k-1}, X_k)$, and $\overline{p}_{\vartheta_{M_0}^*}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$. Let ∇g_{0k}^* denote the derivative of $g_{\vartheta_{M_0,y}}(Y_k, X_k|\overline{\mathbf{Y}}_{k-1}, X_{k-1})$ evaluated at $\vartheta_{M_0,y}^*$, and define ∇q_{0k}^* and $\nabla \overline{p}_{0k}^*$ similarly. Repeating a derivation similar to (13)–(18) but using (51) in place of (15), we obtain

$$\begin{aligned} &\nabla_{\eta_m} \overline{p}_{\psi_m^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) / \overline{p}_{\psi_m^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) \\ &= \sum_{t=1}^{k} \mathbb{E}_{\vartheta^*}\left[\nabla_{\vartheta_{M_0}} \log(g_{0t}^* q_{0t}^*) \middle| \overline{\mathbf{Y}}_0^k \right] - \sum_{t=1}^{k-1} \mathbb{E}_{\vartheta^*}\left[\nabla_{\vartheta_{M_0}} \log(g_{0t}^* q_{0t}^*) \middle| \overline{\mathbf{Y}}_0^{k-1} \right] \\ &= \nabla_{\vartheta_{M_0}} \overline{p}_{\vartheta_{M_0}^*}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) / \overline{p}_{\vartheta_{M_0}^*}(Y_k|\overline{\mathbf{Y}}_0^{k-1}), \end{aligned} \tag{52}$$

$$\nabla_{\lambda_m}\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})=0,\quad \nabla_{\lambda_m\eta_m'}\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})=0, \tag{53}$$

$$\begin{aligned}\frac{\nabla_{\lambda_m\lambda_m{}'}\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} &= \alpha_m(1-\alpha_m)\frac{\nabla_{\theta\theta'}f_{mk}^*}{f_{mk}^*}\mathbb{P}_{\vartheta_{M_0}^*}(X_k\in J_m|\overline{\mathbf{Y}}_0^k)\\ &+\alpha_m(1-\alpha_m)\sum_{t=1}^{k-1}\varrho_m^{k-t}\left(\frac{\nabla_\theta f_{mt}^*}{f_{mt}^*}\frac{\nabla_{\theta'}f_{mk}^*}{f_{mk}^*}+\frac{\nabla_\theta f_{mk}^*}{f_{mk}^*}\frac{\nabla_{\theta'}f_{mt}^*}{f_{mt}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_m^2|\overline{\mathbf{Y}}_0^k).\end{aligned} \tag{54}$$

Define $\tilde{\varrho}:=(\varrho_1,\ldots,\varrho_{M_0})'$, define $t_\lambda(\lambda_m,\pi_m)$ as $t_\lambda(\lambda,\pi)$ in (20) by replacing (λ,π) with (λ_m,π_m), and let

$$t(\psi_m,\pi_m):=\begin{pmatrix}\eta_m-\eta^*\\ t_\lambda(\lambda_m,\pi_m)\end{pmatrix},\ \tilde{s}_{\tilde{\varrho}k}:=\begin{pmatrix}\tilde{s}_{\eta k}\\ \tilde{s}_{\lambda\tilde{\varrho}k}\end{pmatrix},\ \text{ where }\ \tilde{s}_{\eta k}:=\frac{\nabla_{\eta_m}\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})},\ \tilde{s}_{\lambda\tilde{\varrho}k}:=\begin{pmatrix}s_{\lambda\varrho_1 k}^1\\ \vdots\\ s_{\lambda\varrho_{M_0}k}^{M_0}\end{pmatrix}, \tag{55}$$

and $s_{\lambda\varrho_m k}^m:=V(s_{\lambda\lambda\varrho_m k}^m)$, where $s_{\lambda\lambda\varrho_m k}^m$ is defined similarly to (21) as

$$\begin{aligned}s_{\lambda\lambda\varrho_m k}^m &:= \frac{\nabla_{\theta\theta'}f_{mk}^*}{f_{mk}^*}\mathbb{P}_{\vartheta_{M_0}^*}(X_k\in J_m|\overline{\mathbf{Y}}_0^k)\\ &+\sum_{t=1}^{k-1}\varrho_m^{k-t}\left(\frac{\nabla_\theta f_{mt}^*}{f_{mt}^*}\frac{\nabla_{\theta'}f_{mk}^*}{f_{mk}^*}+\frac{\nabla_\theta f_{mk}^*}{f_{mk}^*}\frac{\nabla_{\theta'}f_{mt}^*}{f_{mt}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_m^2|\overline{\mathbf{Y}}_0^k).\end{aligned} \tag{56}$$

Similarly to (23), define

$$\begin{aligned}&\tilde{\mathcal{I}}_\eta:=\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\eta k}\tilde{s}_{\eta k}'),\quad \tilde{\mathcal{I}}_{\lambda\tilde{\varrho}_1\tilde{\varrho}_2}:=\lim_{k\to\infty}\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\lambda\tilde{\varrho}_1 k}\tilde{s}_{\lambda\tilde{\varrho}_2 k}'),\quad \tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}}:=\lim_{k\to\infty}\mathbb{E}_{\vartheta_{M_0}^*}(\tilde{s}_{\lambda\tilde{\varrho}k}\tilde{s}_{\eta k}'),\\ &\tilde{\mathcal{I}}_{\eta\lambda\tilde{\varrho}}:=\tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}}',\quad \tilde{\mathcal{I}}_{\lambda.\eta\tilde{\varrho}_1\tilde{\varrho}_2}:=\tilde{\mathcal{I}}_{\lambda\tilde{\varrho}_1\tilde{\varrho}_2}-\tilde{\mathcal{I}}_{\lambda\eta\tilde{\varrho}_1}\tilde{\mathcal{I}}_\eta^{-1}\tilde{\mathcal{I}}_{\eta\lambda\tilde{\varrho}_2},\quad \tilde{\mathcal{I}}_{\lambda.\eta\varrho_m}^m:=\mathbb{E}_{\vartheta_{M_0}^*}[G_{\lambda.\eta\varrho_m}^m(G_{\lambda.\eta\varrho_m}^m)'],\\ &Z_{\lambda\varrho_m}^m:=(\tilde{\mathcal{I}}_{\lambda.\eta\varrho_m}^m)^{-1}G_{\lambda.\eta\varrho_m}^m,\end{aligned} \tag{57}$$

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 $\tilde{t}_{\lambda\varrho_m}^m$ by

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

The following proposition gives the asymptotic null distribution of the LRTS for testing $H_0:M=M_0$. Under the stated assumptions, the log-likelihood function permits a quadratic approximation in the neighborhood of Υ_m^* similar to the one in Proposition 8. Define $A_{n\varepsilon c}^m(\xi):=\{\vartheta_{M_0+1}\in\Theta_{M_0+1}:\{\ell_n(\psi_m,\pi_m,\xi)-\ell_n(\psi_m^*,\pi_m,\xi)\geq 0\}\wedge|t(\psi_m,\pi_m)|<\varepsilon\}\cup\mathcal{N}_{c/\sqrt{n}}$. Under $H_0:M=M_0$, for all $c>0$, for $m=1,\ldots,M_0$, and uniformly in $\xi\in\Xi$ and $\vartheta_{M_0+1}\in A_{n\varepsilon c}^m(\xi)$,

$$\ell_n(\psi_m,\pi_m,\xi)-\ell_n(\psi_m^*,\pi_m,\xi)\ -\sqrt{n}t(\psi_m,\pi_m)'\nu_n(s_{\varrho_m k})+nt(\psi_m,\pi_m)'\mathcal{I}_{\varrho_m}t(\psi_m,\pi_m)/2=o_{p\varepsilon}(1),$$

where $s_{\varrho_m k} := (\tilde{s}'_{\eta k}, (s^m_{\lambda \varrho_m k})')'$ and $\mathcal{I}_{\varrho_m} = \lim_{k\to\infty} \mathbb{E}_{\vartheta^*_{M_0}}(s_{\varrho_m k} s'_{\varrho_m k})$. 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_{0m}. Denote the parameter space for ϱ_m by Θ_{ϱ_m}, and let $\tilde{\Theta}_\varrho := \Theta_{\varrho_1} \times \ldots \times \Theta_{\varrho_{M_0}}$.

Assumption 10. *$0 < \inf_{\tilde{\varrho}\in\tilde{\Theta}_\varrho} \lambda_{\min}(\tilde{\mathcal{I}}_{\tilde{\varrho}}) \le \sup_{\tilde{\varrho}\in\tilde{\Theta}_\varrho} \lambda_{\max}(\tilde{\mathcal{I}}_{\tilde{\varrho}}) < \infty$ 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 (55).*

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

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 (25). Consider the following reparameterization similar to (26):

$$\begin{pmatrix} \zeta_m \\ \zeta_{m+1} \\ \sigma^2_m \\ \sigma^2_{m+1} \end{pmatrix} = \begin{pmatrix} \nu_{\zeta m} + (1-\alpha_m)\lambda_{\zeta m} \\ \nu_{\zeta m} - \alpha_m \lambda_{\zeta m} \\ \nu_{\sigma m} + (1-\alpha_m)(2\lambda_{\sigma m} + C_1 \lambda^2_{\mu m}) \\ \nu_{\sigma m} - \alpha_m(2\lambda_{\sigma m} + C_2\lambda^2_{\mu m}) \end{pmatrix},$$

where $\nu_{\zeta m} = (\nu_\mu, \nu'_\beta)'$, $\lambda_{\zeta m} = (\lambda_{\mu m}, \lambda'_{\beta m})'$, $C_1 := -(1/3)(1+\alpha_m)$, and $C_2 := (1/3)(2-\alpha_m)$. As in Section 7.1, we collect the reparameterized identified parameters into $\psi_m := (\eta'_m, \lambda'_m)'$, where $\eta_m = (\gamma', \{\theta'_j\}^{m-1}_{j=1}, \nu'_{\zeta m}, \nu_{\sigma m}, \{\theta'_j\}^{M_0+1}_{j=m+2}, \vartheta'_{xm})'$ and $\lambda_m := (\lambda'_{\zeta m}, \lambda_{\sigma m})'$. Similar to (27), define the reparameterized conditional density of y_k as

$$\begin{aligned} g_{\psi_m}(y_k|\overline{\mathbf{y}}_{k-1}, x_k) &= \sum_{j\in \overline{J}_m} q_{kj} f(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \theta_j) \\ &+ \mathbb{I}\{x_k \in J_m\} f\left(y_k|\overline{\mathbf{y}}_{k-1}; \gamma, \nu_{\zeta m} + (q_{km}-\alpha_m)\lambda_{\zeta m}, \nu_{\sigma m} + (q_{km}-\alpha_m)(2\lambda_{\sigma m} + (C_2 - q_{km})\lambda^2_{\mu m})\right). \end{aligned}$$

Let g^*_{mk}, f^*_{mk}, ∇g^*_{mk}, and ∇f^*_{mk} denote $g_{\psi^*_m}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$, $f(Y_k|\overline{\mathbf{Y}}_{k-1}; \gamma^*, \theta^*_m)$, $\nabla g_{\psi^*_m}(Y_k|\overline{\mathbf{Y}}_{k-1}, X_k)$, and $\nabla f(Y_k|\overline{\mathbf{Y}}_{k-1}; \gamma^*, \theta^*_m)$. From (29) and a derivation similar to (51), we obtain the following result that corresponds to (30) in testing homogeneity:

$$\begin{aligned} \mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda^i_{\mu m}} g^*_{mk}/g^*_{mk} \middle| \overline{\mathbf{Y}}^k_{-\infty}\right] &= 0, \quad i = 1,2,3, \\ \mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda^4_{\mu m}} g^*_{mk}/g^*_{mk} \middle| \overline{\mathbf{Y}}^k_{-\infty}\right] &= \alpha_m(1-\alpha_m)b(\alpha_m)(\nabla_{\mu^4} f^*_{mk}/f^*_{mk})\mathbb{P}_{\vartheta^*_{M_0}}(X_k \in J_m|\overline{\mathbf{Y}}^k_{-\infty}) \\ &= b(\alpha_m)\mathbb{E}_{\vartheta^*_{M_0}}\left[\nabla_{\lambda^2_{\sigma m}} g^*_{mk}/g^*_{mk} \middle| \overline{\mathbf{Y}}^k_{-\infty}\right]. \end{aligned} \tag{58}$$

Repeating the calculation leading to (52)–(54) and using (58) gives the following: first, (52) and (53) still hold; second, the elements of $\nabla_{\lambda_m \lambda'_m} \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}^{k-1}_0)/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}^{k-1}_0)$ except for the $(1,1)$th element are given by (54) while adjusting the derivative with respect to $\lambda_{\sigma m}$ by multiplying by 2; third,

$$\frac{\nabla_{\lambda_{\mu m}^2}\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi_m^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \alpha_m(1-\alpha_m)\sum_{t=1}^{k-1}\varrho_m^{k-t}\left(2\frac{\nabla_\mu f_{mt}^*}{f_{mt}^*}\frac{\nabla_\mu f_{mk}^*}{f_{mk}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_m^2|\overline{\mathbf{Y}}_0^k).$$

For $m \geq 0$, define $\zeta_{k,m}^m(\varrho_m) := \sum_{t=-m+1}^{k-1}\varrho_m^{k-t-1}2(\nabla_\mu f_{mt}^*\nabla_\mu f_{mk}^*/f_{mt}^* f_{mk}^*)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_m^2|\overline{\mathbf{Y}}_0^k)$. Similarly to (35), define the elements of the generalized score as

$$\begin{aligned}\begin{pmatrix} * & s_{\lambda_{\mu\beta}\varrho_m k}^m & s_{\lambda_{\mu\sigma}\varrho_m k}^m \\ s_{\lambda_{\beta\mu}\varrho_m k}^m & s_{\lambda_{\beta\beta}\varrho_m k}^m & s_{\lambda_{\beta\sigma}\varrho_m k}^m \\ s_{\lambda_{\sigma\mu}\varrho_m k}^m & s_{\lambda_{\beta\sigma}\varrho_m k}^m & s_{\lambda_{\sigma\sigma}\varrho_m k}^m \end{pmatrix} &:= \frac{\nabla_{\theta\theta'}f_{mk}^*}{f_{mk}^*}\mathbb{P}_{\vartheta_{M_0}^*}(X_k\in J_m|\overline{\mathbf{Y}}_0^k) \\ &+\sum_{t=1}^{k-1}\varrho_m^{k-t}\left(\frac{\nabla_\theta f_{mt}^*}{f_{mt}^*}\frac{\nabla_{\theta'}f_{mk}^*}{f_{mk}^*}+\frac{\nabla_\theta f_{mk}^*}{f_{mk}^*}\frac{\nabla_{\theta'}f_{mt}^*}{f_{mt}^*}\right)\mathbb{P}_{\vartheta_{M_0}^*}((X_t,X_k)\in J_m^2|\overline{\mathbf{Y}}_0^k).\end{aligned} \tag{59}$$

Similarly to (36), define $\tilde{s}_{\tilde{\varrho}k}$ as in (55) with redefining $s_{\lambda\varrho_m k}^m$ in (55) as

$$s_{\lambda\varrho_m k}^m := \left(\zeta_{k,0}^m(\varrho_m)/2 \quad 2s_{\lambda_{\mu\sigma}\varrho_m k}^m \quad 2s_{\lambda_{\sigma\sigma}\varrho_m k}^m \quad (s_{\lambda_{\beta\mu}\varrho_m k}^m)' \quad 2(s_{\lambda_{\beta\sigma}\varrho_m k}^m)' \quad V(s_{\lambda_{\beta\beta}\varrho_h k}^m)'\right)'. \tag{60}$$

Define $\mathcal{I}_{\lambda.\eta\varrho_m}^m$ and $Z_{\lambda\varrho_m}^m$ as in (57) with $s_{\lambda\varrho_m k}^m$ defined in (60). Let $Z_{\lambda 0}^m$ and $\mathcal{I}_{\lambda.\eta 0}^m$ denote $Z_{\lambda\varrho_m}^m$ and $\mathcal{I}_{\lambda.\eta\varrho_m}^m$ evaluated at $\varrho_m = 0$. Define Λ_λ^1 as in (38), and define $\Lambda_{\lambda\varrho_m}^2$ as in (38) with replacing ϱ with ϱ_m. Similar to (39), define $\tilde{t}_\lambda^{m1}$ and $\tilde{t}_{\lambda\varrho_m}^{m2}$ by $r_\lambda(\tilde{t}_\lambda^{m1}) = \inf_{t_\lambda\in\Lambda_\lambda^1} r_\lambda^m(t_\lambda)$ and $r_{\lambda\varrho_m}(\tilde{t}_{\lambda\varrho_m}^{m2}) = \inf_{t_\lambda\in\Lambda_{\lambda\varrho_m}^2} r_{\lambda\varrho_m}^m(t_\lambda)$, where $r_\lambda^m(t_\lambda) := (t_\lambda - Z_{\lambda 0}^m)'\mathcal{I}_{\lambda.\eta 0}^m(t_\lambda - Z_{\lambda 0}^m)$ and $r_{\lambda\varrho_m}^m(t_\lambda) := (t_\lambda - Z_{\lambda\varrho_m}^m)'\mathcal{I}_{\lambda.\eta\varrho_m}^m(t_\lambda - Z_{\lambda\varrho_m}^m)$.

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 $\tilde{s}_{\tilde{\varrho}_m k}$ is given in (60).*

Proposition 16. *Suppose Assumptions 1, 2, 4, 9, and 11 hold and the component density for the j-th regime is given by (25). Then, under $H_0 : m = M_0$, $2[\ell_n(\hat{\vartheta}_{M_0+1},\xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0},\xi_{M_0})] \xrightarrow{d}$ $\max_{m=1,\ldots,M_0}\{\sup_{\varrho_m\in\Theta_\varrho^m}\max\{\mathbb{I}\{\varrho_m = 0\}(\tilde{t}_\lambda^{m1})'\mathcal{I}_{\lambda.\eta 0}^m\tilde{t}_\lambda^{m1}, (\tilde{t}_{\lambda\varrho_m}^{m2})'\mathcal{I}_{\lambda.\eta\varrho_m}^m\tilde{t}_{\lambda\varrho_m}^{m2}\}\}$.*

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 (40).

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

$$\begin{pmatrix} \theta_m \\ \theta_{m+1} \\ \sigma^2 \end{pmatrix} = \begin{pmatrix} \nu_{\theta m} + (1-\alpha_m)\lambda_m \\ \nu_{\theta m} - \alpha_m \lambda_m \\ \nu_{\sigma m} - \alpha_m(1-\alpha_m)\lambda_{\mu m}^2 \end{pmatrix},$$

similar to (41) and following the derivation in Sections 6.3 and 7.2. For brevity, we omit details in derivation. Define $s^m_{\lambda\lambda\varrho_m k}$ as in (56), and denote each element of $s^m_{\lambda\lambda\varrho_m k}$ as

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

Similarly to (46), define $\tilde{s}_{\tilde{\varrho}k}$ as in (55) with redefining $s^m_{\lambda\varrho_m k}$ in (55) as

$$s^m_{\lambda\varrho_m k} := \Big(\zeta^m_{k,0}(\varrho_m)/2 \quad s^m_{\lambda^3_\mu k}/3! \quad s^m_{\lambda^4_\mu k}/4! \quad (s^m_{\lambda_{\beta\mu}\varrho k})' \quad V(s^m_{\lambda_{\beta\beta}\varrho k})'\Big)', \tag{61}$$

where $s^m_{\lambda^i_\mu k} := \mathbb{P}_{\vartheta^*_{M_0}(X_k\in J_m|\overline{\mathbf{Y}}^k_0)} \nabla_{\mu^i} f(Y_k|\overline{\mathbf{Y}}_{k-1};\gamma^*,\theta^*_m)/f(Y_k|\overline{\mathbf{Y}}_{k-1};\gamma^*,\theta^*_m)$ for $i=3,4$.

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

Assumption 12. *Assumption 10 holds when $\tilde{s}_{\tilde{\varrho}_m k}$ is given in (61).*

Proposition 17. *Suppose Assumptions 1, 2, 4, 9, and 12 hold and the component density for the j-th regime is given by (40). Then, under $H_0: m = M_0$, $2[\ell_n(\hat{\vartheta}_{M_0+1},\xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0},\xi_{M_0})] \xrightarrow{d} \max_{m=1,\dots,M_0}\{\sup_{\varrho_m\in\Theta_{\varrho_m\epsilon}} \max\{\mathbb{I}\{\varrho_m=0\}(\tilde{t}^{m1}_\lambda)'\mathcal{I}^m_{\lambda.\eta 0}\tilde{t}^{m1}_\lambda, (\tilde{t}^{m2}_{\lambda\varrho_m})'\mathcal{I}^m_{\lambda.\eta\varrho_m}\tilde{t}^{m2}_{\lambda\varrho_m}\}\}$, where $\tilde{t}^{m1}_\lambda$ and $\tilde{t}^{m2}_{\lambda\varrho_m}$ are defined as in Proposition 16 but in terms of $(Z^m_{\lambda\varrho_m}, \mathcal{I}^m_{\lambda.\eta\varrho_m}, Z^m_{\lambda 0}, \mathcal{I}^m_{\lambda.\eta 0})$ constructed with $s^m_{\lambda\varrho_m k}$ given in (61) and Λ^1_λ and $\Lambda^2_{\lambda\varrho_m}$ defined as in (47) but replacing ϱ with ϱ_m.*

8 Asymptotic distribution under local alternatives

In this section, we derive the asymptotic distribution of our LRTS under local alternatives. We focus on the case of testing $H_0: M=1$ against $H_1: M=2$, but it is straightforward to extend the analysis to the case of testing $H_0: M=M_0$ against $H_1: M=M_0+1$ for $M_0\geq 2$.

Given $\pi\in\Theta_\pi$, we define a local parameter $h := \sqrt{n}t(\psi,\pi)$ so that

$$h = \begin{pmatrix} h_\eta \\ h_\lambda \end{pmatrix} = \begin{pmatrix} \sqrt{n}(\eta-\eta^*) \\ \sqrt{n}t_\lambda(\lambda,\pi) \end{pmatrix},$$

where $t_\lambda(\lambda,\pi)$ differs across different models and is given by (21), (34), and (45). Given $h = (h'_\eta, h'_\lambda)'$ and $\pi\in\Theta_\pi$, we consider the sequence of contiguous local alternatives $\vartheta_n = (\psi'_n,\pi'_n)' = (\eta'_n,\lambda'_n,\pi'_n)' \in \Theta_\eta\times\Theta_\lambda\times\Theta_\pi$ such that

$$h_\eta = \sqrt{n}(\eta_n-\eta^*), \quad h_\lambda = \sqrt{n}t_\lambda(\lambda_n,\pi_n)+o(1), \quad \text{and } \pi_n-\pi = o(1). \tag{62}$$

Let $\mathbb{P}^n_{\vartheta,x_0}$ be the probability measure on $\{Y_k\}_{k=1}^n$ under ϑ conditional on the value of $\overline{\mathbf{Y}}_0$, X_0, and $\mathbf{W}_1^n$. Then, the log likelihood ratio is given by

$$\log \frac{d\mathbb{P}^n_{\vartheta_n,x_0}}{d\mathbb{P}^n_{\vartheta^*,x_0}} = \ell_n(\psi_n,\pi_n,x_0) - \ell_n(\psi^*,\pi,x_0) = \log\left(\frac{\sum_{\mathbf{x}_1^n}\prod_{k=1}^n f_k(\eta_n,\lambda_n)q_{\pi_n}(x_{k-1},x_k)}{\prod_{k=1}^n f_k(\eta^*,0)}\right),$$

where $f_k(\eta,\lambda)$ is defined by the right hand side of (12), (27), and (42) for the models of non-normal distribution, heteroscedastic normal distribution, and homoscedastic normal distribution, respectively. The following result is useful for deriving the asymptotic distribution of the LRTS under $\mathbb{P}^n_{\vartheta_n,x_0}$.

Proposition 18. *Suppose that the assumptions of Propositions 8, 10, and 12 hold for the models of non-normal, heteroscedastic normal, and homoscedastic normal distributions, respectively. Then, uniformly in* $x_0 \in \mathcal{X}$, *(a)* $\mathbb{P}^n_{\vartheta_n,x_0}$ *is mutually contiguous with respect to* $\mathbb{P}^n_{\vartheta^*,x_0}$, *and (b) under* $\mathbb{P}^n_{\vartheta_n,x_0}$, *we have* $\log(d\mathbb{P}^n_{\vartheta_n,x_0}/d\mathbb{P}^n_{\vartheta^*,x_0}) = h'\nu_n(s_{\varrho_n k}) - \frac{1}{2}h'\mathcal{I}_\varrho h + o_p(1)$ *with* $\nu_n(s_{\varrho_n k}) \xrightarrow{d} N(\mathcal{I}_\varrho h, \mathcal{I}_\varrho)$.

This result follows from Le Cam's first and third lemma. Using the result of Proposition 18, we construct the asymptotic distribution of LRTS under the sequence of local alternatives from null asymptotic distribution of LRTS by appropriately shifting the mean of the Gaussian process.

8.1 Non-normal distribution

For non-normal distribution, the sequence of contiguous local alternatives is given by $\lambda_n = \bar{\lambda}/n^{1/4}$ because then $h_\lambda = \sqrt{n}\alpha(1-\alpha)v(\lambda_n) = \alpha(1-\alpha)v(\bar{\lambda})$ holds. The following proposition derives the asymptotic distribution of LRTS for non-normal distribution under $H_{1n} : (\pi_n,\eta_n,\lambda_n) = (\bar{\pi},\eta^*,\bar{\lambda}/n^{1/4})$.

Proposition 19. *Suppose that the assumptions of Proposition 9 hold. For* $\bar{\pi} \in \Theta_\pi$ *and* $\bar{\lambda} \neq 0$, *define* $h_\lambda := \bar{\alpha}(1-\bar{\alpha})v(\bar{\lambda})$. *Then, under* $H_{1n} : (\pi_n,\eta_n,\lambda_n) = (\bar{\pi},\eta^*,\bar{\lambda}/n^{1/4})$, *we have* $2[\ell_n(\hat{\vartheta}_2,\xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho\in\Theta_\varrho}(\tilde{t}_{\lambda\varrho h})'\mathcal{I}_{\lambda.\eta\varrho}\tilde{t}_{\lambda\varrho h}$, *where* $\tilde{t}_{\lambda\varrho h}$ *is defined as in (24) but replacing* $Z_{\lambda\varrho}$ *in (24) with* $(\mathcal{I}_{\lambda.\eta\varrho})^{-1}G_{\lambda.\eta\varrho} + h_\lambda$.

8.2 Heteroscedastic normal distribution

For the model with heteroscedastic normal distribution, the sequences of contiguous local alternatives characterized by (62) include the local alternatives of order $n^{-1/8}$.

Proposition 20. *Suppose that the assumptions of Proposition 11 hold for the model (25). For* $\bar{\varrho} \in (-1,1)$, $\bar{\alpha} \in (0,1)$, *and* $\bar{\lambda} := (\bar{\lambda}_\mu, \bar{\lambda}_\sigma, \bar{\lambda}_\beta')' \neq (0,0,0)'$, *let*

$$H_{1n}^a : (\varrho_n,\alpha_n,\eta_n,\lambda_{\mu n},\lambda_{\sigma n},\lambda_{\beta n}) = (\bar{\varrho}/n^{1/4},\bar{\alpha},\eta^*,\bar{\lambda}_\mu/n^{1/8},\bar{\lambda}_\sigma/n^{3/8},\bar{\lambda}_\beta/n^{3/8}),$$
$$H_{1n}^b : (\varrho_n,\alpha_n,\eta_n,\lambda_{\mu n},\lambda_{\sigma n},\lambda_{\beta n}) = (\bar{\varrho},\bar{\alpha},\eta^*,\bar{\lambda}_\mu/n^{1/4},\bar{\lambda}_\sigma/n^{1/4},\bar{\lambda}_\beta/n^{1/4}),$$

and define

$$h_{\lambda}^{a} := \bar{\alpha}(1-\bar{\alpha}) \times (\bar{\varrho}\bar{\lambda}_{\mu}^{2}, \bar{\lambda}_{\mu}\bar{\lambda}_{\sigma}, b(\bar{\alpha})\bar{\lambda}_{\mu}^{4}/12, \bar{\lambda}_{\beta}'\bar{\lambda}_{\mu}, 0, 0)',$$
$$h_{\lambda}^{b} := \bar{\alpha}(1-\bar{\alpha}) \times (\bar{\varrho}\bar{\lambda}_{\mu}^{2}, \bar{\lambda}_{\mu}\bar{\lambda}_{\sigma}, \bar{\lambda}_{\sigma}^{2}, \bar{\lambda}_{\beta}'\bar{\lambda}_{\mu}, \bar{\lambda}_{\beta}'\bar{\lambda}_{\sigma}, v(\bar{\lambda}_{\beta})')'.$$

Then, for $j \in \{a,b\}$*, under* H_{1n}^{j}*, we have* $2[\ell_n(\hat{\vartheta}_2,\xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho\in\Theta_\varrho} \max\{\mathbb{I}\{\varrho = 0\}(\tilde{t}_{\lambda h}^{1j})'\mathcal{I}_{\lambda.\eta 0}\tilde{t}_{\lambda h}^{1j}, (\tilde{t}_{\lambda\varrho h}^{2j})'\mathcal{I}_{\lambda.\eta\varrho}\tilde{t}_{\lambda\varrho h}^{2j}\}$*, where* $\tilde{t}_{\lambda h}^{1j}$ *and* $\tilde{t}_{\lambda\varrho h}^{2j}$ *are defined as in (39) but replacing* $Z_{\lambda\varrho}$ *with* $(\mathcal{I}_{\lambda.\eta\varrho})^{-1}G_{\lambda.\eta\varrho} + h_{\lambda}^{j}$.

In the local alternative H_{1n}^{a}, ϱ_n converges to 0, and $\lambda_{\mu n}$ converges to 0 at a slower rate than $n^{-1/4}$. Our test has non-trivial power against these local alternatives in the neighborhood of $\varrho = 0$. In contrast, the test of Carrasco et al. (2014) does not have power against the local alternatives in the neighborhood of $\varrho = 0$ as discussed in Section 5 of Carrasco et al. (2014). The test proposed by Qu and Zhuo (2017) assumes that ϱ is bounded away from zero and hence their test rules out H_{1n}^{a}.

8.3 Homoscedastic normal distribution

The local alternatives for the model with homoscedastic distribution also include those of order $n^{-1/8}$ in the neighborhood of $\varrho = 0$.

Proposition 21. *Suppose that the assumptions of Proposition 12 hold for the model (40). For* $\bar{\varrho} \in (-1,1)$*,* $\bar{\alpha} \in (0,1)$*,* $\Delta_\alpha \neq 0$*, and* $\bar{\lambda} := (\bar{\lambda}_\mu, \bar{\lambda}_\beta')' \neq (0,0)'$*, let*

$$H_{1n}^{a} : (\varrho_n, \alpha_n, \eta_n, \lambda_{\mu n}, \lambda_{\beta n}) = (\bar{\varrho}/n^{1/4}, 1/2 + \Delta_\alpha/n^{1/8}, \eta^*, \bar{\lambda}_\mu/n^{1/8}, \bar{\lambda}_\beta/n^{3/8}),$$
$$H_{1n}^{b} : (\varrho_n, \alpha_n, \eta_n, \lambda_{\mu n}, \lambda_{\beta n}) = (\bar{\varrho}, \bar{\alpha}, \eta^*, \bar{\lambda}_\mu/n^{1/4}, \bar{\lambda}_\beta/n^{1/4}),$$

and define $h_{\lambda}^{a} := (1/4)\times(\bar{\varrho}\bar{\lambda}_{\mu}^{2}, \Delta_\alpha\bar{\lambda}_{\mu}^{3}, -\bar{\lambda}_{\mu}^{4}/2, \bar{\lambda}_{\beta}'\bar{\lambda}_{\mu}, 0)'$ *and* $h_{\lambda}^{b} := \bar{\alpha}(1-\bar{\alpha})\times(\bar{\varrho}\bar{\lambda}_{\mu}^{2}, 0, 0, \bar{\lambda}_{\beta}'\bar{\lambda}_{\mu}, v(\bar{\lambda}_{\beta})')'$*. For* $j = \{a,b\}$*, define* $\tilde{t}_{\lambda h}^{1j}$ *and* $\tilde{t}_{\lambda\varrho h}^{2j}$ *as in (39) but replacing* $Z_{\lambda\varrho}$ *with* $(\mathcal{I}_{\lambda.\eta\varrho})^{-1}G_{\lambda.\eta\varrho} + h_{\lambda}^{j}$*, where* $\mathcal{I}_{\lambda.\eta\varrho}$ *and* $G_{\lambda.\eta\varrho}$ *are constructed with* $s_{\varrho k}$ *defined in (46), and* Λ_{λ}^{1} *and* $\Lambda_{\lambda\varrho}^{2}$ *are defined in (47). Then, under* H_{1n}^{j}*, we have* $2[\ell_n(\hat{\vartheta}_2,\xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho\in\Theta_\varrho} \max\{\mathbb{I}\{\varrho = 0\}(\tilde{t}_{\lambda h}^{1j})'\mathcal{I}_{\lambda.\eta 0}\tilde{t}_{\lambda h}^{1j}, (\tilde{t}_{\lambda\varrho h}^{2j})'\mathcal{I}_{\lambda.\eta\varrho}\tilde{t}_{\lambda\varrho h}^{2j}\}$.

9 Parametric bootstrap

We consider the following parametric bootstrap to obtain the bootstrap critical value $c_{\alpha,B}$ and the bootstrap p-value of our LRTS for testing $H_0 : M = M_0$ against $H_1 : M = M_0 + 1$.

1. Using the observed data, estimate $\hat{\vartheta}_{M_0}$ and $\hat{\vartheta}_{M_0+1}$ as $\hat{\vartheta}_M := \arg\max_{\vartheta_M\in\Theta_M} \ell_n(\vartheta_M, \xi_M)$ for some choice of ξ_M for $M = M_0, M_0+1$. Compute $LR_n = 2[\ell_n(\hat{\vartheta}_{M_0+1}, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})]$.

2. Given $\hat{\vartheta}_{M_0}$ and ξ_{M_0}, generate B independent samples $\{Y_1^b, \ldots, Y_n^b\}_{b=1}^{B}$ under H_0 with $\vartheta_{M_0} = \hat{\vartheta}_{M_0}$ conditional on the observed value of $\overline{\mathbf{Y}}_0$ and $\mathbf{W}_1^n$.

3. For each simulated sample $\{Y_k^b\}_{k=1}^n$ with $(\overline{\mathbf{Y}}_0, \mathbf{W}_1^n)$, estimate $\hat{\vartheta}_{M_0}^b$ and $\hat{\vartheta}_{M_0+1}^b$ as in Step 1, and let $LR_n^b = 2[\ell_n(\hat{\vartheta}_{M_0+1}^b, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}^b, \xi_{M_0})]$ for $b = 1, \ldots, B$.

4. Let $c_{\alpha,B}$ be the $(1-\alpha)$ quantile of $\{LR_n^b\}_{b=1}^B$, and define the bootstrap p-value as $B^{-1}\sum_{b=1}^B \mathbb{I}\{LR_n^b > LR_n\}$.

The following proposition shows the consistency of the bootstrap critical values $c_{\alpha,B}$ for testing $H_0 : M_0 = 1$. We omit the result for testing $H_0 : M_0 \geq 2$; it is straightforward to extend the analysis to the case for $M_0 \geq 2$ with more tedious notations.

Proposition 22. *Suppose that the assumptions of Propositions 8, 10, and 12 hold for the models of non-normal, heteroscedastic normal, and homoscedastic normal distributions, respectively. Then, the bootstrap critical values $c_{\alpha,B}$ converge to the asymptotic critical values in probability as n and B go to infinity under H_0 and under the local alternatives described in Propositions 19, 20, and 21.*

10 Simulations and Empirical Application

10.1 Simulations

We consider the following two models:

$$\text{Model 1}: \quad Y_k = \mu_{X_k} + \beta Y_{k-1} + \varepsilon_k, \quad \varepsilon_k \sim N(0, \sigma^2), \tag{63}$$

$$\text{Model 2}: \quad Y_k = \mu_{X_k} + \beta Y_{k-1} + \varepsilon_k, \quad \varepsilon_k \sim N(0, \sigma_{X_k}^2), \tag{64}$$

where $X_k \in \{1, \ldots, M\}$ with $p_{ij} = p(X_k = i | X_{k-1} = j)$. Model 1 in (63) is a model with switching intercept, where variance parameter σ^2 does not switch across regimes. In Model 2 in (64), both intercept and variance parameters switch across regimes.

We investigate the size and power property of our bootstrap LRT and compare the LRT with the QLR test of Cho and White (2007) and supTS test of Carrasco et al. (2014), where the critical values are computed by bootstrap. In the supTS test, we set $\rho \in [-0.9.0.9]$, and in the QLR test, we set the parameter set $\Theta_\mu = [-2, 2]$. Note that this comparison favors the LRT over the supTS test because the supTS test is designed to detect general parameter variation including Markov chain.

We first examine the rejection frequency of $H_0 : M = 1$ against $H_1 : M = 2$ when the data are generated by $H_0 : M = 1$ with $(\beta, \mu, \sigma) = (0.5, 0, 1)$. Columns (1) and (2) in Table 1 report the rejection frequency of the bootstrap tests at the nominal 5% level over 3000 replications with $n = 200$ and 500. Overall, our bootstrap LRT has good sizes.

Table 2 examines the power of our bootstrap LRT for testing the null hypothesis of $M = 1$ at nominal level of 5%. We generate 3000 data sets for $n = 500$ under the alternative hypothesis of $M = 2$ by setting $\mu_1 = 0.2, 0.6$, and 1.0 and $\mu_2 = -\mu_1$ while $(p_{11}, p_{22}) = (0.25, 0.25)$, $(0.50, 0.50)$, $(0.70, 0.70)$, and $(0.90, 0.90)$. We set $\sigma = 1$ for Model 1 and $(\sigma_1^2, \sigma_2^2) = (1.1, 0.9)$ for Model 2.

In Table 2, our LRT performs better than the supTS and QLR tests for Model 1 except for the case with $(p_{11}, p_{22}) = (0.25, 0.25)$, where the supTS performs very well, and the case with $(p_{11}, p_{22}) = (0.5, 0.5)$, where the QLRT test outperforms LRT, because the true dgp is finite mixture in this case.

The last three columns of Table 2 reports the power of the LRT to detect alternative models with switching variances (i.e., Model 2 with $M = 2$).

We also examine the power of our LRT for testing the null hypothesis of $M = 2$ in Table 3. We generate 1000 data sets for $n = 500$ under the alternative hypothesis of $M = 3$ across different values of (μ_1, μ_2, μ_3) and (p_{11}, p_{22}, p_{33}) with $p_{ij} = (1 - p_{ii})/2$ for $j \neq i$, where we set $(\beta, \sigma) = (0.5, 1.0)$ for Model 1 and $(\beta, \sigma_1, \sigma_2) = (0.5, 0.9, 1.2)$ for Model 2, and compute the rejection frequencies for testing the null hypothesis of $M = 2$ at nominal level of 5%. In Table 3, the powers of our LRT for testing $H_0 : M = 2$ against $H_1 : M = 3$ increase when the alternative is further away from H_0 or when latent regimes become more persistent.

11 Empirical example

Using the U.S. GDP per capita quarterly growth rate data from 1960Q1 to 2014Q4, we estimate the regime switching models with common variance (i.e., Model 1 in (63)) and with switching variances (i.e., Model 2 in (64)) for $M = 1$, 2, 3, and 4 and sequentially test the null hypothesis of $M = M_0$ against the alternative hypothesis $M = M_0 + 1$ for $M_0 = 1$, 2, 3, and 4.[6] We also report the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) as a reference although, to our best knowledge, the consistency of AIC and BIC for selecting the number of regimes has not been established in the existing literature.

Table 5 reports the result of the selected number of regimes by AIC, BIC, and our LRT. For the model (63) with common variance, our LRT selects $M = 4$ while AIC and BIC select $M = 3$ and $M = 1$, respectively. For the model (64) with switching variance, both our LRT and AIC select $M = 3$ while BIC selects $M = 2$.

Panel A of Table 4 and Figure 1 report the estimated parameter values and the posterior probabilities of being each regime for the model with common variance for $M = 2$, 3, and 4. Across different specifications in M, the estimated values of μ_1, μ_2, ..., μ_M are well separated in the common variance model, indicating that each regime represents booming or thorough period with different degrees. In Figure 1, when the number of regimes is specified as $M = 2$, the posterior probability of "recession" regime (Regime 1) against that of "booming" regime (Regime 2) sharply rises during the collapse of Lehman Brothers in 2008 and then declines after 2009. When the number of regimes is specified as $M = 3$, in addition to "recession" and "booming" regimes corresponding to Regime 1 and 2, respectively, the regime with a rapid change in the growth rates from low to high is captured by Regime 3; for the model with $M = 3$ in Figure 1, the posterior probability of Regime 3 rises in

[6] For both models, we restrict the parameter values for transition probabilities by setting $\epsilon = 0.05$ to prevent the issue of unbounded likelihood.

late 2009 when the U.S. economy started to recover from the Lehman shock. When the number of regimes is specified as $M = 4$, Regime 1 now captures a rapid change in the growth rates from high to low, where the posterior probability of Regime 1 becomes high when the growth rate of the U.S. economy rapidly declined in the middle of the Lehman shock. Our LRT selects the model with four regimes, which capture rapid changes in growth rates of the U.S. GDP per capita during the Lehman shock period.

The estimated parameter values and the posterior probabilities of being each regime for the model with switching variance are reported in Panel B of Table 4 and Figure 2, respectively. When the number of regimes is specified as $M = 2$ in the switching variance model, the estimated values of variance parameter are very different between two regimes while the estimated intercept values are similar, indicating that Regime 1 is "low volatility" regime while Regime 2 is "high volatility" regime.[7] When the number of regimes is specified as $M = 3$, different regimes capture different states of the U.S. economy in terms of both growth rates and volatilities. Regime 1 is characterized by the negative value of intercept with high volatility, capturing a recession period. Regime 2 is characterized by the positive value of intercept with low volatility, capturing booming/stable economy. Regime 3 is characterized by high value of intercept and high value of variance, capturing both a rapid recovery in the growth rates and high volatility in the aftermath of the Lehman shock in 2009. Our LRT selects the model with three regimes when the model is specified with switching variance.

12 Appendix

Henceforth, for notational brevity, we suppress $\mathbf{W}_a^b$ from the conditioning variables and conditional densities when doing so does not cause confusion.

12.1 Proof of Propositions and Corollaries

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 $\mathbf{W}_1^n$, but our Lemma 10(a) extends Corollary 1 of DMR to accommodate W_k's. Consequently, the argument of the proof of DMR goes through. $\square$

Proof of Proposition 2. Define $h_{\vartheta k x_0} := \sqrt{l_{\vartheta k x_0}} - 1$. Using the Taylor expansion of $2\log(1 + x) =$

[7] We may test the null hypothesis of $\sigma_1 = \sigma_2$ in the model with switching variance given $M = 2$ by standard LRT with the critical value obtained from the chi-square distribution with 2 degrees of freedom. With $LRT = 2 \times (-307.99 + 321.27) = 26.56$, the null hypothesis of $\sigma_1 = \sigma_2$ is rejected at 1 percent significance level, suggesting that the model with switching variance is more appropriate than the model with common variance when we specified the number of regimes as $M = 2$.

$2x - x^2(1+o(1))$ for small x, we have, uniformly in $x_0 \in \mathcal{X}$ and $\vartheta \in \mathcal{N}_{c/\sqrt{n}}$,

$$\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) = 2\sum_{k=1}^{n} \log(1 + h_{\vartheta k x_0}) = nP_n(2h_{\vartheta k x_0} - [1 + o_p(1)]h_{\vartheta k x_0}^2). \tag{65}$$

The stated result holds if we show that

$$\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} \left| nP_n(h_{\vartheta k x_0}^2) - nt_\vartheta' \mathcal{I}_\pi t_\vartheta / 4 \right| = o_p(1) \quad \text{and} \tag{66}$$

$$\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} |nP_n(h_{\vartheta k x_0}) - \sqrt{n} t_\vartheta' \nu_n(s_{\pi k})/2 + n t_\vartheta \mathcal{I}_\pi t_\vartheta'/8| = o_p(1), \tag{67}$$

because then the right hand side of (65) is equal to $\sqrt{n} t_\vartheta' \nu_n(s_{\pi k}) - t_\vartheta \mathcal{I}_\pi t_\vartheta'/2 + o_p(1)$ uniformly in $x_0 \in \mathcal{X}$ and $\vartheta \in \mathcal{N}_{c/\sqrt{n}}$.

We first show (66). 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

$$\max_{1 \le k \le n} \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} |m_{\vartheta k}| = \max_{1 \le k \le n} \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} |t_\vartheta' s_{\pi k} + r_{\vartheta k}| = o_p(1), \tag{68}$$

from Assumption 3(a)(c) and Lemma 9. Write $4P_n(h_{\vartheta k x_0}^2)$ as

$$4P_n(h_{\vartheta k x_0}^2) = P_n\left(\frac{4(l_{\vartheta k x_0} - 1)^2}{(\sqrt{l_{\vartheta k x_0}} + 1)^2}\right) = P_n(l_{\vartheta k x_0} - 1)^2 - P_n\left((l_{\vartheta k x_0} - 1)^3 \frac{(\sqrt{l_{\vartheta k x_0}} + 3)}{(\sqrt{l_{\vartheta k x_0}} + 1)^3}\right). \tag{69}$$

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

$$P_n(l_{\vartheta k x_0} - 1)^2 = t_\vartheta' P_n(s_{\pi k} s_{\pi k}') t_\vartheta + 2t_\vartheta' P_n[s_{\pi k}(r_{\vartheta k} + u_{\vartheta k x_0})] + P_n(r_{\vartheta k} + u_{\vartheta k x_0})^2 = t_\vartheta' P_n(s_{\pi k} s_{\pi k}') t_\vartheta + \zeta_{\vartheta n x_0}, \tag{70}$$

where $\zeta_{\vartheta n x_0}$ satisfies $\sup_{x_0 \in \mathcal{X}} |\zeta_{\vartheta n x_0}| = O_p(|t_\vartheta|^2 |\psi - \psi^*|) + O_p(n^{-1}|t_\vartheta||\psi - \psi^*|) + O_p(n^{-1}|\psi - \psi^*|^2)$. Then, (66) holds because $\sup_{\pi \in \Theta_\pi} |P_n(s_{\pi k} s_{\pi k}') - \mathcal{I}_\pi| = o_p(1)$ and the second term on the right of (69) is bounded by, from (68), $P_n(m_{\vartheta k}^2) = t_\vartheta' \mathcal{I}_\pi t_\vartheta + o_p(|t_\vartheta|^2)$, and Assumption 3(e),

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

We proceed to show (67). Consider the following expansion of $h_{\vartheta k x_0}$:

$$h_{\vartheta k x_0} = (l_{\vartheta k x_0} - 1)/2 - h_{\vartheta k x_0}^2/2 = (t_\vartheta' s_{\pi k} + r_{\vartheta k} + u_{\vartheta k x_0})/2 - h_{\vartheta k x_0}^2/2. \tag{71}$$

Then, (67) follows from (66), (71), and Assumption 3(d)(e), and the stated result follows. □

Proof of Proposition 3. For part (a), it follows from $\log(1+x) \le x$ and $h_{\vartheta k x_0} = (l_{\vartheta k x_0} - 1)/2 - h^2_{\vartheta k x_0}/2$ (see (71)) that

$$\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) = 2\sum_{k=1}^{n} \log(1 + h_{\vartheta k x_0}) \le 2nP_n(h_{\vartheta k x_0}) = \sqrt{n}\nu_n(l_{\vartheta k x_0} - 1) - nP_n(h^2_{\vartheta k x_0}). \tag{72}$$

Observe that $h^2_{\vartheta k x_0} = (l_{\vartheta k x_0} - 1)^2/(\sqrt{l_{\vartheta k x_0}} + 1)^2 \ge \mathbb{I}\{l_{\vartheta k x_0} \le \kappa\}(l_{\vartheta k x_0} - 1)^2/(\sqrt{\kappa}+1)^2$ for any $\kappa > 0$. Therefore,

$$P_n(h^2_{\vartheta k x_0}) \ge (\sqrt{\kappa} + 1)^{-2} P_n\left(\mathbb{I}\{l_{\vartheta k x_0} \le \kappa\}(l_{\vartheta k x_0} - 1)^2\right). \tag{73}$$

Substituting (70) into the right hand side of (73) gives

$$P_n(h^2_{\vartheta k x_0}) \ge (\sqrt{\kappa} + 1)^{-2} t'_\vartheta \left[P_n(s_{\pi k} s'_{\pi k}) - P_n(\mathbb{I}\{l_{\vartheta k x_0} > \kappa\} s_{\pi k} s'_{\pi k})\right] t_\vartheta + \zeta_{\vartheta n x_0}. \tag{74}$$

From Hölder's inequality, we have $P_n(\mathbb{I}\{l_{\vartheta k x_0} > \kappa\}|s_{\pi k}|^2) \le [P_n(\mathbb{I}\{l_{\vartheta k x_0} > \kappa\})]^{\delta/(2+\delta)}[P_n(|s_{\pi k}|^{2+\delta})]^{2/(2+\delta)}$. The right hand side is no larger than $\kappa^{-\delta/(2+\delta)} O_p(1)$ uniformly in $x_0 \in \mathcal{X}$ and $\vartheta \in \mathcal{N}_\varepsilon$ because (i) it follows from $\kappa \mathbb{I}\{l_{\vartheta k x_0} > \kappa\} \le l_{\vartheta k x_0}$ that $P_n(\mathbb{I}\{l_{\vartheta k x_0} > \kappa\}) \le \kappa^{-1} P_n(l_{\vartheta k x_0})$ and $\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_\varepsilon} |P_n(l_{\vartheta k x_0}) - 1| = o_p(1)$ from Assumption 3(d)(e)(f)(g), and (ii) $P_n(\sup_{\pi \in \Theta_\pi} |s_{\pi k}|^{2+\delta}) = O_p(1)$ from Assumption 3(a). Consequently, $\mathbb{P}(\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_\varepsilon} P_n(\mathbb{I}\{l_{\vartheta k x_0} > \kappa\}|s_{\pi k}|^2) \ge \lambda_{\min}/4) \to 0$ as $\kappa \to \infty$, and hence we can write (74) as $P_n(h^2_{\vartheta k x_0}) \ge \eta(1 + o_p(1)) t'_\vartheta \mathcal{I}_\pi t_\vartheta + O_p(|t_\vartheta|^2 |\psi - \psi^*|) + O_p(n^{-1})$ for $\eta = (\sqrt{\kappa}+1)^{-2}/2 > 0$ by taking κ sufficiently large. Because $\sqrt{n}\nu_n(l_{\vartheta k x_0} - 1) = \sqrt{n} t'_\vartheta \nu_n(s_{\pi k}) + O_p(1)$ from Assumption 3(d)(e), it follows from (72) that, uniformly in $x_0 \in \mathcal{X}$ and $\vartheta \in \mathcal{N}_\varepsilon$,

$$0 \le \ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) \le \sqrt{n} t'_\vartheta \nu_n(s_{\pi k}) - \eta(1 + o_p(1)) n t'_\vartheta \mathcal{I}_\pi t_\vartheta + O_p(n|t_\vartheta|^2|\psi - \psi^*|) + O_p(1). \tag{75}$$

Let $T_n := \mathcal{I}_\pi^{1/2} \sqrt{n} t_\vartheta$. From (75), Assumption 3(b)(g), and the fact $\psi - \psi^* \to 0$ if $t_\vartheta \to 0$, we obtain the following result: For any $\delta > 0$, there exist $\varepsilon > 0$ and $M, n_0 < \infty$ such that

$$\mathbb{P}\left(\inf_{x_0 \in \mathcal{X}} \inf_{\vartheta \in \mathcal{N}_\varepsilon} \left(|T_n| M - (\eta/2)|T_n|^2 + M\right) \ge 0\right) \ge 1 - \delta, \quad \text{for all} \quad n > n_0. \tag{76}$$

Rearranging the terms inside $\mathbb{P}(\cdot)$ gives $\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_\varepsilon} (|T_n| - (M/\eta))^2 \le 2M/\eta + (M/\eta)^2$. Taking its square root gives $\mathbb{P}(\sup_{x_0 \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}_\varepsilon} |T_n| \le M_1) \ge 1 - \delta$ for a constant M_1, and part (a) follows. Part (b) follows from part (a) and Proposition 2. □

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

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

Proof of Proposition 5. First, parts (a) and (b) hold when the right hand side is replaced with $K_j(k+m)^7 \rho^{\lfloor (k+m-1)/24 \rfloor}$ and $K_j(k+m)^7 \rho^{\lfloor (k+m-1)/1340 \rfloor}$ by using Proposition 4 and Lemma 3 and noting that $q_1 = 6q_0, q_2 = 5q_0, q_3 = 4q_0, \ldots, q_6 = q_0$. For example, when $j = 2$, we can bound $\sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\nabla^2 \ell_{k,m,x}(\vartheta) - \nabla^j 2\overline{\ell}_{k,m}(\vartheta)|$ from $\nabla^2 \ell_{k,m,x}(\vartheta) = \Delta_{1,k,m,x}^2(\vartheta) + \Delta_{2,k,m,x}^{1,1}(\vartheta)$, $\sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Delta_{j,k,m,x}^{\mathcal{I}(j)}(\vartheta) - \overline{\Delta}_{j,k,m}^{\mathcal{I}(j)}(\vartheta)| \le K_{\mathcal{I}(j)}(k+m)^7 \rho^{\lfloor (k+m-1)/24 \rfloor}$, $K_{\mathcal{I}(j)} \in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$, $r_{(2)} = q_2 = 5q_0$, and $r_{(1,1)} = q_1/2 = 3q_0$. Second, letting $\rho_* = \rho^{1/1340} \mathbb{I}\{\rho > 0\}$ and redefining K_j gives parts (a) and (b). Parts (c) and (d) follow Proposition 4 and Lemma 3. □

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

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

where

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

In view of Lemma 4 and the Hölder's inequality, part (a) holds if, for $j = 1, \ldots, 6$, there exist random variables $(\{A_{j,k}\}_{k=1}^n, B_j) \in L^{q_0}(\mathbb{P}_{\vartheta^*})$ and $\rho_* \in (0,1)$ such that, for all $1 \le k \le n$ and $m \ge 0$,

$$(A) \quad \sup_{m \ge 0} \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Psi_{k,m,x}^j(\vartheta)| \le A_{j,k}, \quad (B) \quad \sup_{x \in \mathcal{X}} \sup_{\vartheta \in \mathcal{N}^*} |\Psi_{k,m,x}^j(\vartheta) - \overline{\Psi}_{k,m}^j(\vartheta)| \le B_j(k+m)^7 \rho_*^{k+m-1}. \tag{77}$$

We show (A) and (B). From (96) 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 $\nabla^j \bar{\ell}_{k,m}$ replacing $\nabla^j \ell_{k,m,x}$. Therefore, (A) of (77) follows from Proposition 5(c) and the Hölder's inequality. (B) of (77) 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.

For part (c), the bound on $\nabla l^j_{k,m,x}(\vartheta)$ follows from writing $\nabla l^j_{k,m,x}(\vartheta) = [\overline{p}_\vartheta(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, X_{-m} = x)/\overline{p}_{\vartheta^*}(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, X_{-m} = x)]\Psi^j_{k,m,x}(\vartheta)$ and using (77) and Lemma 4. $\overline{\nabla}^j l_{k,m}(\vartheta)$ is bounded by a similar argument. Part (d) follows from parts (a)(b)(c), the completeness of $L^q(\mathbb{P}_{\vartheta^*})$, Markov's inequality, and Borel-Cantelli Lemma. Part (e) follows from combining parts (a) and (b) and letting $m' \to \infty$ in part (b). □

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 $\mathbb{E}_{\vartheta^*_1} \log f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma, \theta)$ from Assumption 5(c), and (ii) $\sup_{\vartheta_1 \in \Theta_1} |n^{-1}\ell_{0,n}(\vartheta_1) - \mathbb{E}_{\vartheta^*_1} \log f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma, \theta)| \xrightarrow{p} 0$ and $\mathbb{E}_{\vartheta^*_1} \log f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma, \theta)$ is continuous because (Y_k, W_k) is strictly stationary and ergodic from Assumption 1(e) and $\mathbb{E}_{\vartheta^*_1} \sup_{\vartheta_1 \in \Theta_1} |\log f(Y_1|\overline{\mathbf{Y}}_0, W_1; \gamma, \theta)| < \infty$ from Assumption 2(c).

We proceed to prove the consistency of $\hat{\vartheta}_2$. Define, similarly to pp. 2265–2266 in DMR, $\Delta_{k,m,x}(\vartheta_2) := \log p_{\vartheta_2}(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m}, X_{-m} = x)$, $\Delta_{k,m}(\vartheta_2) := \log p_{\vartheta_2}(Y_k|\overline{\mathbf{Y}}^{k-1}_{-m}, \mathbf{W}^k_{-m})$, $\Delta_{k,\infty}(\vartheta_2) := \lim_{m\to\infty} \Delta_{k,m}(\vartheta_2)$, 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 10(a) 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 5(d) because $\mathbb{E}_{\vartheta^*_1}[\log p_{\vartheta_2}(Y_1|\overline{\mathbf{Y}}^0_{-m}, \mathbf{W}^1_{-m})]$ 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) $\sup_{\xi_2} \sup_{\vartheta_2 \in \Theta_2} |n^{-1}\ell_n(\vartheta_2, \xi_2) - \ell(\vartheta_2)| \xrightarrow{p} 0$ 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, $\inf_{\vartheta_2 \in \Gamma^*} |\hat{\vartheta}_2 - \vartheta_2| \xrightarrow{p} 0$ 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 $q = \dim(\psi)$. For a $k \times 1$ vector a, define $a^{\otimes p} := a \otimes a \otimes \cdots \otimes a$ (p times) and $\nabla_{a^{\otimes p}} := \nabla_a \otimes \nabla_a \otimes \cdots \otimes \nabla_a$ (p times). Recall that the $(p+1)$-th order Taylor expansion of $f(x)$ with $x \in \mathbb{R}^q$ around $x = x^*$ is given by

$$f(x) = f(x^*) + \sum_{j=1}^{p} \frac{1}{j!} \nabla_{(x^{\otimes j})'} f(x^*)(x - x^*)^{\otimes j} + \frac{1}{(p+1)!} \nabla_{(x^{\otimes (p+1)})'} f(\overline{x})(x - x^*)^{\otimes (p+1)},$$

where $\overline{x}$ lies between x and x^*, and $\overline{x}$ may differ from element to element of $\nabla_{x^{\otimes(p+1)}} f(\overline{x})$.

Choose $\epsilon > 0$ sufficiently small so that $\mathcal{N}_\epsilon$ is a subset of $\mathcal{N}^*$ in Assumption 4. For $m \geq 0$ and $j = 1, 2, \ldots$, let

$$\Lambda^j_{k,m,x_{-m}}(\psi, \pi) := \frac{\nabla_{\psi^{\otimes j}} p_{\psi\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_{-m}, x_{-m})}{j! p_{\psi^*\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_{-m}, x_{-m})}, \quad \Lambda^j_{k,m}(\psi, \pi) := \frac{\nabla_{\psi^{\otimes j}} p_{\psi\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_{-m})}{j! p_{\psi^*\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_{-m})},$$

and $\Delta\psi := \psi - \psi^*$. With this notation, expanding $l_{\vartheta k x_0} - 1$ three times around ψ^* while fixing π gives, with $\overline{\psi} \in [\psi, \psi^*]$,

$$\begin{aligned} l_{\vartheta k x_0} - 1 &= \Lambda^1_{k,0,x_0}(\psi^*, \pi)' \Delta\psi + \Lambda^2_{k,0,x_0}(\psi^*, \pi)'(\Delta\psi)^{\otimes 2} + \Lambda^3_{k,0,x_0}(\overline{\psi}, \pi)'(\Delta\psi)^{\otimes 3} \\ &= \Lambda^1_{k,0}(\psi^*, \pi)' \Delta\psi + \Lambda^2_{k,0}(\psi^*, \pi)'(\Delta\psi)^{\otimes 2} + \Lambda^3_{k,0}(\overline{\psi}, \pi)'(\Delta\psi)^{\otimes 3} + u_{k x_0}(\psi, \pi), \end{aligned} \tag{78}$$

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

Noting that $\nabla_\lambda \overline{p}_{\psi^*\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_0) = 0$ and $\nabla_{\lambda\eta'} \overline{p}_{\psi^*\pi}(Y_k | \overline{\mathbf{Y}}^{k-1}_0) = 0$ from (16), we may rewrite (78) as

$$l_{k\vartheta x_0} - 1 = t(\psi, \pi)' s_{\varrho k} + r_{k,0}(\psi, \pi) + u_{kx_0}(\psi, \pi), \tag{79}$$

where $s_{\varrho k}$ is defined in (20), $r_{k,0}(\psi, \pi) := \widetilde{\Lambda}_{k,0}(\pi)'(\Delta\eta)^{\otimes 2} + \Lambda^3_{k,0}(\overline{\psi}, \pi)'(\Delta\psi)^{\otimes 3}$, where $\widetilde{\Lambda}_{k,0}(\pi)$ denotes the part of $\Lambda^2_{k,0}(\psi^*, \pi)$ corresponding to $(\Delta\eta)^{\otimes 2}$.

For $m \geq 0$, define $v_{k,m}(\vartheta) := (\Lambda^1_{k,m}(\psi, \pi)', \Lambda^2_{k,m}(\psi, \pi)', \Lambda^3_{k,m}(\psi, \pi)')'$, and define $v_{k,\infty}(\vartheta) := \lim_{m\to\infty} v_{k,m}(\vartheta)$. In order to apply Corollary 1 to $l_{\vartheta k x_0} - 1$, we first show

$$\sup_{\vartheta \in \mathcal{N}_\epsilon} \left| P_n[v_{k,0}(\vartheta) v_{k,0}(\vartheta)'] - \mathbb{E}_{\vartheta^*}[v_{k,\infty}(\vartheta) v_{k,\infty}(\vartheta)'] \right| = o_p(1), \tag{80}$$

$$\nu_n(v_{k,0}(\vartheta)) \Rightarrow W(\vartheta), \tag{81}$$

where $W(\vartheta)$ is a mean-zero continuous Gaussian process with $\mathbb{E}_{\vartheta^*}[W(\vartheta_1)W(\vartheta_2)'] = \mathbb{E}_{\vartheta^*}[v_{k,\infty}(\vartheta_1) v_{k,\infty}(\vartheta_2)']$. (80) holds because $\sup_{\vartheta \in \mathcal{N}_\epsilon} P_n[v_{k,0}(\vartheta) v_{k,0}(\vartheta)' - v_{k,\infty}(\vartheta) v_{k,\infty}(\vartheta)'] = o_p(1)$ 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 in ϑ from the continuity of $\nabla^j l_{k,m,x}(\vartheta)$ and Proposition 6, and $\mathbb{E}_{\vartheta^*}\sup_{\vartheta\in\mathcal{N}_\epsilon}|v_{k,\infty}(\vartheta)|^2<\infty$ from Proposition 6. (81) holds because $\sup_{\vartheta\in\mathcal{N}_\epsilon}\nu_n(v_{k,0}(\vartheta)-v_{k,\infty}(\vartheta))=o_p(1)$ 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 martingale CLT because $v_{k,\infty}(\vartheta)$ is a stationary $L^2(\mathbb{P}_{\vartheta^*})$ martingale difference sequence for all $\vartheta\in\mathcal{N}_\epsilon$ from Proposition 6, 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 (79) satisfies Assumption 3(a)–(g). Observe that $t(\psi,\pi)=0$ if and only if $\psi=\psi^*$. First, $s_{\varrho k}$ satisfies Assumption 3(a)(b)(g) by Proposition 6, (80), (81), and Assumption 6. Second, $r_{k,0}(\psi,\pi)$ satisfies Assumption 3(c)(d) from Proposition 6 and (81). Third, $u_{kx_0}(\psi,\pi)$ satisfies Assumption 3(e)(f) 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

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

Then, we can write (22) as

$$\sup_{\xi\in\Xi}\sup_{\vartheta\in A_{n\varepsilon c}(\xi)}\left|2\left[\ell_n(\psi,\pi,\xi)-\ell_n(\psi^*,\pi,\xi)\right]-A_n(\sqrt{n}t_{\eta.\lambda\varrho})-B_{\varrho n}(\sqrt{n}t_\lambda)\right|=o_p(1), \tag{82}$$

where

$$\begin{aligned}A_n(t_{\eta.\lambda\varrho})&=2t_{\eta.\lambda\varrho}'G_{\eta n}-t_{\eta.\lambda\varrho}'\mathcal{I}_\eta t_{\eta.\lambda\varrho},\\ B_{\varrho n}(t_\lambda)&=2t_\lambda'G_{\lambda.\eta\varrho n}-t_\lambda'\mathcal{I}_{\lambda.\eta\varrho}t_\lambda=Z_{\lambda\varrho n}'\mathcal{I}_{\lambda.\eta\varrho}Z_{\lambda\varrho n}-(t_\lambda-Z_{\lambda\varrho n})'\mathcal{I}_{\lambda.\eta\varrho}(t_\lambda-Z_{\lambda\varrho n}).\end{aligned} \tag{83}$$

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 (82), we obtain, uniformly in $\pi\in\Theta_\pi$,

$$2[\ell_n(\hat{\psi}_\pi,\pi,\xi)-\ell_{0n}(\hat{\vartheta}_0)]=B_{\varrho n}(\sqrt{n}\hat{t}_\lambda)+o_p(1). \tag{84}$$

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 $\pi \in \Theta_\pi$ 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 (84), 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 $\hat{\psi}$, (82), 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 (81) 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 $\alpha \in \Theta_\alpha$. Therefore, $\sup_{\varrho \in \Theta_\varrho} B_{\varrho n}(\sqrt{n}\tilde{t}_\lambda) \xrightarrow{d} \sup_{\varrho \in \Theta_\varrho}(\tilde{t}_{\lambda\varrho}' \mathcal{I}_{\lambda.\eta\varrho}\tilde{t}_{\lambda\varrho})$ follows from Theorem 1(c) of Andrews (2001). □

Proof of Proposition 10. The proof is similar to that of Proposition 8. Define $\Lambda_{k,m,x_{-m}}^j(\psi,\pi)$ and $\Lambda_{k,m}^j(\psi,\pi)$ 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 $\overline{\psi} \in [\psi, \psi^*]$,

$$l_{k\vartheta x_0} - 1 = \sum_{j=1}^{4} \Lambda_{k,0}^j(\psi^*,\pi)'(\Delta\psi)^{\otimes j} + \Lambda_{k,0}^5(\overline{\psi},\pi)'(\Delta\psi)^{\otimes 5} + u_{kx_0}(\psi,\pi), \tag{85}$$

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

Define $\overline{p}_{\psi\pi k,0} := \overline{p}_{\psi\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$. Observe that $s_{\varrho k}$ defined in (36) satisfies

$$s_{\varrho k} := \begin{pmatrix} \nabla_\eta \overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0} \\ \zeta_{k,0}(\varrho)/2 \\ \nabla_{\lambda_\mu\lambda_\sigma}\overline{p}_{\psi^*\pi k,0}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \\ \nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi k,0}/2\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \\ \nabla_{\lambda_\beta\lambda_\mu}\overline{p}_{\psi^*\pi k,0}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \\ \nabla_{\lambda_\beta\lambda_\sigma}\overline{p}_{\psi^*\pi k,0}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \\ V(\nabla_{\lambda_\beta\lambda_\beta}\overline{p}_{\psi^*\pi k,0})/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \end{pmatrix}.$$

Noting that $\nabla_\lambda \overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$ and $\nabla_{\lambda\eta'}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$ from (16) and (17), we may rewrite (85) as, with $t(\psi,\pi)$ and $s_{\varrho k}$ defined in (33) and (36),

$$l_{\vartheta k x_0} - 1 = t(\psi,\pi)'s_{\varrho k} + r_{k,0}(\pi) + u_{kx_0}(\psi,\pi), \tag{86}$$

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

The stated result follows from Corollary 1 if the terms on the right hand side of (86) satisfy Assumption 3. Similarly to the proof of Proposition 9, define $v_{k,m}(\vartheta) := (\zeta_{k,m}(\varrho), \Lambda_{k,m}^1(\psi,\pi)', \ldots, \Lambda_{k,m}^5(\psi,\pi)')'$. Note that $\zeta_{k,m}(\varrho)$ satisfies Proposition 6 because the mean value theorem and $\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}) = 0$ gives $\zeta_{k,m}(\varrho) = [\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\varrho\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) - \nabla_{\lambda_\mu^2}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})]/[\varrho\alpha(1-\alpha)\overline{p}_{\psi^*\varrho\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})] = \nabla_\varrho\nabla_{\lambda_\mu^2}\overline{p}_{\psi^*\alpha\bar{\varrho}}(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1})/[\alpha(1-\alpha)\overline{p}_{\psi^*\bar{\varrho}\alpha}(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1})]$ for $\bar{\varrho} \in [0,\varrho]$. 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 (80)–(81) from repeating the argument in the proof of Proposition 9.

We proceed to show that the terms on the right hand side of (86) satisfy Assumption 3. Observe that $t(\psi,\pi) = 0$ if and only if $\psi = \psi^*$. $s_{\varrho k}$ and $u_{kx_0}(\psi,\pi)$ satisfy Assumption 3 by noting that $s_{\varrho k}$ is a linear function of $v_{k,0}(\vartheta)$ and using 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, $\Lambda_{k,0}^5(\overline{\psi},\pi)'(\Delta\psi)^{\otimes 5}$ satisfies Assumption 3(c)(d) from Proposition 6, (81) and $\lambda_\mu^5 = (12\lambda_\mu/b(\alpha))[\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12] - 12(\lambda_\sigma/b(\alpha))\lambda_\mu\lambda_\sigma = O(|\psi||t(\psi,\pi)|)$. Second, $\lambda_\mu^4[\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi k,0} - b(\alpha)\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi k,0}]/\overline{p}_{\psi^*\pi k,0}$ satisfies Assumption 3(c)(d) from Lemma 6(b). Third, for $\widetilde{\Delta}_{k,0}(\pi)'\tau(\psi)$, observe that $\nabla_{\lambda\eta^j}\overline{p}_{\psi^*\pi k,0} = 0$ for any $j \geq 1$ in view of (27)–(30). Therefore, $\widetilde{\Delta}_{k,0}(\pi)'\tau(\psi)$ is written as, with $\Delta\eta := \eta - \eta^*$,

$$\widetilde{\Delta}_{k,0}(\pi)'\tau(\psi) = \nabla_{(\eta^{\otimes 2})'}\overline{p}_{\psi^*\pi k,0}(\Delta\eta)^{\otimes 2}/2!\overline{p}_{\psi^*\pi k,0} + R_{3k\vartheta} + R_{4k\vartheta}, \tag{87}$$

where $R_{3k\vartheta} := \nabla_{(\psi^{\otimes 3})'}\overline{p}_{\psi^*\pi k,0}(\Delta\psi)^{\otimes 3}/3!\overline{p}_{\psi^*\pi k,0}$ and

$$R_{4k\vartheta} := [\nabla_{(\psi^{\otimes 4})'}\overline{p}_{\psi^*\pi k,0}(\Delta\psi)^{\otimes 4} - \nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi k,0}\lambda_\mu^4]/4!\overline{p}_{\psi^*\pi k,0}. \tag{88}$$

The first term in (87) clearly satisfies Assumption 3(c)(d). The terms in $R_{3k\vartheta}$ belong 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. These terms satisfy Assumption 3(c)(d) because the term (i) is bounded by $|\psi||t(\psi,\pi)|$ because $\lambda_\sigma^3 = \lambda_\sigma[\lambda_\sigma^2 + b(\alpha)\lambda_\mu^4/12] - (\lambda_\mu^3 b(\alpha))\lambda_\mu\lambda_\sigma/12$, the term (ii) is bounded by $\varrho\lambda_\mu^3$ from Lemma 6(a), and the terms in (iii) are bounded by $|\psi||t(\psi,\pi)|$ because they either contain $\Delta\eta$ 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 Assumption 3(c)(d) because they either contain $\Delta\eta$ or a term of the form $\lambda_\mu^i\lambda_\sigma^j\lambda_\beta^k$ with $i+j+k=4$ and $i \neq 4$. 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 $(\hat{\psi}_\alpha, \hat{\varrho}_\alpha) := \arg\max_{(\psi,\varrho)\in\Theta_\psi\times\Theta_\varrho} \ell_n(\psi,\varrho,\alpha,\xi)$ 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 $(\hat{\psi}_\alpha^j, \hat{\varrho}_\alpha^j) := \arg\max_{(\psi,\varrho)\in\Theta_\psi\times\Theta_\varrho,\lambda\in\Theta_\lambda^j} \ell_n(\psi,\varrho,\alpha,\xi)$. Then, uniformly in α,

$$\ell_n(\hat{\psi}_\alpha, \hat{\varrho}_\alpha, \alpha, \xi) = \max\left\{\ell_n(\hat{\psi}_\alpha^1, \hat{\varrho}_\alpha^1, \alpha, \xi), \ell_n(\hat{\psi}_\alpha^2, \hat{\varrho}_\alpha^2, \alpha, \xi)\right\}.$$

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 (83) in the proof of Proposition 9 but using $t(\psi,\pi)$ and $s_{\varrho k}$ defined in (33) and (36) and replacing t_λ in (83) with $t_\lambda(\lambda,\varrho,\alpha)$. Observe that the proof of Proposition 9 goes through up to (84) 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 $\hat{\varrho}^1(\hat{\lambda}^1_\mu)^2 = O_p(n^{-1/2})$ from Proposition 10(a) and $|\hat{\lambda}^1_\mu| \geq n^{-1/8}(\log n)^{-1}$. 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,

$$2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{0n}(\hat{\vartheta}_0)] = \max\{B_{0n}(\sqrt{n}t_\lambda(\hat{\lambda}^1, \hat{\varrho}^1, \alpha)), B_{\hat{\varrho}^2 n}(\sqrt{n}t_\lambda(\hat{\lambda}^2, \hat{\varrho}^2, \alpha))\} + o_p(1). \tag{89}$$

We proceed to construct parameter spaces $\tilde{\Lambda}^1_{\lambda\alpha}$ and $\tilde{\Lambda}^2_{\lambda\alpha\varrho}$ that are locally equal to the cones Λ^1_λ and $\Lambda^2_{\lambda\varrho}$ defined in (38). Define $c(\alpha) := \alpha(1-\alpha)$, and denote the elements of $t_\lambda(\hat{\lambda}^j, \hat{\varrho}^j, \alpha)$ corresponding to (34) by

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

Note that $\hat{\lambda}^1_\sigma = O_p(n^{-3/8}\log n)$ 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, $\hat{t}^2_{\sigma^2} = c(\alpha)(\hat{\lambda}^2_\sigma)^2 + o_p(n^{-1/2})$ because $|\hat{\lambda}^2_\mu| < n^{-1/8}(\log n)^{-1}$. Consequently,

$$\begin{aligned} &\hat{t}^1_{\beta\sigma} = o_p(n^{-1/2}), \quad \hat{t}^1_{v(\beta)} = o_p(n^{-1/2}), \quad \hat{t}^1_{\sigma^2} = c(\alpha)b(\alpha)(\hat{\lambda}^1_\mu)^4/12 + o_p(n^{-1/2}), \\ &\hat{t}^2_{\sigma^2} = c(\alpha)(\hat{\lambda}^2_\sigma)^2 + o_p(n^{-1/2}). \end{aligned} \tag{90}$$

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

$$2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{0n}(\hat{\vartheta}_0)] = W_n(\alpha) + o_p(1), \tag{91}$$

uniformly in $\alpha \in \Theta_\alpha$ because (i) $W_n(\alpha) \geq 2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{0n}(\hat{\vartheta}_0)] + o_p(1)$ in view of the definition of $(\tilde{\lambda}^1_\alpha, \tilde{\varrho}^1_\alpha, \tilde{\lambda}^2_{\alpha\varrho})$, (89), and (90), and (ii) $2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{0n}(\hat{\vartheta}_0)] \geq$ $\max\{2[\max_\eta \ell_n(\eta, \tilde{\lambda}^1_\alpha, \tilde{\varrho}^1_\alpha, \alpha, \xi), \sup_{\varrho\in\Theta_\varrho} \max_\eta \ell_n(\eta, \tilde{\lambda}^2_{\alpha\varrho}, \varrho, \alpha, \xi)\} - 2\ell_{0n}(\hat{\vartheta}_0) + o_p(1) = W_n(\alpha) + o_p(1)$ 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 (81) 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, $W_n(\alpha) \stackrel{d}{\to} \sup_{\varrho\in\Theta_\varrho} \max\{\mathbb{I}\{\varrho = 0\}(\tilde{t}^1_\lambda)'\mathcal{I}_{\lambda.\eta 0}\tilde{t}^1_\lambda, (\tilde{t}^2_{\lambda\varrho})'\mathcal{I}_{\lambda.\eta\varrho}\tilde{t}^2_{\lambda\varrho}\}$ uniformly in α from Theorem 1(c) of Andrews (2001), and the stated result follows from (91). □

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

$$l_{\vartheta k x_0} - 1 = t(\psi, \pi)' s_{\varrho k} + r_{k,0}(\pi) + u_{kx_0}(\psi, \pi), \tag{92}$$

where $t(\psi, \pi)$ is defined in (45), $s_{\varrho k}$ is defined in (46) and satisfies

$$s_{\varrho k} := \begin{pmatrix} \nabla_\eta \overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0} \\ \zeta_{k,0}(\varrho)/2 \\ \nabla_{\mu^3} f^*_k/3! f^*_k \\ \nabla_{\mu^4} f^*_k/4! f^*_k \\ \nabla_{\lambda_\beta\lambda_\mu}\overline{p}_{\psi^*\pi k,0}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \\ \widetilde{\nabla}_{v(\lambda_\beta)}\overline{p}_{\psi^*\pi k,0}/\alpha(1-\alpha)\overline{p}_{\psi^*\pi k,0} \end{pmatrix},$$

and

$$\begin{aligned} r_{k,0}(\pi) := &\ \widetilde{\Lambda}_{k,0}(\pi)'\tau(\psi) + \Lambda^5_{k,0}(\overline{\psi}, \pi)'(\Delta\psi)^{\otimes 5} \\ &+ \lambda^3_\mu[\nabla_{\lambda^3_\mu}\overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0} - \alpha(1-\alpha)(1-2\alpha)\nabla_{\mu^3} f^*_k/f^*_k]/3! \\ &+ \lambda^4_\mu[\nabla_{\lambda^4_\mu}\overline{p}_{\psi^*\pi k,0}/\overline{p}_{\psi^*\pi k,0} - \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\nabla_{\mu^4} f^*_k/f^*_k]/4!, \end{aligned}$$

where $u_{kx_0}(\psi, \pi)$, $\overline{p}_{\psi\pi k,m}$, and the terms in the definition of $r_{k,0}(\pi)$ 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 (92) satisfy Assumption 3. $t(\psi, \pi) = 0$ if and only if $\psi = \psi^*$. $s_{\varrho k}$ 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, $\Lambda^5_{k,0}(\overline{\psi}, \pi)'(\Delta\psi)^{\otimes 5}$ satisfies Assumption 3(c)(d) from a similar argument to the proof of Proposition 10; λ^5_μ is dominated by λ^3_μ or λ^4_μ because $\inf_{0\leq\alpha\leq 1}\max\{|1-2\alpha|, |1-6\alpha+6\alpha^2|\} > 0$. Second, similar to (87) in the proof of Proposition 10, write $\widetilde{\Lambda}_{k,0}(\pi)'\tau(\psi) = \nabla_{(\eta^{\otimes 2})'}\overline{p}_{\psi^*\pi k,0}(\Delta\eta)^{\otimes 2}/2!\overline{p}_{\psi^*\pi k,0} + \tilde{R}_{3k\vartheta} + R_{4k\vartheta}$, where

$\tilde{R}_{3k\vartheta} := [\nabla_{(\psi^{\otimes 3})'}\overline{p}_{\psi^*\pi k,0}(\Delta\psi)^{\otimes 3} - \nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi k,0}\lambda_\mu^3]/3!\overline{p}_{\psi^*\pi k,0}$, and $R_{4k\vartheta}$ is defined as $R_{4k\vartheta}$ in (88). The term $\nabla_{(\eta^{\otimes 2})'}\overline{p}_{\psi^*\pi k,0}(\Delta\eta)^{\otimes 2}/2!\overline{p}_{\psi^*\pi k,0}$ clearly satisfies Assumption 3(c)(d). The terms in $\tilde{R}_{3k\vartheta}$ satisfy Assumption 3(c)(d) because they contain either $\Delta\eta$ or $\lambda_\mu^2\lambda_\beta$ or $\lambda_\mu\lambda_\beta^2$ or λ_β^3. The terms in $R_{4k\vartheta}$ satisfy Assumption 3(c)(d) because they either contain $\Delta\eta$ or a term of the form $\lambda_\mu^i\lambda_\beta^{4-i}$ with $1 \leq i \leq 3$. The last two terms in $r_{k,0}(\pi)$ satisfy Assumption 3(c)(d) from Lemma 7. 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 $(\hat{\psi}, \hat{\varrho}, \hat{\alpha}) := \arg\max_{(\psi,\varrho,\alpha)\in\Theta_\psi\times\Theta_\varrho\times\Theta_\alpha}\ell_n(\psi,\varrho,\alpha,\xi)$ denote the MLE of (ψ,ϱ,α). Consider the sets $\Theta_\lambda^1 := \{\lambda \in \Theta_\lambda : |\lambda_\mu| \geq n^{-1/6}(\log n)^{-1}\}$ and $\Theta_\lambda^2 := \{\lambda \in \Theta_\lambda : |\lambda_\mu| < n^{-1/6}(\log n)^{-1}\}$, so that $\Theta_\lambda = \Theta_\lambda^1 \cup \Theta_\lambda^2$. For $j = 1, 2$, define $(\hat{\psi}^j, \hat{\varrho}^j, \hat{\alpha}^j) := \arg\max_{(\psi,\varrho,\alpha)\in\Theta_\psi\times\Theta_\varrho\times\Theta_\alpha,\lambda\in\Theta_\lambda^j}\ell_n(\psi,\varrho,\alpha,\xi)$, so that $\ell_n(\hat{\psi},\hat{\varrho},\hat{\alpha},\xi) = \max_{j\in\{1,2\}}\ell_n(\hat{\psi}^j,\hat{\varrho}^j,\hat{\alpha}^j,\xi)$.

Define $B_{\varrho n}(t_\lambda(\lambda,\varrho,\alpha))$ as in (83) in the proof of Proposition 9 but using $t(\psi,\pi)$ and $s_{\varrho k}$ defined in (45) and (46) and replacing t_λ in (83) with $t_\lambda(\lambda,\varrho,\alpha)$. 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 (89), we obtain

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

We proceed to construct parameter spaces that are locally equal to the cones Λ_λ^1 and $\Lambda_{\lambda\varrho}^2$ defined in (47). Define $c(\alpha) := \alpha(1-\alpha)$, and denote the elements of $t_\lambda(\hat{\lambda}^j,\hat{\varrho}^j,\hat{\alpha}^j)$ corresponding to (45) by

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

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

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

In view of this, let $t_\lambda(\lambda,\varrho,\alpha) := (t_{\varrho\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}', t_{v(\beta)}')' \in \mathbb{R}^{q_\lambda}$, and consider the following sets:

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

Define $(\tilde{\lambda}^1, \tilde{\varrho}^1, \tilde{\alpha}^1)$ and $\tilde{\lambda}^2_{\alpha\varrho}$ by $B_{0n}(\sqrt{n}t_\lambda(\tilde{\lambda}^1, \tilde{\varrho}^1, \tilde{\alpha}^1)) = \max_{t_\lambda(\lambda,\varrho,\alpha)\in\tilde{\Lambda}^1_\lambda} B_{0n}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha))$ and $B_{\varrho n}(\sqrt{n}t_\lambda(\tilde{\lambda}^2_{\alpha\varrho}, \varrho, \alpha)) = \max_{t_\lambda(\lambda,\varrho,\alpha)\in\tilde{\Lambda}^2_{\lambda\alpha\varrho}} B_{\varrho n}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha))$. $\tilde{\Lambda}^1_\lambda$ is locally (in a neighborhood of $\varrho = 0$, $\lambda = 0$) equal to the cone Λ^1_λ because, when $|1-2\alpha| \geq 1/2$, we have $t_{\mu^4}/t_{\mu^3} \to 0$ as $\lambda_\mu \to 0$, and when $|1-2\alpha| \leq 1/2$, we have $1-6\alpha+6\alpha^2 < 0$. $\tilde{\Lambda}^2_{\lambda\alpha\varrho}$ is locally equal to the cone $\Lambda^2_{\lambda\varrho}$ uniformly in $\varrho \in \Theta_\varrho$.

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

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

Next, $\ell_n(\psi_m, \pi_m, \xi_{M_0+1}) - \ell_n(\psi^*_m, \pi_m, \xi_{M_0+1})$ admits the same expansion as $\ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi)$ in (22) or (37). Therefore, the stated result follows from applying the proof of Propositions 9, 11, and 13 to $\ell_n(\hat{\psi}_m, \pi_m, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})$ for each m and combining the results to derive the joint asymptotic distribution of $\{\ell_n(\hat{\psi}_m, \pi_m, \xi_{M_0+1}) - \ell_n(\hat{\vartheta}_{M_0}, \xi_{M_0})\}_{m=1}^{M_0}$. □

Proof of Proposition 18. Observe that Proposition 2 holds under $\mathbb{P}^n_{\vartheta^*,x_0}$ under the assumptions of Proposition 8, 10, and 12. Because $\vartheta_n = (\eta'_n, \lambda'_n, \pi'_n)' \in \mathcal{N}_{c/\sqrt{n}}$ by choosing $c > |h|$, it follows from Proposition 2 that

$$\sup_{x_0\in\mathcal{X}} \left| \log \frac{d\mathbb{P}^n_{\vartheta_n,x_0}}{d\mathbb{P}^n_{\vartheta^*,x_0}} - h'\nu_n(s_{\varrho_n k}) + \frac{1}{2}h'\mathcal{I}_{\varrho_n}h \right| = o_{\mathbb{P}^n_{\vartheta^*,x_0}}(1), \tag{93}$$

where $s_{\varrho k}$ is given by (20), (36), and (46) for the models of non-normal distribution, heteroscedastic normal distribution, and homoscedastic normal distribution, respectively. Furthermore, $\nu_n(s_{\varrho_n k}) \Rightarrow G_\varrho$ under $\mathbb{P}^n_{\vartheta^*,x_0}$, where G_ϱ is a mean zero Gaussian process with $cov(G_{\varrho_1}, G_{\varrho_2}) = \mathcal{I}_{\varrho_1\varrho_2} := \lim_{k\to\infty} \mathbb{E}_{\vartheta^*}(s_{\varrho_1 k}s'_{\varrho_2 k})$. Therefore, $d\mathbb{P}^n_{\vartheta_n,x_0}/d\mathbb{P}^n_{\vartheta^*,x_0}$ converges in distribution under $\mathbb{P}^n_{\vartheta^*,x_0}$ to $\exp\left(N(\mu,\sigma^2)\right)$ with $\mu = -(1/2)h'\mathcal{I}_\varrho h$ and $\sigma^2 = h'\mathcal{I}_\varrho h$ so that $E(\exp\left(N(\mu,\sigma^2)\right)) = 1$. Consequently, part (a) follows from Le Cam's first lemma (see, e.g., Corollary 12.3.1 of Lehmann and Romano (2005)). Part (b) follows from Le Cam's third lemma (see, e.g., Corollary 12.3.2 of Lehmann and Romano (2005)) because part (a) and (93) imply that

$$\begin{pmatrix} \nu_n(s_{\varrho_n k}) \\ \log \frac{d\mathbb{P}^n_{\vartheta_n,x_0}}{d\mathbb{P}^n_{\vartheta^*,x_0}} \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ -\frac{1}{2}h'\mathcal{I}_\varrho h \end{pmatrix}, \begin{pmatrix} \mathcal{I}_\varrho & \mathcal{I}_\varrho h \\ h'\mathcal{I}_\varrho & h'\mathcal{I}_\varrho h \end{pmatrix} \right) \quad \text{under } \mathbb{P}^n_{\vartheta^*,x_0}.$$

□

Proof of Proposition 19. The proof follows the argument in the proof of Proposition 9. Observe that $h_\eta = 0$ and $h_\lambda = \sqrt{n} t_\lambda(\lambda_n, \pi_n)$ hold under H_{1n}. Therefore, Proposition 18 holds under $\mathbb{P}^n_{\vartheta_n, x_0}$ implied by H_{1n}, and, in conjunction with Theorem 12.3.2(a) of Lehmann and Romano (2005), Propositions 6 and 8 hold under $\mathbb{P}^n_{\vartheta_n, x_0}$. Consequently, the proof of Proposition 9 goes through if we replace $G_{\lambda.\eta\varrho n} \Rightarrow G_{\lambda.\eta\varrho}$ with $G_{\lambda.\eta\varrho n} \Rightarrow G_{\lambda.\eta\varrho} + (\mathcal{I}_{\lambda\varrho\varrho} - \mathcal{I}_{\lambda\eta\varrho}\mathcal{I}_\eta^{-1}\mathcal{I}_{\eta\lambda\varrho})h_\lambda = G_{\lambda.\eta\varrho} + \mathcal{I}_{\lambda.\eta\varrho} h_\lambda$, and the stated result follows. □

Proof of Propositions 20 and 21. The proof is similar to the proof of Proposition 19. Observe that, for $j \in \{a, b\}$, $h^j_\eta = 0$ and $h^j_\lambda = \sqrt{n} t_\lambda(\lambda_n, \pi_n) + o(1)$ hold under H^j_{1n}. Therefore, Proposition 18 holds under $\mathbb{P}^n_{\vartheta_n, x_0}$ implied by H^j_{1n}, and the stated result follows from repeating the argument of proof of Proposition 19. □

Proof of Proposition 22. We only provide the proof for the models of non-normal distribution with $M_0 = 1$ because the proof for the other models is similar. The proof follows the argument in the proof of Theorem 15.4.2 in Lehmann and Romano (2005). Define $\boldsymbol{C}_\eta$ as the set of sequences $\{\eta_n\}$ satisfying $\sqrt{n}(\eta_n - \eta^*) \to h_\eta$ for some finite h_η. Denote the MLE of the one-regime model parameter by $\hat{\eta}_n$. For the MLE under H_0, $\sqrt{n}(\hat{\eta}_n - \eta^*)$ converges in distribution to a $\mathbb{P}_{\vartheta^*}$-a.s. finite random variable by the standard argument. Then, by the Almost Sure Representation Theorem (e.g., Theorem 11.2.19 of Lehmann and Romano (2005)), there exists random variable**s** $\tilde{\eta}_n$ and $\tilde{h}_\eta$ defined on a common probability space such that $\hat{\eta}_n$ and $\tilde{\eta}_n$ have the same distribution and $\sqrt{n}(\tilde{\eta}_n - \eta^*) \to \tilde{h}_\eta$ almost surely. Therefore, $\{\tilde{\eta}_n\} \in \boldsymbol{C}_\eta$ with probability one, and the stated result under H_0 follows from Lemma 8 because $\hat{\eta}_n$ and $\tilde{\eta}_n$ have the same distribution.

For the MLE under H_{1n}, note that the proof of Proposition 19 goes through when h_η is finite even if $h_\eta \neq 0$. Therefore, $\sqrt{n}(\hat{\eta}_n - \eta^*)$ converges in distribution to a $\mathbb{P}_{\vartheta_n}$-a.s. finite random variable under H_{1n}. Hence, the stated result follows from Lemma 8 and repeating the argument in the case of H_0. □

12.2 Auxiliary results

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

Define

$$
\begin{aligned}
\Phi^{\mathcal{I}(5)}_{\vartheta\mathcal{T}(5)}[\mathcal{F}] &:= \frac{1}{|\sigma(\mathcal{I}(5))|} \sum_{(\ell_1,\ldots,\ell_5)\in\sigma(\mathcal{I}(5))} \Bigg(\mathbb{E}^c_\vartheta \left[\phi^{\ell_1}_{\theta t_1} \phi^{\ell_2}_{\theta t_2} \phi^{\ell_3}_{\theta t_3} \phi^{\ell_4}_{\theta t_4} \phi^{\ell_5}_{\theta t_5} \middle| \mathcal{F} \right] \\
&\qquad - \sum_{(\{a,b,c\},\{d,e\})\in\sigma_5} \mathbb{E}^c_\vartheta \left[\phi^{\ell_a}_{\theta t_a} \phi^{\ell_b}_{\theta t_b} \phi^{\ell_c}_{\theta t_c} \middle| \mathcal{F} \right] \mathbb{E}^c_\vartheta \left[\phi^{\ell_d}_{\theta t_d} \phi^{\ell_e}_{\theta t_e} \middle| \mathcal{F} \right] \Bigg), \\
\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}] &:= \mathbb{E}^c_\vartheta \left[\phi_{\theta t_1} \phi_{\theta t_2} \phi_{\theta t_3} \phi_{\theta t_4} \phi_{\theta t_5} \phi_{\theta t_6} | \mathcal{F} \right] - \sum_{(\{a,b,c,d\},\{e,f\})\in\sigma_{61}} \mathbb{E}^c_\vartheta \left[\phi_{\theta t_a} \phi_{\theta t_b} \phi_{\theta t_c} \phi_{\theta t_d} | \mathcal{F} \right] \mathbb{E}^c_\vartheta \left[\phi_{\theta t_e} \phi_{\theta t_f} \middle| \mathcal{F} \right] \\
&\qquad - \sum_{(\{a,b,c\},\{d,e,f\})\in\sigma_{62}} \mathbb{E}^c_\vartheta \left[\phi_{\theta t_a} \phi_{\theta t_b} \phi_{\theta t_c} | \mathcal{F} \right] \mathbb{E}^c_\vartheta \left[\phi_{\theta t_d} \phi_{\theta t_e} \phi_{\theta t_f} \middle| \mathcal{F} \right] \\
&\qquad + 2 \sum_{(\{a,b\},\{c,d\},\{e,f\})\in\sigma_{63}} \mathbb{E}^c_\vartheta \left[\phi_{\theta t_a} \phi_{\theta t_b} | \mathcal{F} \right] \mathbb{E}^c_\vartheta \left[\phi_{\theta t_c} \phi_{\theta t_d} | \mathcal{F} \right] \mathbb{E}^c_\vartheta \left[\phi_{\theta t_e} \phi_{\theta t_f} \middle| \mathcal{F} \right],
\end{aligned}
\tag{94}
$$

where

$$
\begin{aligned}
\sigma_5 &:= \text{the set of } \binom{5}{3} = 10 \text{ partitions of } \{1,2,3,4,5\} \text{ of the form } \{a,b,c\},\{d,e\}, \\
\sigma_{61} &:= \text{the set of } \binom{6}{4} = 15 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b,c,d\},\{e,f\}, \\
\sigma_{62} &:= \text{the set of } \binom{6}{3}/2 = 10 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b,c\},\{d,e,f\}, \\
\sigma_{63} &:= \text{the set of } \binom{6}{2}\binom{4}{2}/6 = 15 \text{ partitions of } \{1,2,3,4,5,6\} \text{ of the form } \{a,b\},\{c,d\},\{e,f\}.
\end{aligned}
\tag{95}
$$

12.2.2 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

$\nabla^j_\vartheta P(Y;\vartheta)/P(Y;\vartheta) = \mathbb{E}[\nabla^j_\vartheta P(Y,X;\vartheta)/P(Y,X;\vartheta)|Y]$ and

$$
\begin{aligned}
\nabla \log f &= \nabla f/f, \quad \nabla^2 \log f = \nabla^2 f/f - (\nabla \log f)^2, \\
\nabla^3 \log f &= \nabla^3 f/f - 3\nabla^2 f \nabla f/f^2 + 2(\nabla f/f)^3, \\
\nabla^4 \log f &= \nabla^4 f/f - 4\nabla^3 f\nabla f/f^2 - 3(\nabla^2 f/f)^2 + 12\nabla^2 f(\nabla f)^2/f^3 - 6(\nabla f/f)^4, \\
\nabla^5 \log f &= \nabla^5 f/f - 5\nabla^4 f \nabla f/f^2 - 10\nabla^3 f\nabla^2 f/f^2 + 20\nabla^3 f(\nabla f)^2/f^3 \\
&\quad + 30(\nabla^2 f)^2 \nabla f/f^3 - 60\nabla^2 f(\nabla f)^3/f^4 + 24(\nabla f/f)^5, \\
\nabla^6 \log f &= \nabla^6 f/f - 6\nabla^5 f\nabla f/f^2 - 15\nabla^4 f\nabla^2 f/f^2 + 30\nabla^4 f(\nabla f)^2/f^3 - 10(\nabla^3 f)^2/f^2 \\
&\quad + 120\nabla^3 f \nabla^2 f\nabla f/f^3 - 120\nabla^3 f(\nabla f)^3/f^4 + 30(\nabla^2 f)^3/f^3 \\
&\quad - 270(\nabla^2 f)^2(\nabla f)^2/f^4 + 360\nabla^2 f(\nabla f)^4/f^5 - 120(\nabla f)^6/f^6, \qquad (96) \\
\nabla^3 f/f &= \nabla^3 \log f + 3\nabla^2 \log f \nabla \log f + (\nabla \log f)^3, \\
\nabla^4 f/f &= \nabla^4 \log f + 4\nabla^3 \log f\nabla \log f + 3(\nabla^2 \log f)^2 + 6\nabla^2 \log f(\nabla \log f)^2 + (\nabla \log f)^4, \\
\nabla^5 f/f &= \nabla^5 \log f + 5\nabla^4 \log f\nabla \log f + 10\nabla^3 \log f\nabla^2 \log f + 10\nabla^3 \log f(\nabla \log f)^2 \\
&\quad + 15(\nabla^2 \log f)^2 \nabla \log f + 10\nabla^2 \log f(\nabla \log f)^3 + (\nabla \log f)^5, \\
\nabla^6 f/f &= \nabla^6 \log f + 6\nabla^5 \log f\nabla \log f + 15\nabla^4 \log f\nabla^2 \log f + 15\nabla^4 \log f(\nabla \log f)^2 \\
&\quad + 10(\nabla^3 \log f)^2 + 60\nabla^3 \log f\nabla^2 \log f\nabla \log f + 20\nabla^3 \log f(\nabla \log f)^3 \\
&\quad + 15(\nabla^2 \log f)^3 + 45(\nabla^2 \log f)^2(\nabla \log f)^2 + 15\nabla^2 \log f(\nabla \log f)^4 + (\nabla \log f)^6.
\end{aligned}
$$

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

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

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

12.2.3 Auxiliary Lemmas

Henceforth, we suppress the conditioning variable $\mathbf{W}^n_{-m}$ from the conditioning sets and conditional densities unless confusions might arise. The following Lemma provides bounds on $\Phi^{\mathcal{I}(j)}_{\vartheta\mathcal{T}(j)}[\mathcal{F}]$ defined in (7) and (94) and is used in the proof of Lemma 3. For $j = 2, \ldots, 6$, define $\|\phi^i_t\|_\infty := \sup_{\vartheta \in \mathcal{N}^*} \sup_{x,x'} |\phi^i(\vartheta, Y_t, x, \overline{\mathbf{Y}}_{t-1}, x')|$ and $\|\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}\|_\infty := \sum_{(\ell_1,\ldots,\ell_j)\in\sigma(\mathcal{I}(j))} \|\phi^{\ell_1}_{t_1}\|_\infty \cdots \|\phi^{\ell_j}_{t_j}\|_\infty$.

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 $j = 2, \ldots, 6$,*

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

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

Recall that $\phi^i(\vartheta, \overline{\mathbf{Z}}^t_{t-1})$ depends on X_t and X_{t-1}. Parts (c) and (d) follow from Lemma 10(a) and the fact that, for any two probability measures μ_1 and μ_2, $\sup_{f(x):\max_x|f(x)|\leq 1}|\int f(x)d\mu_1(x) - \int f(x)d\mu_2(x)| = 2\|\mu_1 - \mu_2\|_{TV}$ (see, e.g., Levin et al. (2009, Proposition 4.5)). Similarly, parts (e) and (f) for $t_j \leq n-1$ follow from Lemma 10(b), and parts (e) and (f) for $t_j = n$ follow from $|\Phi^{\mathcal{I}(j)}_{\vartheta\mathcal{T}(j)}[\cdot]| \leq 2^j\|\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}\|_\infty$.

We proceed to show parts (a) and (b). The results for $j = 2$ and $j = 3$ follow from Lemma 10(c) and

$$
\begin{aligned}
E(X_{t_1} - EX_{t_1})\cdots(X_{t_j} - EX_{t_j}) &= \mathrm{cov}[X_{t_1}, (X_{t_2} - EX_{t_2})\cdots(X_{t_j} - EX_{t_j})]\\
&= \mathrm{cov}[(X_{t_1} - EX_{t_1})\cdots(X_{t_{j-1}} - EX_{t_{j-1}}), X_{t_j}].
\end{aligned} \tag{97}
$$

Before proving the results for $j \geq 4$, we collect some results. For a conditioning set $\mathcal{F} = \overline{\mathbf{Y}}^n_{-m}$ or $\{\overline{\mathbf{Y}}^n_{-m}, X_m = x\}$, Lemma 10(c) and (97) imply that

$$
|\mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_j-t_{j-1}-1)_+}\|\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}\|_\infty, \tag{98}
$$

$$
\begin{aligned}
&|\mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\cdots\phi^{\ell_k}_{t_k}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_{k+1}}_{t_{k+1}}\cdots\phi^{\ell_j}_{t_j}|\mathcal{F}]|\\
&\quad = |\mathrm{cov}_\vartheta[\tilde{\phi}^{\ell_1}_{t_1}\cdots\tilde{\phi}^{\ell_k}_{t_k}, \tilde{\phi}^{\ell_{k+1}}_{t_{k+1}}\cdots\tilde{\phi}^{\ell_j}_{t_j}|\mathcal{F}]| \leq \mathcal{C}\rho^{(t_{k+1}-t_k-1)_+}\|\phi^{\mathcal{I}(j)}_{\mathcal{T}(j)}\|_\infty \quad \text{for any } 2 \leq k \leq j-2.
\end{aligned} \tag{99}
$$

Parts (a)–(b) hold for $j = 4$ because $\Phi^{\mathcal{I}(4)}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] \leq \mathcal{C}\rho^{(t_2-t_1-1)_+\vee(t_4-t_3-1)_+}\|\phi^{\mathcal{I}(4)}_{\mathcal{T}(4)}\|_\infty$ from (98) and we have $\Phi^{\mathcal{I}(4)}_{\vartheta\mathcal{T}(4)}[\mathcal{F}] \leq \mathcal{C}\rho^{(t_3-t_2-1)_+}\|\phi^{\mathcal{I}(4)}_{\mathcal{T}(4)}\|_\infty$ from writing $\tilde{\Phi}^{\ell_1\ell_2\ell_3\ell_4}_{\vartheta\mathcal{T}(4)}$ defined in (7) as $\tilde{\Phi}^{\ell_1\ell_2\ell_3\ell_4}_{\vartheta\mathcal{T}(4)} = \mathrm{cov}_\vartheta[\tilde{\phi}^{\ell_1}_{t_1}\tilde{\phi}^{\ell_2}_{t_2}, \tilde{\phi}^{\ell_3}_{t_3}\tilde{\phi}^{\ell_4}_{t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\phi^{\ell_3}_{t_3}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_2}_{t_2}\phi^{\ell_4}_{t_4}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi^{\ell_1}_{t_1}\phi^{\ell_4}_{t_4}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi^{\ell_2}_{t_2}\phi^{\ell_3}_{t_3}|\mathcal{F}]$ and applying (99). Parts (a)–(b) for $j = 5$ follows from a similar argument.

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 (98). Second,

write $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}] = A_1 + A_2$, where $A_1 = \mathbb{E}^c_\vartheta\left[\phi_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}\right] - \mathbb{E}^c_\vartheta\left[\phi_{t_1}\phi_{t_2}\phi_{t_3}|\mathcal{F}\right]\mathbb{E}^c_\vartheta\left[\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}\right]$ and A_2 denotes all the terms on the right hand side of $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}]$ in (94) except for A_1. A_1 is bounded by $\mathcal{C}\rho^{(t_4-t_3-1)_+}\|\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}\|_\infty$ from (99), and A_2 is bounded by $\mathcal{C}\rho^{(t_4-t_3-1)_+}\|\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}\|_\infty$ from (98). 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, write $\Phi^{\mathcal{I}(6)}_{\vartheta\mathcal{T}(6)}[\mathcal{F}] = B_1 + B_2 + B_3$, where $B_1 = \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}] - \mathbb{E}^c_\vartheta[\phi_{t_3}\phi_{t_4}\phi_{t_5}\phi_{t_6}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}|\mathcal{F}]$, $B_2 = -\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\mathbb{E}^c_\vartheta[\phi_1\phi_2\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}] + 2\sum_{(\{a,b\},\{c,d\},\{e,f\})\in X_{63}}\mathbb{E}^c_\vartheta[\phi_{t_a}\phi_{t_b}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]$, 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
$X_{63} := \{(\{1,2\},\{3,4\},\{5,6\}),(\{1,2\},\{3,5\},\{4,6\}),(\{1,2\},\{3,6\},\{4,5\})\}$, and B_3 denotes 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. B_1 is bounded by $\mathcal{C}\rho^{(t_3-t_2-1)_+}\|\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}\|_\infty$ from (99). We can write B_2 as $\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\{-\mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}] + \mathbb{E}^c_\vartheta[\phi_{t_1}\phi_{t_2}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_c}\phi_{t_d}|\mathcal{F}]\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]\} = -\sum_{(\{1,2,c,d\},\{e,f\})\in X_{61}}\mathbb{E}^c_\vartheta[\phi_{t_e}\phi_{t_f}|\mathcal{F}]\text{cov}_\vartheta[\tilde{\phi}_{\theta t_1}\tilde{\phi}_{\theta t_2},\tilde{\phi}_{\theta t_c}\tilde{\phi}_{\theta t_d}|\mathcal{F}]$, then this is bounded by $\mathcal{C}\rho^{(t_3-t_2-1)_+}\|\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}\|_\infty$ from (99). Finally, B_3 is bounded by $\mathcal{C}\rho^{(t_3-t_2-1)_+}\|\phi^{\mathcal{I}(6)}_{\mathcal{T}(6)}\|_\infty$ from (98). 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 $i_1 = i_2 = i_3 = i_4 = i_5$; $(q_{i_1}/3 \wedge q_{i_2}/6)$ 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 $K_{\mathcal{I}(j)}, \{M_{\mathcal{I}(j),k}\}_{k=1}^n \in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ such that, for all $1 \le k \le n$ and $m' \ge m \ge 0$,*

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

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

(c) $\sup_{m\ge0}\sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)| + \sup_{m\ge0}\sup_{\vartheta\in\mathcal{N}^*}|\overline{\Delta}^{\mathcal{I}(j)}_{j,k,m}(\vartheta)| \le M_{\mathcal{I}(j),k}$ $\mathbb{P}_{\vartheta^*}$-a.s., (d) Uniformly in $\vartheta \in \mathcal{N}^*$ and $x \in \mathcal{X}$, $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$ 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. First, we prove parts (a) and (b). Recall $\mathcal{T}(j) = (t_1,\ldots,t_j)$. For part (a), define, suppressing the dependence of $A_{\mathcal{T}(j)}$ on ϑ and $\mathcal{I}(j)$,

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

$$A_{\mathcal{T}(j,\ell,k)} := A_{t_1t_2\cdots t_{j-\ell}\underbrace{k\cdots k}_{\ell\text{ times}}}, \quad \text{where } \mathcal{T}(j,\ell,k) := (\mathcal{T}(j-\ell), \underbrace{k,\cdots,k}_{\ell\text{ times}}).$$

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

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

and $\Delta_b := 0$ 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

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

From $(t-1)_+\ge\lfloor t/2\rfloor$ and Lemma 12, $B_{j,k,m}$ 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 $i_1 = i_2 = \cdots = i_j$, it follows from Lemma 13 that $M^{\mathcal{I}(j)}_{j,k,m}\le(k+m)^{j+1}\|\phi^{i_1}\|^j_\infty$, and $\|\phi^{i_1}\|^j_\infty\in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ from Assumption 4. In the other cases, observe that, if $x,y,z\ge 0$, we have $xy\le x^2+y^2$, $xyz\le x^3+y^3+z^3$, $xy\le x^4+y^{4/3}$, and $xy\le x^3+y^{3/2}$ from Young's inequality. Using this result and Lemma 13, we can bound $M^{\mathcal{I}(j)}_{j,k,m}$ by

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

Therefore, from Assumption 4, Δ_a is bounded by the right hand side of part (a). From Lemmas 2 and 12, Δ_b is bounded by $\mathcal{C}\sum_{\ell=1}^{j-1}\sum_{-m+1\le t_1\le\cdots\le t_{j-\ell}\le k-1}(\rho^{(m+t_1-1)_+}\wedge\rho^{(t_2-t_1-1)_+}\wedge\cdots\wedge\rho^{(k-t_{j-\ell}-1)_+})M^{\mathcal{I}(j)}_{j,k+1,m} \le \mathcal{C}\rho^{\lfloor(k+m-1)/4(j-1)\rfloor}M^{\mathcal{I}(j)}_{j,k+1,m}$. Similarly, Δ_c is bounded by

$\mathcal{C}\rho^{\lfloor (k+m-1)/4(j-1)\rfloor} M^{\mathcal{I}(j)}_{j,k+1,m}$, and part (a) of the lemma follows.

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

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

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

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

By the same argument as part (a), $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\theta) - \Delta_d$ is bounded by the right hand of part (a). For Δ_e, observe that, with $M_j := \max_{1\le\ell\le j}\binom{j}{\ell}$,

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

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 14.

For part (c), observe that $\sup_{m\ge0}\sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)| \le A+B$, where $A := \sup_{m\ge0}\sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta) - \Delta^{\mathcal{I}(j)}_{j,k,0,x}(\vartheta)|$ and $B := \sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,k,0,x}(\vartheta)|$. A is bounded by $K_{\mathcal{I}(j)}k^7\rho^{\lfloor (k-1)/1340\rfloor}$ from part (b). B does not depend on m and is distributionally equivalent to $\sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,1,k-1,x}(\vartheta)|$. This is bounded by $\sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,1,k-1,x}(\vartheta) - \Delta^{\mathcal{I}(j)}_{j,1,0,x}(\vartheta)| + \sup_{x\in\mathcal{X}}\sup_{\vartheta\in\mathcal{N}^*}|\Delta^{\mathcal{I}(j)}_{j,1,0,x}(\vartheta)|$. The first term is in $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ from part (b), and the second term is in $L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ from the definition of $\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)$. Therefore, there exists $M_{\mathcal{I}(j),k} \in L^{r_{\mathcal{I}(j)}}(\mathbb{P}_{\vartheta^*})$ such that $A+B \le M_{\mathcal{I}(j),k}$, and part (c) holds in view of part (a). Part (d) follows from parts (a)–(c) because parts (a)–(c) imply that $\{\Delta^{\mathcal{I}(j)}_{j,k,m,x}(\vartheta)\}_{m\ge0}$ and $\{\Delta^{\mathcal{I}(j)}_{j,k,m}(\vartheta)\}_{m\ge0}$ 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. □

Lemma 4. *Under Assumptions 1, 2, and 4, there exist random variables* $\{K_k\}_{k=1}^n \in$

$L^{(1+\varepsilon)q_\vartheta/\varepsilon}(\mathbb{P}_{\vartheta^*})$ *and* $\rho \in (0,1)$ *such that, for all* $1 \le k \le n$ *and* $m' \ge m \ge 0$,

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

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

Proof of Lemma 4. The first result follows from noting that $\overline{p}_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}) = \sum_{(x_{k-1},x_k)\in\mathcal{X}^2} g_\vartheta(Y_k|\overline{\mathbf{Y}}_{k-1},x_k) \times q_{\vartheta_x}(x_{k-1},x_k)\mathbb{P}_\vartheta(x_{k-1}|\overline{\mathbf{Y}}_{-m}^{k-1}) \in [\sigma_- G_{\vartheta k}, \sigma_+ G_{\vartheta k}]$ and using Assumption 4(b). For the second result, observe that $|p_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, X_{-m}=x) - \overline{p}_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1})| \le \sum_{(x_{k-1},x_k)\in\mathcal{X}^2} g_\vartheta(Y_k|\overline{\mathbf{Y}}_{k-1},x_k) q_{\vartheta_x}(x_{k-1},x_k) \times |\mathbb{P}_\vartheta(x_{k-1}|\overline{\mathbf{Y}}_{-m}^{k-1}, X_{-m}=x) - \mathbb{P}_\vartheta(x_{k-1}|\overline{\mathbf{Y}}_{-m}^{k-1})| \le \rho^{k+m-1}\sigma_+ G_{\vartheta k}/\sigma_-$, where the second inequality follows from Lemma 10(a). The second result then follows from writing the left hand side as

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

noting that $p_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m}^{k-1}, X_{-m}=x) \ge \sigma_- G_{\vartheta k}$, and using the derived bounds. The results with $p_\vartheta(Y_k|\overline{\mathbf{Y}}_{-m'}^{k-1}, X_{-m'}=x)$ and $p_{\vartheta^*}(Y_k|\overline{\mathbf{Y}}_{-m'}^{k-1}, X_{-m'}=x)$ 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 $\nabla_{u^j} f(c_1\lambda_\mu, c_2u^2)|_{\lambda_\mu=u=0}$ is 0 if $j=1,3$, is $2c_2\nabla_h f(0,h(0))$ if $j=2$, and is $12c_2^2\nabla_{h^2} f(0,h(0))$ if $j=4$. Therefore, the stated result follows. $\square$

Lemma 6. *Suppose the assumptions of Proposition 10 hold. Then, there exist* $\bar{\varrho}_1, \bar{\varrho}_2, \bar{\varrho}_3 \in (0,\varrho)$

such that, for all $k \geq 1$,

$$(a)\quad \frac{\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \varrho\frac{\nabla_\varrho\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\bar{\varrho}_1\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\bar{\varrho}_1\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})},$$

$$(b)\quad \frac{\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} - b(\alpha)\frac{\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}$$
$$= \varrho\frac{\nabla_\varrho\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*\bar{\varrho}_2\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\bar{\varrho}_2\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} - \varrho\frac{\nabla_\varrho\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*\bar{\varrho}_3\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^*\bar{\varrho}_3\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}.$$

Proof of Lemma 6. Part (a) holds if

$$\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0, \tag{100}$$

because (i) $\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\varrho\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) - \nabla_{\lambda_\mu^3}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \nabla_\varrho\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\bar{\varrho}\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})\varrho$ for $\bar{\varrho} \in (0, \varrho)$ from the mean value theorem and (ii) $\overline{p}_{\psi^*\varrho\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})$ does not depend on the value of ϱ.

We proceed to show (100). Note that$\nabla_{\lambda_\mu^3}\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \nabla_{\lambda_\mu^3}\log\overline{p}_{\psi^*\pi}(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0) - \nabla_{\lambda_\mu^3}\log\overline{p}_{\psi^*\pi}(\mathbf{Y}_1^{k-1}|\overline{\mathbf{Y}}_0)$ from (96) and $\nabla_\lambda\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$. Let $\nabla^i\ell_t^* := \nabla_{\lambda_\mu^i}\log g_t^*$ with $\nabla\ell_t^* = \nabla^1\ell_t^*$. Observe that

$$\begin{aligned}
\nabla_{\lambda_\mu^3}\log\overline{p}_{\psi^*0\alpha}(\mathbf{Y}_1^k|\overline{\mathbf{Y}}_0) &= \sum_{t=1}^k \mathbb{E}_{\psi^*0\alpha}\left[\nabla^3\ell_t^*\middle|\overline{\mathbf{Y}}_0^k\right] + 3\sum_{t_1=1}^k\sum_{t_2=1}^k \mathbb{E}_{\psi^*0\alpha}\left[\nabla^2\ell_{t_1}^*\nabla\ell_{t_2}^*\middle|\overline{\mathbf{Y}}_0^k\right] \\
&\quad + \sum_{t_1=1}^k\sum_{t_2=1}^k\sum_{t_3=1}^k \mathbb{E}_{\psi^*0\alpha}\left[\nabla\ell_{t_1}^*\nabla\ell_{t_2}^*\nabla\ell_{t_3}^*\middle|\overline{\mathbf{Y}}_0^k\right] \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*0\alpha}\left[\nabla^3\ell_t^* + 3\nabla^2\ell_t^*\nabla\ell_t^* + \nabla\ell_t^*\nabla\ell_t^*\nabla\ell_t^*\middle|\overline{\mathbf{Y}}_0^k\right] \\
&= \sum_{t=1}^k \mathbb{E}_{\psi^*0\alpha}\left[\nabla_{\lambda_\mu^3}g_t^*/g_t^*\middle|\overline{\mathbf{Y}}_0^k\right],
\end{aligned} \tag{101}$$

where the first equality follows from Lemma 1, the second equality holds because (i) X_t is serially independent when $\varrho = 0$ and (ii) $\nabla\ell_t^* = d_{1t}\nabla_\mu f_t^*/f_t^*$ and $\nabla^2\ell_t^* = d_{2t}\nabla_\mu^2 f_t^*/f_t^* - (d_{1t}\nabla_\mu f_t^*/f_t^*)^2$, and (iii) $\mathbb{E}_{\psi^*0\alpha}[d_{1t}|\overline{\mathbf{Y}}_0^k] = \mathbb{E}_{\psi^*0\alpha}[d_{2t}|\overline{\mathbf{Y}}_0^k] = 0$ from (30), and the third equality follows from (96) The right hand side is 0 from (30), and hence part (a) is proven.

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

$$\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = b(\alpha)\nabla_{\lambda_\sigma^2}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}). \tag{102}$$

Observe that $\nabla_{\lambda_\mu^4}\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \nabla_{\lambda_\mu^4}\log\overline{p}_{\psi^*0\alpha}(\mathbf{Y}_0^k|\overline{\mathbf{Y}}_0) - \nabla_{\lambda_\mu^4}\log\overline{p}_{\psi^*0\alpha}(\mathbf{Y}_0^{k-1}|\overline{\mathbf{Y}}_0)$ from (96), $\nabla_\lambda\overline{p}_{\psi^*\pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$, and $\nabla_{\lambda_\mu^2}\log\overline{p}_{\psi^*0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = 0$. A similar derivation to (101) gives

$$\nabla_{\lambda_\mu^4} \log \overline{p}_{\psi^* 0\alpha}(\mathbf{Y}_0^k | \overline{\mathbf{Y}}_0) = \sum_{t=1}^{k} \mathbb{E}_{\psi^* 0\alpha} \left[\nabla_{\lambda_\mu^4} g_t^* / g_t^* \middle| \overline{\mathbf{Y}}_0^k \right]. \tag{103}$$

(102) follows from (103) because (i) $\nabla_{\lambda_\sigma^2} \overline{p}_{\psi^* 0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})/\overline{p}_{\psi^* 0\alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1}) = \mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2} g_k^* | \overline{\mathbf{Y}}_0^k]$ from a similar argument to (18) and (ii) $\mathbb{E}_{\psi^* 0\alpha}[\nabla_{\lambda_\mu^4} g_t^* / g_t^* | \overline{\mathbf{Y}}_0^k] = b(\alpha)\mathbb{E}_{\vartheta^*}[\nabla_{\lambda_\sigma^2} g_k^* | \overline{\mathbf{Y}}_0^k]$ from (30). Therefore, part (b) is proven. □

Lemma 7. *Suppose the assumptions of Proposition 12 hold. Then, there exist $\bar{\varrho}_1, \bar{\varrho}_2 \in (0, \varrho)$ such that, for all $k \geq 1$,*

$$(a) \quad \frac{\nabla_{\lambda_\mu^3} \overline{p}_{\psi^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \alpha(1-\alpha)(1-2\alpha)\frac{\nabla_{\mu^3} f_k^*}{f_k^*} + \varrho \frac{\nabla_\varrho \nabla_{\lambda_\mu^3} \overline{p}_{\psi^* \bar{\varrho}_1 \alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^* \bar{\varrho}_1 \alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})},$$

$$(b) \quad \frac{\nabla_{\lambda_\mu^4} \overline{p}_{\psi^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^* \pi}(Y_k|\overline{\mathbf{Y}}_0^{k-1})} = \alpha(1-\alpha)(1-6\alpha+6\alpha^2)\frac{\nabla_{\mu^4} f_k^*}{f_k^*} + \varrho \frac{\nabla_\varrho \nabla_{\lambda_\mu^4} \overline{p}_{\psi^* \bar{\varrho}_2 \alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}{\overline{p}_{\psi^* \bar{\varrho}_2 \alpha}(Y_k|\overline{\mathbf{Y}}_0^{k-1})}.$$

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

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

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

Observe that equalities (101) and (103) in the proof of Lemma 6 still hold under the assumptions of Proposition 12 if we use (44) in place of (30). Consequently, (A) and (B) follow from (43), (44), and the argument of the proof of Lemma 6, and the stated result follows. □

Lemma 8. *Suppose that the assumptions of Propositions 9 hold. Let $\boldsymbol{C}_\eta$ be a set of sequences $\{\eta_n\}$ satisfying $\sqrt{n}(\eta_n - \eta^*) \to h_\eta$ for some finite h_η. Let $\mathbb{P}_{\eta_n}^n := \prod_{k=1}^n f_k(\eta_n, 0)$ denote the probability measure under η_n with $\lambda_n = 0$. Then, for every sequence $\{\eta_n\} \in \boldsymbol{C}_\eta$, the LRTS under $\{\mathbb{P}_{\eta_n}^n\}$ converges in distribution $\sup_{\varrho \in \Theta_\varrho} \left(\tilde{t}_{\lambda\varrho}' \mathcal{I}_{\lambda.\eta\varrho} \tilde{t}_{\lambda\varrho} \right)$ given in Propositions 9.*

Proof of Lemma 8. Observe that $\vartheta_n := (\pi_n, \eta_n, \lambda_n) = (\pi, \eta^* + h_\eta/\sqrt{n}, 0)$ satisfies the assumptions of Proposition 18. Therefore, Proposition 18 holds under ϑ_n with $\nu_n(s_{\varrho n k}) \to_d N(\mathcal{I}_\varrho h, \mathcal{I}_\varrho)$ with $h = (h_\eta', 0)'$ under $\mathbb{P}_{\vartheta_n}^n$. Furthermore, the log-likelihood function of the one-regime model admits a similar expansion, and $\log(d\mathbb{P}_{\eta_n}^n / d\mathbb{P}_{\eta^*}^n) = h_\eta' \nu_n(s_{\eta k}) - (1/2)h_\eta' \mathcal{I}_\eta h_\eta + o_p(1)$ holds under $\mathbb{P}_{\eta_n}^n$. Therefore, the proof of Proposition 9 goes through by replacing $G_{\varrho n}$ with $G_{\varrho n}^h = \begin{bmatrix} G_{\eta n}^h \\ G_{\lambda \varrho n}^h \end{bmatrix} := G_{\varrho n} + \mathcal{I}_\varrho h$. In view of $G_{\eta n}^h = G_{\eta n} + \mathcal{I}_\eta h_\eta$ and $G_{\lambda \varrho n}^h = G_{\lambda \varrho n} + \mathcal{I}_{\lambda \eta \varrho} h_\eta$, we have $G_{\lambda.\eta \varrho n}^h := G_{\lambda \varrho n}^h - \mathcal{I}_{\lambda \eta \varrho} \mathcal{I}_\eta^{-1} G_{\eta n}^h = G_{\lambda \varrho n} - \mathcal{I}_{\lambda \eta \varrho} \mathcal{I}_\eta^{-1} G_{\eta n} = G_{\lambda \varrho n}$. Therefore, the asymptotic distribution of the LRTS under $\mathbb{P}_{\eta^n}^n$ is the same as that under $\mathbb{P}_{\eta^*}^n$, and the stated result follows. □

12.2.4 Bounds on difference in state probabilities and conditional moments

Lemma 9. *Suppose* $X_1, \ldots, X_n$ *are random variables with* $\max_{1\le i\le n} \mathbb{E}|X_i|^q < C$ *for some* $q > 0$ *and* $C \in (0, \infty)$*. Then,* $\max_{1\le i\le n} |X_i| = o_p(n^{1/q})$*.*

Proof of Lemma 9. For any $\varepsilon > 0$, we have $\mathbb{P}(\max_{1\le i\le n} |X_i| > \varepsilon n^{1/q}) \le \sum_{1\le i\le n} \mathbb{P}(|X_i| > \varepsilon n^{1/q})$ $\le \varepsilon^{-q} n^{-1} \sum_{1\le i\le n} \mathbb{E}(|X_i|^q \mathbb{I}\{|X_i| > \varepsilon n^{1/q}\})$ by a version of Markov inequality. As $n \to \infty$, the right hand side tends to 0 by the Dominated Convergence Theorem. □

The following Lemma extends 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 $\mathbf{W}_{-m}^n$ 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 $\sup_{f(x):0\le f(x)\le 1} |\int f(x) d\mu_1(x) - \int f(x) d\mu_2(x)| = \|\mu_1 - \mu_2\|_{TV}$. In the following, we define $\overline{\mathbf{V}}_{-m}^n := (\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n)$, and we let ~~$x_{\overline{m}}$~~ $\boldsymbol{x_{-m}}$ denote "$X_{-m} = x_{-m}$."

Lemma 10. *Suppose Assumptions 1-2 hold and* $\vartheta_x \in \Theta_x$*. Then,*
(a) For all $-m \le t_1 \le t_2$ *with* $-m < n$ *and all probability measures* μ_1 *and* μ_2 *on* $\mathcal{B}(\mathcal{X})$*,*

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

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

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

The same bound holds when x_{-m} *is dropped from the conditioning variables.*
(c) For all $-m \le t_1 \le t_2 < t_3 \le t_4$ *with* $-m < n$*,*

$$\left\| \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in \cdot, \mathbf{X}_{t_3}^{t_4} \in \cdot | \overline{\mathbf{V}}_{-m}^n, x_{-m}) - \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in \cdot | \overline{\mathbf{V}}_{-m}^n, x_{-m}) \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_3}^{t_4} \in \cdot | \overline{\mathbf{V}}_{-m}^n, x_{-m}) \right\|_{TV} \le \rho^{t_3-t_2}.$$

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

Proof of Lemma 10. We prove part (a) first. We assume $t_1 > -m$ because the stated result holds trivially when $t_1 = -m$. Observe that Lemma 1 of DMR still holds when $\mathbf{W}_{-m}^n$ is added to the conditioning variable because Assumption 1 implies that $\{(X_k, \overline{\mathbf{Y}}_k)\}_{k=0}^{\infty}$ is a Markov chain given $\{W_k\}_{k=0}^{\infty}$. Therefore, $\{X_t\}_{t\ge -m}$ is a Markov chain when conditioned on $\{\overline{\mathbf{Y}}_{-m}^n, \mathbf{W}_{-m}^n\}$, and hence $\mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in A | \overline{\mathbf{V}}_{-m}^n, x_{-m}) = \sum_{x_{t_1}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in A | X_{t_1} = x_{t_1}, \overline{\mathbf{V}}_{-m}^n) p_{\vartheta_x}(x_{t_1} | \overline{\mathbf{V}}_{-m}^n, x_{-m})$ holds. From applying this result and the property of the total variation distance, we can bound the left hand side of the lemma by $\| \sum_{x_{-m}\in\mathcal{X}} p_{\vartheta_x}(X_{t_1} \in \cdot | \overline{\mathbf{V}}_{-m}^n, x_{-m}) \mu_1(x_{-m}) - \sum_{x_{-m}\in\mathcal{X}} p_{\vartheta_x}(X_{t_1} \in \cdot | \overline{\mathbf{V}}_{-m}^n, x_{-m}) \mu_2(x_{-m}) \|_{TV}$. This is bounded by ρ^{t_1+m} from Corollary 1 of DMR, which holds when $\mathbf{W}_{-m}^n$ is added to the conditioning variable. Therefore, part (a) is proven.

We proceed to prove part (b). Observe that the time-reversed process $\{Z_{n-k}\}_{0\le k\le n+m}$ is Markov when conditioned on $\mathbf{W}_{-m}^n$ and that W_k is independent of $(\mathbf{X}_0^{k-1}, \overline{\mathbf{Y}}_0^{k-1})$ given $\mathbf{W}_0^{k-1}$.

Consequently, for $k = n, n-1$, we have $\mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in A|\overline{\mathbf{V}}_{-m}^k, x_{-m}) = \sum_{x_{t_2}\in\mathcal{X}} \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in A|X_{t_2} = x_{t_2}, \overline{\mathbf{V}}_{-m}^{t_2}, x_{-m}) p_{\vartheta_x}(x_{t_2}|\overline{\mathbf{V}}_{-m}^k, x_{-m})$. Therefore, from the property of the total variation distance, the left hand side of the lemma is bounded by $\|\mathbb{P}_{\vartheta_x}(X_{t_2} \in \cdot|\overline{\mathbf{V}}_{-m}^n, x_{-m}) - \mathbb{P}_{\vartheta_x}(X_{t_2} \in \cdot|\overline{\mathbf{V}}_{-m}^{n-1}, x_{-m})\|_{TV}$. This is bounded by ρ^{n-1-t_2} because equation (39) of DMR p. 2294 holds when $\mathbf{W}_{-m}^n$ is added to the conditioning variables, and the stated result follows. When x_{-m} is dropped from the conditioning variables, part (b) follows from a similar argument with using Lemma 9 and an analogue of Corollary 1 of DMR in place of equation (39) of DMR.

Part (c) follows immediately from writing the left hand side of lemma as $\sup_{A,B} |\mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_1}^{t_2} \in A|\overline{\mathbf{V}}_{-m}^n, x_{-m})[\mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_3}^{t_4} \in B|\overline{\mathbf{V}}_{-m}^n, \mathbf{X}_{t_1}^{t_2} \in A) - \mathbb{P}_{\vartheta_x}(\mathbf{X}_{t_3}^{t_4} \in B|\overline{\mathbf{V}}_{-m}^n, x_{-m})]|$ and applying part (a). □

12.2.5 The sums of powers of ρ

Lemma 11. *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 11. 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 12. *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 12. When $k = 1$, the stated result follows from Lemma 11 with $c = 1$. We first show that the following holds for $k \geq 2$:

$$\sum_{t_1 \leq t_2 \leq \cdots \leq t_k \leq n} \left(\rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \cdots \wedge \rho^{\lfloor (t_k-t_{k-1})/q \rfloor} \wedge \rho^{\lfloor (n-t_k)/q \rfloor} \right) \leq \frac{q^{k-1}(k+1)!\rho^{\lfloor (n-t_1)/kq \rfloor}}{(1-\rho)^{k-1}}. \tag{104}$$

We prove (104) by induction. When $k=2$, it follows from Lemma 11 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 (104). Suppose (104) holds when $k=\ell$. Then (104) holds when $k=\ell+1$ because, from Lemma 11,

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

and hence (104) holds for all $k \geq 2$. We proceed to show the stated result. Observe that

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

where the first inequality holds by symmetry, and the subsequent equality follows from $n-t_k \geq t_1$. From (104), the right hand side is no larger than $q^{k-1}(k+1)!(1-\rho)^{(1-k)} \sum_{t_1=0}^{n/2} \rho^{\lfloor (n-t_1)/kq \rfloor} \leq q^k k(k+1)!(1-\rho)^{-k}\rho^{\lfloor n/2kq \rfloor}$, 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 13. *Let $a_j > 0$ for all j. For all positive integer $\ell \geq 1$ and all $k \geq 1$ and $m \geq 0$, we have $\max_{-m+1 \leq t_1,\ldots,t_\ell \leq k} a_{t_1} \cdots a_{t_\ell} \leq (k+m)^{\ell+1} A_\ell$, where $A_\ell := \sum_{t=-\infty}^{\infty} (|t| \vee 1)^{-2} a_t^\ell$.*

Proof of Lemma 13. When $\ell = 1$, the stated result follows from $\max_{-m+1 \leq t \leq k} a_t \leq \sum_{t=-m+1}^{k} a_t = \sum_{t=-m+1}^{k} (|t| \vee 1)^2 (|t| \vee 1)^{-2} a_t \leq (k+m)^2 \sum_{t=-\infty}^{\infty} (|t| \vee 1)^{-2} a_t$. When $\ell \geq 2$, from the Hölder's inequality, we have $\max_{-m+1\le t_1\le\ldots\le t_\ell\le k} a_{t_1}a_{t_2}\cdots a_{t_\ell} \le (\sum_{t=-m+1}^k a_t)^\ell = [\sum_{t=-m+1}^k (|t|\vee 1)^{2/\ell}(|t|\vee 1)^{-2/\ell}a_t]^\ell \le [\sum_{t=-m+1}^k (|t|\vee 1)^{2/(\ell-1)}]^{(\ell-1)} \sum_{t=-m+1}^k (|t|\vee 1)^{-2}a_t^\ell \le [(k+m)^{1+2/(\ell-1)}]^{\ell-1}A_\ell = (k+m)^{\ell+1}A_\ell$. □

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

Lemma 14. *For $\alpha > 0$, $q > 0$, and $c_{jt} \ge 0$, define $c_{jq}^\infty(\rho^\alpha) := \sum_{t=-\infty}^\infty \rho^{\lfloor \alpha|t|/q\rfloor} c_{jt}$. For all $\rho \in (0,1)$, $k \ge 1$, and $0 \le m \le m'$,*

$$\sum_{t_1=-m'+1}^{-m} \sum_{t_1\le t_2\le t_3\le t_4\le t_5\le t_6\le k} \Big(\rho^{\lfloor (k-1-t_6)/q\rfloor} \wedge \rho^{\lfloor (t_6-t_5)/q\rfloor} \wedge \rho^{\lfloor (t_5-t_4)/q\rfloor} \wedge \rho^{\lfloor (t_4-t_3)/q\rfloor} \wedge$$
$$\rho^{\lfloor (t_3-t_2)/q\rfloor} \wedge \rho^{\lfloor (t_2-t_1)/q\rfloor}\Big) \prod_{j=1}^6 c_{jt_j} \le \rho^{\lfloor (k-1+m)/2qa_7\rfloor} c_{1q}^\infty\left(\rho^{1/2a_7}\right) \prod_{j=2}^6 c_{jq}^\infty\left(\rho^{1/4a_j}\right), \tag{105}$$

where (a_j, b_j) are defined recursively with $(a_2, b_2) = (1,1)$ and, for $j \ge 3$, $a_{j+1} = 4a_j(a_j+b_j)/(2a_j-1)$ and $b_{j+1} = a_j(4b_j-1)/(2a_j-1)$. a_j and b_j satisfy $a_j, b_j \ge 3/2$ for all j. Direct calculations using Matlab produce $a_7 \doteq 334.5406$.

Proof of Lemma 14. First, observe that the following result holds for $a, b > 1/4$, $t_1 \le 0$, and $t_j, t_{j+1} \ge t_1$:

$$\begin{aligned} &(a) \text{ if } t_j \le \frac{at_{j+1}+t_1}{a+b}, \quad \text{then } \frac{|t_j|}{4a} \le \frac{a(4a+1)t_{j+1}+(2a-1)t_1}{4a(a+b)} - t_j, \\ &(b) \text{ if } t_j \ge \frac{at_{j+1}+t_1}{a+b}, \quad \text{then } \frac{|t_j|}{4a} \le \frac{b}{a}t_j - \frac{a(4b-1)t_{j+1}+(2a+4b+1)t_1}{4a(a+b)}. \end{aligned} \tag{106}$$

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

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

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

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

$$
\begin{aligned}
&\sum_{t_j=-m'+1}^{k} \left(\rho^{\lfloor (t_{j+1}-t_j)/q \rfloor} \wedge \rho^{\lfloor (b_j t_j - t_1)/a_j q \rfloor}\right) c_{jt_j} \\
&\leq \rho^{\lfloor \frac{a_j(4b_j-1)t_{j+1}-(2a_j-1)t_1}{4a_j(a_j+b_j)q} \rfloor} \left(\sum_{t_j \leq \bar{t}_j} \rho^{\lfloor \frac{a_j(4a_j+1)t_{j+1}+(2a_j-1)t_1}{4a_j(a_j+b_j)q} - \frac{t_j}{q} \rfloor} + \sum_{t_j \geq \bar{t}_j} \rho^{\lfloor \frac{b_j}{a_j q} t_j - \frac{a_j(4b_j-1)t_{j+1}+(2a_j+4b_j+1)t_1}{4a_j(a_j+b_j)q} \rfloor} \right) c_{jt_j} \\
&\leq \rho^{\lfloor \frac{a_j(4b_j-1)t_{j+1}-(2a_j-1)t_1}{4a_j(a_j+b_j)q} \rfloor} c_{jq}^{\infty} \left(\rho^{1/4a_j}\right) \\
&= \rho^{\lfloor \frac{b_{j+1}t_{j+1}-t_1}{a_{j+1}} \rfloor} c_{jq}^{\infty} \left(\rho^{1/4a_j}\right). \qquad (107)
\end{aligned}
$$

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 (106) and (107) to the left hand side of (105) sequentially for $j = 2, 3, \ldots, 6$. Consequently, the left hand side of (105) 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. □

12.2.6 Derivation of $\vartheta_{M_0+1,x} = (\vartheta_{xm}', \pi_{xm}')'$ and $\pi_{xm} = (\varrho_m, \alpha_m, \phi_m')'$

Define $\overline{J}_{m0} := \{1, \ldots, M_0\} \setminus J_m$, 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.

We parameterize the transition probability of X_k in terms of its stationary distribution and the first to the $(m-1)$th rows and the $(m+1)$th to the (M_0+1)th rows of its transition matrix.[8] For $i \in \overline{J}_m$, we reparameterize $(p_{im}, p_{i,m+1})$ to $p_{iJ} = p_{im} + p_{i,m+1} = \mathbb{P}_{\vartheta_{M_0+1}}(X_k \in J_m | X_{k-1} = i)$ and $p_{im|iJ} = p_{im}/(p_{im} + p_{i,m+1})$. Furthermore, we reparameterize (p_m, p_{m+1}) in the stationary distribution to $p_J = p_m + p_{m+1} = \mathbb{P}_{\vartheta_{M_0+1}}(X_k \in J_m)$ and $p_{m|J} = p_m/(p_m + p_{m+1}) = \mathbb{P}_{\vartheta_{M_0+1}}(X_k = m | X_k \in J_m)$. Therefore, with $\wedge$ and $\vee$ denoting "and" and "or," the transition probability of X_k is summarized by $\vartheta_{M_0+1,x} := (\{p_{ij}\}_{i \in \overline{J}_m \wedge j \in \overline{J}_{m0}}, \{p_{iJ}, p_{im|iJ}\}_{i \in \overline{J}_m}, \{p_{m+1,j}\}_{j=1}^{M_0}, \{p_j\}_{j \in \overline{J}_{m0}}, p_J, p_{m|J})$.

[8]Suppose a Markov process has a transition probability P and stationary distribution π whose elements are strictly positive. If π and all the rows of P except for one are identified, then the remaining row of P is identified from the relation $\pi P = \pi$.

Split $\vartheta_{M_0+1,x}$ as $\vartheta_{M_0+1,x} = (\vartheta'_{xm}, \pi'_{xm})'$, where

$$\vartheta_{xm} := (\{p_{ij}\}_{i\in\overline{J}_m \wedge j\in\overline{J}_{m0}}, \{p_{iJ}\}_{i\in\overline{J}_m}, \{p_j\}_{j\in\overline{J}_{m0}}, p_J) \quad \text{and}$$
$$\pi_{xm} := (\{p_{im|iJ}\}_{i\in\overline{J}_m}, \{p_{m+1,j}\}_{j=1}^{M_0}, p_{m|J}).$$

When the mth and $(m+1)$th regimes are combined into one regime, the transition probability of X_k equals the transition probability of X_k under $\vartheta^*_{M_0,x}$ if and only if $\vartheta_{xm} = \vartheta^*_{xm} := \{p_{ij} = p^*_{ij}$ for $i \in \overline{J}_m \wedge (1 \le j \le m-1); p_{ij} = p^*_{i,j-1}$ for $i \in \overline{J}_m \wedge (m+2 \le j \le M_0); p_{iJ} = p^*_{im}$ for $i \in \bar{J}_m; p_j = p^*_j$ for $1 \le j \le m-1; p_j = p^*_{j-1}$ for $m+2 \le j \le M_0; p_J = p^*_m\}$. π_{xm} is the part of $\vartheta_{M_0+1,x}$ that is not identified under H_{0m}.

We proceed to derive the reparameterization of some elements of π_{xm} in terms of (α_m, ϱ_m). First, map $p_{m+1,m}$ and $p_{m+1,m+1}$ to $p_{m+1,J} := p_{m+1,m} + p_{m+1,m+1} = \mathbb{P}_{\vartheta_{M_0+1}}(X_k \in J|X_{k-1} = m+1)$ and $p_{m+1,m|J} := p_{m+1,m}/p_{m+1,J} = \mathbb{P}_{\vartheta_{M_0+1}}(X_k = m|X_k \in J, X_{k-1} = m+1)$. Let P_J and π_J denote the transition matrix and stationary distribution of X_k restricted to lie in J_m. The second row of P_J is given by $(p_{m+1,m|J}, 1 - p_{m+1,m|J})$, and π_J is given by $(p_{m|J}, 1 - p_{m|J})$. From the relation $\pi_J = \pi_J P_J$, we can obtain the first row of P_J as a function of $p_{m+1,m|J}$ and $p_{m|J}$. Finally, the elements of P_J are mapped to (ϱ_m, α_m) as in Section 6.

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.

Andrews, D. W. K. and Ploberger, W. (1994), "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

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.

Chesher, A. (1984), "Testing for Neglected Heterogeneity." *Econometrica*, 52, 865–872.

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.

— (2017), "Asymptotic Properties of the Maximum Likelihood Estimator in Regime Switching Econometric Models," Preprint, University of British Columbia.

Lee, L.-F. and Chesher, A. (1986), "Specification Testing when Score Test statistics are Identically Zero." *Journal of Econometrics*, 31, 121–149.

Lehmann, E. L. and Romano, J. P. (2005), *Testing Statistical Hypotheses*, Springer, third edition ed.

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 Zhuo, F. (2017), "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.

Table 1: Rejection frequencies under the null hypothesis

		$H_0 : M = 1$		$H_0 : M = 2$	
		(1)	(2)	(3)	(4)
	Test Statistic	Model 1	Model 2	Model 1	Model 2
$n = 200$	LRT	5.83	6.50	4.50	3.73
	supTS	5.10	4.67	—	—
	QLRT	4.97	—	—	—
$n = 500$	LRT	4.37	5.33	4.00	3.80
	supTS	4.57	5.07	—	—
	QLRT	4.43	—	—	—

Notes: Nominal level of 5%. We use 199 bootstrap samples and 3000 replications for $H_0 : M = 1$, and 1000 replications for $H_0 : M = 2$. For testing $H_0 : M = 2$ using Models 1 and 2, we generate the data under $(\beta, \mu_1, \mu_2, \sigma, p_{11}, p_{22}) = (0.5, -1, 1, 1, 0.7, 0.7)$ and $(\beta, \mu_1, \mu_2, \sigma_1, \sigma_2, p_{11}, p_{22}) = (0.5, -1, 1, 0.9, 1.2, 0.7, 0.7)$, respectively.

Table 2: Rejection frequencies for testing $H_0 : M = 1$ under the alternative hypothesis

		Model 1			Model 2		
(p_{11}, p_{22})	Test Statistic	$\mu_1 = 0.20$	$\mu_1 = 0.6$	$\mu_1 = 1.0$	$\mu_1 = 0.20$	$\mu_1 = 0.6$	$\mu_1 = 1.0$
$(0.25, 0.25)$	LRT	4.87	46.90	99.63	9.63	55.90	99.60
	supTS	6.23	56.43	95.90	16.37	70.97	95.37
	QLRT	5.10	8.00	55.27	—	—	—
$(0.50, 0.50)$	LRT	3.80	7.03	67.87	8.77	22.30	75.13
	supTS	4.07	4.40	4.60	14.70	35.77	35.30
	QLRT	4.90	9.40	82.50	—	—	—
$(0.70, 0.70)$	LRT	4.10	10.23	91.07	9.13	27.37	92.10
	supTS	4.57	7.40	26.37	14.90	36.20	43.43
	QLRT	5.13	8.53	58.73	—	—	—
$(0.90, 0.90)$	LRT	5.33	46.87	99.97	10.23	58.37	99.97
	supTS	6.77	13.90	4.40	19.10	41.17	35.30
	QLRT	4.83	5.63	5.97	—	—	—

Notes: Nominal level of 5% and $n = 500$. We use 199 bootstrap samples and 3000 replications. We set $\mu_2 = -\mu_1$ for both models, $(\beta, \sigma) = (0.5, 1.0)$ for Model 1 and $(\beta, \sigma_1, \sigma_2) = (0.5, 1.1, 0.9)$ for Model 2.

Table 3: Rejection frequencies for testing $H_0 : M = 2$ under the alternative hypothesis

	Model 1		Model 2	
	(μ_1, μ_2, μ_3) $= (1, 0, -1)$	(μ_1, μ_2, μ_3) $= (2, 0, -2)$	(μ_1, μ_2, μ_3) $= (1, 0, -1)$	(μ_1, μ_2, μ_3) $= (2, 0, -2)$
$(p_{11}, p_{22}, p_{33}) = (0.5, 0.5, 0.5)$	5.7	31.2	7.2	39.5
$(p_{11}, p_{22}, p_{33}) = (0.7, 0.7, 0.7)$	7.5	91.3	8.1	98.1

Notes: Nominal level of 5% and $n = 500$. We set $B = 199$, 1000 replications. We set $(\beta, \sigma) = (0.5, 1.0)$ for Model 1 and $(\beta, \sigma_1, \sigma_2, \sigma_3) = (0.5, 0.6, 0.9, 1.2)$ for Model 2. For both Model 1 and 2, we set $p_{ij} = (1 - p_{ii})/2$ for $j \neq i$ so that, for example, $(p_{12}, p_{13}) = (0.15, 0.15)$ when $p_{11} = 0.7$.

Table 4: Estimated parameters: the U.S. GDP per capita growth, 1960Q1-2014Q4

	$M=2$		$M=3$		$M=4$	
Panel A: Model 1 with common variance						
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
μ_1	-0.634	0.200	-0.823	0.151	-2.348	0.649
μ_2	0.951	0.176	0.692	0.172	-0.330	0.179
μ_3	–	–	2.023	0.236	0.532	0.161
μ_4	–	–	–	–	2.025	0.184
σ	0.913	0.053	0.752	0.052	0.832	0.040
β	0.787	0.041	0.773	0.046	0.639	0.053
Panel B: Model 2 with switching variance						
	$M=2$		$M=3$		$M=4$	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
μ_1	0.377	0.121	-0.629	0.298	-0.693	0.287
μ_2	0.428	0.175	0.624	0.167	0.614	0.179
μ_3	–	–	1.838	0.301	1.454	0.223
μ_4	–	–	–	–	2.244	0.369
σ_1	0.634	0.058	1.085	0.163	1.008	0.176
σ_2	1.495	0.135	0.579	0.053	0.466	0.077
σ_3	–	–	0.867	0.140	0.384	0.070
σ_4	–	–	–	–	0.874	0.156
β	0.865	0.035	0.780	0.047	0.687	0.051

Table 5: Selection of the number of regimes: the U.S. GDP per capita growth, 1960Q1-2014Q4

	Model 1 with common variance					Model 2 with switching variance				
				LRT					LRT	
M_0	log-like.	AIC	BIC	LR$_n$	p-val.	log-like.	AIC	BIC	LR$_n$	p-val.
1	-331.70	669.39	**679.58**	20.86	0.000	-331.70	669.39	679.58	47.41	0.000
2	-321.27	656.54	680.29	27.77	0.000	-307.99	631.99	**659.14**	22.29	0.010
3	-307.39	**640.77**	684.89	15.23	0.020	-296.85	**623.70**	674.61	**11.528**	**0.477**
4	-299.77	641.54	712.81	**6.57**	**0.523**	-291.09	630.17	711.62	15.02	0.296
5	-296.49	654.97	760.17	–	–	-283.58	637.15	755.93	–	–

Figure 1: The posterior probabilities of each regime (Model 1 with common variance): the U.S. GDP per capita growth, 1960Q1–2014Q4

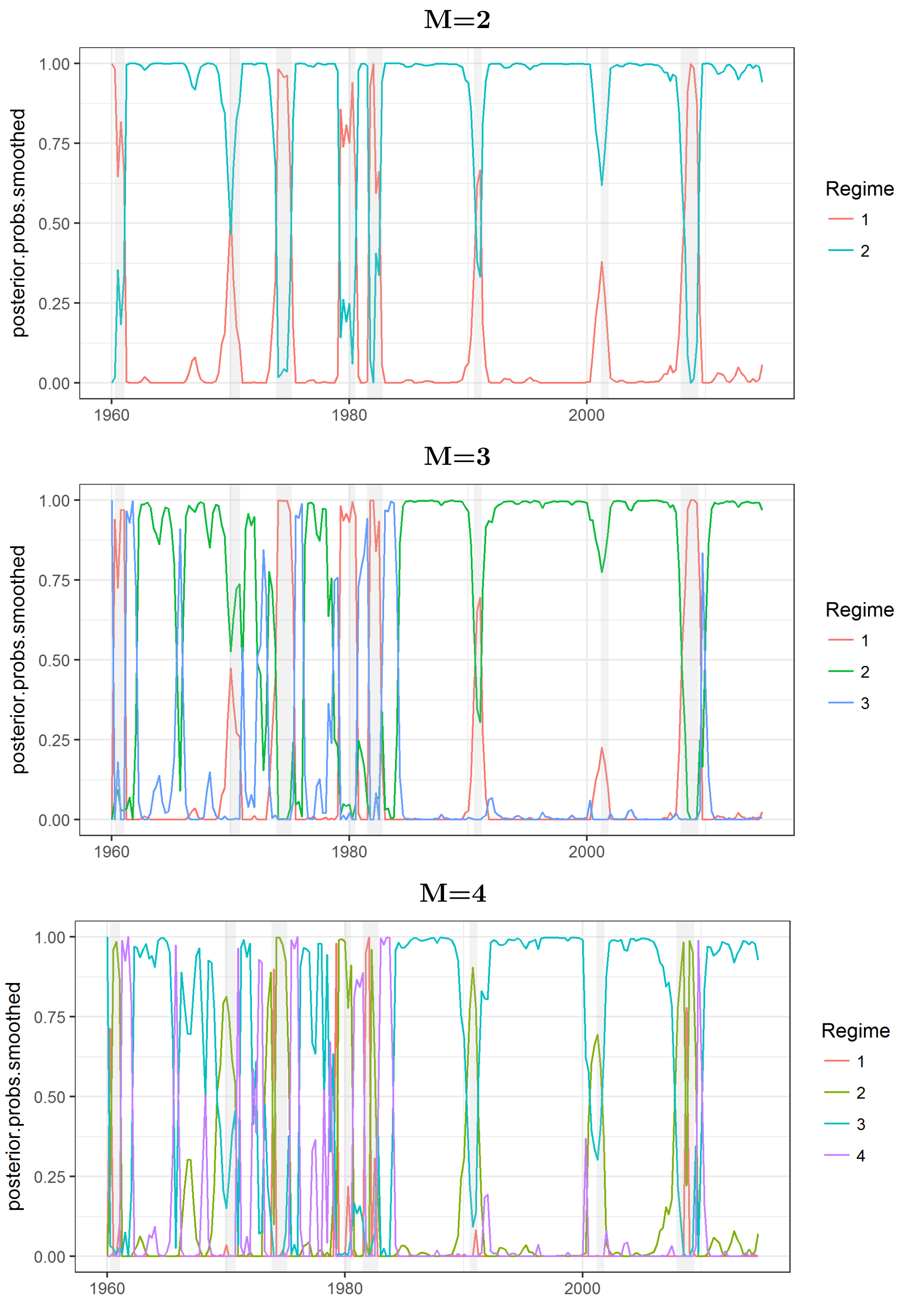

Figure 2: The posterior probabilities of each regime (Model 2 with switching variance): the U.S. GDP per capita growth, 1960Q1–2014Q4