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

April 18, 2017

Abstract

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

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

1 Introduction

Since Hamilton (1989)'s seminal contribution, Markov regime switching models have been widely used in numerous empirical applications in economics and finance because it can capture many important features in time series, such as structural changes, nonlinearity, high persistence, fat tails, leptokurtosis, and asymmetric dependence (see, e.g., Evans and Wachtel, 1993; Hamilton and Susmel, 1994; Gray, 1996; Sims and Zha, 2006; Inoue and Okimoto, 2008; Ang and Bekaert, 2002; Okimoto, 2008; Dai et al., 2007).

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

*This research is supported by the Natural Science and Engineering Research Council of Canada and JSPS Grant-in-Aid for Scientific Research (C) No. 26380267.
true parameter being on the boundary of the parameter space, and the degeneracy of Fisher information matrix. Analyzing the asymptotic properties of LRTS for testing the number of regimes for Markov regime switching models with normal component density, which is popular in empirical applications, is especially difficult because the normal density has an undesirable mathematical property that the second-order derivative with respect to the mean parameter is linearly dependent of the first derivative with respect to the variance parameter, leading to a further singularity.

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

Cho and White (2007) derive the asymptotic distribution of the quasi-likelihood ratio test statistic (Q-LRTS) for testing single regime against two regimes in the model with scalar switching parameter by rewriting the model as a two-component mixture models, thereby ignoring the temporal dependence of the regimes. Qu and Fan (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 while allowing for multiple switching parameters under some restriction on transition probabilities of latent regimes. Both Cho and White (2007) and Qu and Fan (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 more than two regimes remains unknown. Dannemann and Holtzmann (2008) analyze the Q-LRTS for testing the null of two regimes against three.

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

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

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

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

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

The remainder of this paper is organized as follows. After introducing notation and assumptions in section 2, we discuss the degeneracy of Fisher information matrix and the loss of identifiability in regime switching model in section 3. Section 4 establishes the DQM-type expansion. Section 5 presents the uniform convergence for the derivatives of density-ratios. Sections 6 and 7 derives the asymptotic distribution of the LRTS under $H_0$. Section 8 derives the asymptotic distribution
under local alternatives. Section 9 establishes the consistency of parametric bootstrap. Section 8 collects the proofs and the auxiliary results.

2 Notation and assumptions

Let := denote “equals by definition.” Let ⇒ denote weak convergence of a sequence of stochastic processes indexed by \( \pi \) for some space \( \Pi \). For a matrix \( B \), let \( \lambda_{\text{min}}(B) \) and \( \lambda_{\text{max}}(B) \) 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 \( |x| := \sqrt{x^T x} \) and \( |B| := \sqrt{\lambda_{\text{max}}(B^T B)} \). For \( k \times 1 \) vector \( a = (a_1, \ldots, a_k) \) and a function \( f(a) \), let \( \nabla^a f(a) \) denote a collection of derivatives of the form \( (\partial^j / \partial a_{i_1} \partial a_{i_2} \ldots \partial a_{i_j}) f(a) \).

The notation \(| \cdot | \) is used for the \( L^2 \) norm. Let \( \{ A \} \) denote an indicator function that takes value 1 when \( A \) is true and 0 otherwise. \( C \) denotes a generic nonnegative finite constant whose value may change from one expression to another. Let \([x] \) 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 \( \nu_n(f_k) := n^{-1/2} \sum_{k=1}^n [f_k - \mathbb{E} f^*(f_k)] \).

The proof of all the propositions and lemmas is presented in the appendix.

Consider the Markov regime switching process defined by a discrete-time stochastic process \( \{(X_k, Y_k, W_k)\} \), where \( (X_k, Y_k, W_k) \) takes values in a set \( \mathcal{X}_M \times \mathcal{Y} \times \mathcal{W} \) with the associated Borel \( \sigma \)-field \( \mathcal{B}(\mathcal{X}_M \times \mathcal{Y} \times \mathcal{W}) \). For a stochastic process \( \{ Z_k \} \) and \( a < b \), define \( Z^a_b := (Z_a, Z_{a+1}, \ldots, Z_b) \). Denote \( \overline{Y}_{k-1} := (Y_{k-1}, \ldots, Y_{k-s}) \) for a fixed integer \( s \) and \( \overline{Y}_a := (\overline{Y}_a, \overline{Y}_{a+1}, \ldots, \overline{Y}_b) \).

**Assumption 1.** (a) \( \{ X_k \}_{k=0}^\infty \) is a first-order Markov chain with the state space \( \mathcal{X}_M := \{1, 2, \ldots, M\} \). (b) For each \( k \geq 1 \), \( X_k \) is independent of \( (X_0, Y_0, W_0) \) given \( X_{k-1} \). (c) For each \( k \geq 1 \), \( Y_k \in \mathcal{Y} \subset \mathbb{R}^q_Y \) is conditionally independent of \( (X_0^{k-1}, Y_0^{k-1}, W_0^{k-1}, W_0^{\infty}) \) given \( (Y_{k-1}, W_k, X_k) \). (d) \( W_1^{\infty} \) is conditionally independent of \( (Y_0, X_0) \) given \( W_0 \). (e) \( \{(X_k, Y_k, W_k)\}_{k=0}^\infty \) is a strictly stationary ergodic process. [To Katsumi: (d1) and (d2) are equivalent under (a)-(c). See DRM_mem09.pdf.]

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

\[
p_{\vartheta}(y_k, x_k | \overline{Y}_{k-1}, w_k, x_{k-1}) := q_{\vartheta_{x}}(x_{k-1}, x_k) g_{\vartheta_{y}}(y_k | \overline{Y}_{k-1}, w_k, x_k) = q_{\vartheta_{x}}(x_{k-1}, x_k) \sum_{j \in \mathcal{X}_M} \mathbb{I}\{x_k = j\} f(y_k | \overline{Y}_{k-1}, w_k; \gamma, \theta_j).
\]
The parameter \( \vartheta_M \) belongs to \( \Theta_M = \Theta_{M,x} \times \Theta_{M,y} \), a compact subset of \( \mathbb{R}^{qM} \), and the true parameter value is denoted by \( \vartheta_M^* \).

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

**Assumption 2.** (a) \( 0 < \sigma_- := \inf_{\vartheta_M \in \Theta_M} \min_{x,x' \in \mathcal{X}_M} \vartheta_M(x, x') \) and \( \sigma_+ := \sup_{\vartheta_M \in \Theta_M} \max_{x,x' \in \mathcal{X}_M} \vartheta_M(x, x') < 1 \) for each \( M \). (b) For all \( y' \in \mathcal{Y} \), \( \overline{y} \in \mathcal{Y}^s \), and \( w \in \mathcal{W} \), \( 0 < \inf_{\vartheta_M \in \Theta_M} \sum_{x \in \mathcal{X}_M} \vartheta_M(y'|\overline{y}, w, x) \) and \( \sup_{\vartheta_M \in \Theta_M} \sum_{x \in \mathcal{X}_M} \vartheta_M(y|x, y, w) < \infty \). (c) \( b_+ := \sup_{\vartheta_M \in \Theta_M} \sup_{y_0, y_1, w, x} \vartheta_M(y_1|y_0, w, x) < \infty \) and \( E_{\vartheta^*}(\| \log b_-(Y_0, W, Y_1) \|) < \infty \), where \( b_-(y_0, w, y_1) := \inf_{\vartheta_M \in \Theta_M} \sum_{x \in \mathcal{X}_M} \vartheta_M(y_1|y_0, w, x) \).

As discussed in p. 2260 of DMR, Assumption 2(a) implies that the Markov chain \( \{X_k\} \) has a unique invariant distribution and uniformly ergodic for all \( \vartheta_M \in \Theta_M \). For notational brevity, we drop the subscript \( M \) from \( X_M, \vartheta_M \), etc., unless it is important to clarify the specific value of \( M \). Assumption 1(b)(c) imply that \( \{Z_k\}_{k=0}^\infty := \{(X_k, Y_k)\}_{k=0}^\infty \) is a Markov chain on \( \mathcal{Z} := \mathcal{X} \times \mathcal{Y}^s \) given \( \{W_k\}_{k=0}^\infty \), and \( Z_k \) is conditionally independent of \( \{Z_k, W_k\}_{k=0}^{k-1}, W_{k+1}^\infty \) given \( \{Z_k, W_k\}_{k=-\infty}^0 \). Consequently, Lemma 1 of the associated expectation of \( \{(Z_k, W_k)\}_{k=0}^\infty \) under stationarity by \( \mathbb{P}_\vartheta \) and \( E_{\vartheta} \), respectively.

Under Assumption 1(a)-(d), the density function of \( Y_1^n \) given \( X_0 = x_0, \overline{Y}_0 \) and \( W_1^n \) for the model with \( M \) regimes is given by

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

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

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

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

Note that

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

and \( \mathbb{P}_{\vartheta}(x_{k-1}|\overline{Y}_0^{k-1}, W_1^n) = \sum_{x_0 \in \mathcal{X}} \mathbb{P}_{\vartheta}(x_{k-1}|\overline{Y}_0^{k-1}, W_1^n, x_0) \mathbb{P}_{\vartheta}(x_0|\overline{Y}_0^{k-1}, W_1^n) \). Let \( \rho := 1 - \sigma_-/\sigma_+ \in (0, 1) \).

---

1DMR use \( \mathbb{P}_{\vartheta} \) and \( E_{\vartheta} \) to denote probability and expectation under stationarity on \( \{Z_k\}_{k=0}^{\infty} \), because their Section 7 deals with the case when \( Z_0 \) is drawn from an arbitrary distribution. Because we assume \( \{(Z_k, W_k)\}_{k=0}^{\infty} \) is stationary throughout this paper, we use notations such as \( \mathbb{P}_{\vartheta} \) and \( E_{\vartheta} \) without an overline for simplicity.
(0, 1). Lemma 10 in the appendix implies that, for all probability measures \( \mu_1 \) and \( \mu_2 \) on \( \mathcal{B}(\mathcal{X}) \) and all \( (Y_0^{k-1}, W_1^n) \),

\[
\sup_A \left| \sum_{x_0 \in \mathcal{X}} \mathbb{P}_\theta(X_{k-1} \in A|Y_0^{k-1}, W_1^n, x_0)\mu_1(x_0) - \sum_{x_0 \in \mathcal{X}} \mathbb{P}_\theta(X_{k-1} \in A|Y_0^{k-1}, W_1^n, x_0)\mu_2(x_0) \right| \leq \rho^{k-1}.
\]

Consequently, \( p_{\theta}(Y_k|Y_0^{k-1}, W_1^n, x_0) - p_{\theta}(Y_k|Y_0^{k-1}, W_1^n) \) goes to zero at an exponential rate as \( k \to \infty \). Therefore, as shown in the following proposition, the difference between \( \ell_n(\vartheta, x_0) \) and \( \ell_n(\vartheta) \) 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 DRM, the stationary density \( p_{\theta}(Y_k|Y_0^{k-1}, W_1^n) \) is not available in closed form for some models with autoregression. For this reason, we consider the log-likelihood function when the initial distribution of \( X_0 \) follows some arbitrary distribution \( \xi_M \in \Xi_M := \{ \xi(x_0)_{x_0 \in \mathcal{X}^M} : \xi(x_0) \geq 0 \text{ and } \sum_{x_0 \in \mathcal{X}^M} \xi(x_0) = 1 \} \).

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

\[
\ell_n(\vartheta_M, \xi_M) := \log \left( \sum_{x_0=1}^{M} p_{\vartheta_M}(Y_1^n|Y_0, W_1^n, x_0)\xi_M(x_0) \right),
\]

where \( p_{\vartheta_M}(Y_1^n|Y_0, W_1^n, x_0) \) is given in (1). We define the number of regimes by the smallest number \( M \) such that the data density admits the representation (3). Our objective is to test

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

Define the likelihood ratio test statistic (LRTS, hereafter) for testing \( H_0 \) as

\[
2[\max_{\vartheta_{M_0+1} \in \Theta_{M_0+1}} \ell_n(\vartheta_{M_0+1}, \xi_{M_0+1}) - \max_{\vartheta_{M_0} \in \Theta_{M_0}} \ell_n(\vartheta_{M_0}, \xi_{M_0})].
\]

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

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

\(^2\)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
because $Y_k$ has the same distribution across regimes. Furthermore, Section 6 shows that, when $\theta_1 = \theta_2$, the scores with respect to $\theta_1$ and $\theta_2$ are linearly dependent so that the Fisher information matrix is degenerate.

The log-likelihood function of Markov switching models with normal density has further degeneracy. For example, in a two-regime model where $Y_k$ in the $j$-th regime follows $N(\mu_j, \sigma_j^2)$, the model reduces to a heteroskedastic normal mixture model when $p_1(X_k = 1|X_{k-1} = 1) = p_2(X_k = 1|X_{k-1} = 2)$, i.e., $p_{11} = 1 - p_{22}$. Kasahara and Shimotsu (2015) show that, in a heteroskedastic normal mixture model, the first and second derivatives of the log-likelihood function are linearly dependent and the score function is a function of the fourth-order derivative. Consequently, one needs to expand the log-likelihood function four times to derive the score function.

4 Quadratic expansion under loss of identifiability

When testing the number of regimes by the LRT, a part of $\vartheta$ is not identified under the null hypothesis. Let $\pi$ denote the part of $\vartheta$ that is not identified under the null, split $\vartheta$ 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 $p_\vartheta$ and $p_{\psi\pi}$ interchangeably.

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

$$N_\varepsilon := \{\vartheta \in \Theta : |t_\vartheta| < \varepsilon\}.$$  

When the MLE is consistent, the asymptotic distribution of the LRTS is determined by the local properties of the likelihood functions in $N_\varepsilon$.

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

Denote the conditional density-ratio by

$$l_{\vartheta_{k,x_0}} := \frac{p_{\psi\pi}(Y_k|Y_{0}^{k-1}, W_n^{1}, x_0)}{p_{\psi^*\pi}(Y_k|Y_{0}^{k-1}, W_n^{1}, x_0)},$$

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

zero (Gassiat and Keribin, 2000).
Assumption 3. For all $k = 1, \ldots, n$, $l_{\theta k x_0} - 1$ admits an expansion
\begin{equation}
  l_{\theta k x_0} - 1 = t_0 s_{\pi k} + r_{\theta k} + u_{\theta k x_0},
\end{equation}

where $(s_{\pi k}, r_{\theta k}, u_{\theta k x_0})$ satisfy, for some $C \in (0, \infty)$, $\varepsilon > 0$, and $p \in (0, 1)$, (a) $\mathbb{E}_{\theta} \sup_{\pi \in \Theta_{\varepsilon}} |s_{\pi k}| < C$, (b) $\sup_{\pi \in \Theta_{\varepsilon}} |P_n(s_{\pi k} s'_{\pi k}) - I_\pi| = o_p(1)$, where $\sup_{\pi \in \Theta_{\varepsilon}} \lambda_{\max}(I_\pi) < \infty$, (c) $\mathbb{E}_{\theta} \sup_{\theta \in \mathcal{N}_\epsilon} |r_{\theta k}/(|t_\theta||\psi - \psi^*)|^2 < \infty$, (d) $\sup_{\theta \in \mathcal{N}_\epsilon} [\sup_{\psi \in \mathcal{X}_0} (1 - \sup_{\psi \in \mathcal{X}_0} |\psi|)|\psi - \psi^*|] = O_p(1)$, (e) $\max_{x_0 \in \chi} \mathbb{E}_{\theta} \sup_{\theta \in \mathcal{N}_\epsilon} (|u_{\theta k x_0}|/|\psi - \psi^*|)^2 \leq C' k^{-1}$, (f) $0 < \inf_{\pi \in \Theta_{\varepsilon}} \lambda_{\min}(I_\pi)$, (g) there exists a stochastic process $Z_n(\theta)$ such that $\sup_{\theta \in \mathcal{N}_\epsilon} |\nu_n(s_{\pi k}) - Z_n(\theta)| = o_p(1)$ and $\sup_{\theta \in \mathcal{N}_\epsilon} |Z_n(\theta)| = O_p(1)$.

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

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

**Proposition 2.** Suppose that Assumption 3(a)-(e) holds. Then, for any $x_0 \in \mathcal{X}$ and for all $c > 0$,
\begin{equation}
  \sup_{\vartheta \in \mathcal{N}_\epsilon/\sqrt{n}} \left| \ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) - \sqrt{nt_\theta} \nu_n(s_{\pi k}) + nt_\theta I_\pi t_\theta/2 \right| = o_p(1).
\end{equation}

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

**Proposition 3.** Suppose that Assumption 3 holds. Then, for any $x_0 \in \mathcal{X}$, (a) $\sup_{\theta \in A_n(x_0)} |t_\theta| = O_p(n^{-1/2})$; (b)
\begin{equation}
  \sup_{\vartheta \in A_n(x_0)} \left| \ell_n(\psi, \pi^*, x_0) - \ell_n(\psi, \pi, x_0) - \sqrt{nt_\theta} Z_n(\theta) + nt_\theta I_\pi t_\theta/2 \right| = o_p(1).
\end{equation}

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

**Corollary 1.** (a) Under the assumptions of Proposition 2, we have
\begin{equation}
  \sup_{\xi \in \mathcal{N}_\epsilon} \sup_{\vartheta \in \mathcal{N}_\epsilon/\sqrt{n}} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{nt_\theta} \nu_n(s_{\pi k}) + nt_\theta I_\pi t_\theta/2 \right| = o_p(1) \text{ for all } c > 0. \quad (b)
\end{equation}

Under the assumptions of Proposition 3, we have $\sup_{\xi \in \mathcal{N}_\epsilon} \sup_{\vartheta \in A_n} |t_\theta| = O_p(n^{-1/2})$ and $\sup_{\xi \in \mathcal{N}_\epsilon} \sup_{\vartheta \in A_n} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{nt_\theta} Z_n(\theta) + nt_\theta I_\pi t_\theta/2 \right| = o_p(1)$.
5 Uniform convergence of the derivatives of the log-density and the density-ratios

In this section, we establish approximations that enable us to apply Propositions 2 and 3 and Corollary 1 to the log-likelihood function of regime switching models. Because of the presence of singularity, the expansion (5) involves higher-order derivatives of the density-ratios \( \nabla^j \psi_l \theta x_0 \) with \( j \geq 2 \). First, we express \( \nabla^j \psi_l \theta x_0 \) 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 \( \{ \nabla^j \psi_l \theta x_0 \}_{k=0}^{\infty} \) can be approximated by a stationary martingale difference sequence by conditioning on the infinite past \( Y_{k-1}^{-\infty} \) in place of \( Y_{k-1}^{0} \). The leading term satisfies the assumptions on \( s_{\pi k} \) in (5) because it is a stationary martingale difference sequence, and the resulting approximation error is sufficiently small to satisfy the assumptions on the remainder terms \( r_{\theta k} \) and \( u_{\theta k x_0} \).

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

\[
\phi^i(\theta, Z_{k-1}) := \nabla^i \log p_{\theta}(Y_k, X_k | Y_{k-1}, W_k, X_{k-1}), \quad i \geq 1.
\]

(6)

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

\[
E_\theta^c [V_1, \ldots, V_q | F] := E_\theta [(V_1 - E_\theta[V_1 | F]) \cdots (V_q - E_\theta[V_q | F]) | F],
\]

so that \( E_\theta^c [\phi_{\theta k_1} \phi_{\theta k_2} \phi_{\theta k_3} | F] := E_\theta [(\phi_{\theta k_1} - E_\theta[\phi_{\theta k_1} | F]) (\phi_{\theta k_2} - E_\theta[\phi_{\theta k_2} | F]) (\phi_{\theta k_3} - E_\theta[\phi_{\theta k_3} | F]) | F] \).

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

$$
\Phi_{\theta \mathcal{T}(1)}(\mathcal{F}) := \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \middle| \mathcal{F} \right], \quad \Phi_{\theta \mathcal{T}(2)}(\mathcal{F}) := \frac{1}{|\mathcal{I}(2)|} \sum_{(t_1, t_2) \in \sigma(\mathcal{T}(2))} \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_2}^2 \middle| \mathcal{F} \right],
$$

$$
\Phi_{\theta \mathcal{T}(3)}(\mathcal{F}) := \frac{1}{|\mathcal{I}(3)|} \sum_{(t_1, t_2, t_3) \in \sigma(\mathcal{T}(3))} \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_2}^2 \phi_{\theta t_3}^3 \middle| \mathcal{F} \right],
$$

$$
\Phi_{\theta \mathcal{T}(4)}(\mathcal{F}) := \frac{1}{|\mathcal{I}(4)|} \sum_{(t_1, t_2, t_3, t_4) \in \sigma(\mathcal{T}(4))} \left( \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_2}^2 \phi_{\theta t_3}^3 \phi_{\theta t_4}^4 \middle| \mathcal{F} \right] - \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_2}^2 \phi_{\theta t_3}^3 \middle| \mathcal{F} \right] \mathbb{E}_{\theta} \left[ \phi_{\theta t_3}^3 \phi_{\theta t_2}^4 \middle| \mathcal{F} \right] - \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_3}^3 \phi_{\theta t_4}^4 \middle| \mathcal{F} \right] \mathbb{E}_{\theta} \left[ \phi_{\theta t_2}^2 \phi_{\theta t_4}^4 \middle| \mathcal{F} \right] \mathbb{E}_{\theta} \left[ \phi_{\theta t_2}^2 \phi_{\theta t_3}^3 \phi_{\theta t_4}^4 \middle| \mathcal{F} \right] \mathbb{E}_{\theta} \left[ \phi_{\theta t_1}^1 \phi_{\theta t_2}^2 \phi_{\theta t_3}^3 \phi_{\theta t_4}^4 \middle| \mathcal{F} \right] \right),
$$

and Section 10.1 in appendix defines $\Phi_{\theta \mathcal{T}(5)}(\mathcal{F})$ and $\Phi_{\theta \mathcal{T}(6)}(\mathcal{F})$. Note that these moments are symmetric with respect to $(t_1, \ldots, t_j)$. Define, for $j = 1, 2, \ldots, 6$, $k \geq 1$, $m \geq 0$ and $x \in \mathcal{X}$,

$$
\Delta_{\mathcal{T}(j) j, k, m, x}(\theta) := \sum_{T(j) \in \{-m+1, \ldots, k\}^j} \Phi_{\theta \mathcal{T}(j)} \left[ \tilde{X}^k_{-m}, W^n_{m+1}, X_m = x \right] - \sum_{T(j) \in \{-m+1, \ldots, -k\}^j} \Phi_{\theta \mathcal{T}(j)} \left[ \tilde{X}^{k-1}_{-m}, W^n_{m+1}, X_m = x \right],
$$

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

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

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

de note the stationary density of $Y^k_{-m}$ conditional on $\{\tilde{Y}_{-m}, W^n_{m+1}\}$ where $X_m$ is drawn from its true conditional stationary distribution $\overline{p}_{\theta^*}(X_m | \tilde{Y}^{k-1}_{-m}, W^n_{m+1})$. Let $\overline{p}_{\theta}(Y^k_{-m}, W^n_{m+1}) := \overline{p}_{\theta}(Y^k_{-m} | \tilde{Y}_{-m}, W^n_{m+1}) / \overline{p}_{\theta}(Y^{k-1}_{-m} | \tilde{Y}_{-m}, W^n_{m+1})$ denote the associated conditional density of $Y^k_{-m}$ given $\{\tilde{Y}^{k-1}_{-m}, W^n_{m+1}\}$.

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

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

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

\[3\text{Note that DRM use the same notation \overline{p}_{\theta}(Y^k_{-m}) for a different purpose. On p. 2263 and in some other (but not all) places, DRM use \overline{p}_{\theta}(Y_k | \tilde{Y}^{k-1}_{-m}) to denote an (ordinary) stationary conditional distribution of Y_k.}\]
The following proposition expresses the derivatives of log densities, \( \nabla^j \ell_{k,m,x}(\vartheta) \)'s, in terms of the conditional expectation of the central moments of derivatives of the complete data log-density. The first two equations are also given in DMR (p. 2272 and pp. 2276-7).

**Proposition 4.** For all \( 1 \leq k \leq n, \; m \geq 0, \; \text{and} \; x \in X \),

\[
\nabla^1 \ell_{k,m,x}(\vartheta) = \Delta^1_{1,k,m,x}(\vartheta), \quad \nabla^2 \ell_{k,m,x}(\vartheta) = \Delta^2_{1,k,m,x}(\vartheta) + \Delta^1_{2,k,m,x}(\vartheta), \\
\nabla^3 \ell_{k,m,x}(\vartheta) = \Delta^3_{1,k,m,x}(\vartheta) + 3\Delta^2_{2,k,m,x}(\vartheta) + \Delta^1_{3,k,m,x}(\vartheta), \\
\nabla^4 \ell_{k,m,x}(\vartheta) = \Delta^4_{1,k,m,x}(\vartheta) + 4\Delta^3_{2,k,m,x}(\vartheta) + 3\Delta^2_{3,k,m,x}(\vartheta) + 6\Delta^1_{4,k,m,x}(\vartheta) + \Delta^1_{11,k,m,x}(\vartheta), \\
\nabla^5 \ell_{k,m,x}(\vartheta) = \Delta^5_{1,k,m,x}(\vartheta) + 5\Delta^4_{2,k,m,x}(\vartheta) + 10\Delta^3_{2,k,m,x}(\vartheta) + 10\Delta^2_{3,k,m,x}(\vartheta) + 15\Delta^1_{4,k,m,x}(\vartheta) \\
\quad + 10\Delta^2_{5,k,m,x}(\vartheta) + \Delta^1_{11,11,11}(\vartheta), \\
\nabla^6 \ell_{k,m,x}(\vartheta) = \Delta^6_{1,k,m,x}(\vartheta) + 6\Delta^5_{2,k,m,x}(\vartheta) + 15\Delta^4_{2,k,m,x}(\vartheta) + 10\Delta^3_{3,k,m,x}(\vartheta) + 15\Delta^2_{4,k,m,x}(\vartheta) \\
\quad + 60\Delta^3_{3,k,m,x}(\vartheta) + 5\Delta^2_{5,k,m,x}(\vartheta) + 20\Delta^1_{4,k,m,x}(\vartheta) + 45\Delta^2_{2,k,m,x}(\vartheta) + 15\Delta^1_{4,5,k,m,x}(\vartheta) + \Delta^1_{11,11,11}(\vartheta).
\]

Further, the above holds when \( \nabla^j \ell_{k,m,x}(\vartheta) \) and \( \Delta^j_{j,k,m,x}(\vartheta) \) are replaced with \( \nabla^j \ell_{k,m,x}(\vartheta) \) and \( \Delta^j_{j,k,m,x}(\vartheta) \).

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

**Assumption 4.** There exists a positive real \( \delta \) such that on \( N^* := \{ \vartheta \in \Theta : |\vartheta - \vartheta^*| \leq \delta \} \) the following conditions hold: (a) For all \( (y, w, x) \in Y^s \times \mathcal{W} \times X \), the function \( \vartheta \mapsto g_{\vartheta}(y|y,w,x) \) is six times continuously differentiable on \( N^* \). (b) \( E_{\vartheta^*}[\sup_{\vartheta \in N^*} \sup_{x \in X} |\nabla^j_{\vartheta} \log g_{\vartheta}(y|Y_0, W, X)|^{2q}] < \infty \) for \( j = 1, \ldots, 6 \) and \( E_{\vartheta^*} \sup_{\vartheta \in N^*} |G_{\vartheta_k}/G_{\vartheta_k}|^{q_0} < \infty \) with \( q_0 = 6q_1, q_2 = 5q_0, \ldots, q_6 = q_0 \), where \( q_0 = (1 + \varepsilon) \max\{2, \dim(\vartheta)\} \) and \( q_9 = (1 + \varepsilon) \max\{2, \dim(\vartheta)\}/\varepsilon \) for some \( \varepsilon > 0 \). (c) For almost all \( (y, w, x) \in Y^s \times \mathcal{W} \times X \), there exists a function \( f_{y,w,x} : Y \to \mathbb{R}^+ \) such that \( \sup_{y' \in Y^s, w, x} g_{\vartheta_k}(y'|y, w, x) \leq f_{y,w,x}(x) < \infty \) and, for almost all \( (y, w, x) \in Y^s \times \mathcal{W} \times X \), for \( j = 1, \ldots, 6 \), there exist functions \( f_{y,w,x}^j : Y \to \mathbb{R}^+ \) in \( L^1 \) such that \( |\nabla^j_{\vartheta} g_{\vartheta}(y'|y, w, x)| \leq f_{y,w,x}^j(y') \) for all \( \vartheta \in N^* \).

Lemma 3 in the appendix shows that, for all \( x \in X \) and \( 1 \leq k \leq n \) and a suitably defined \( r_{I(j)} \), \( \{\Delta^j_{j,k,m,x}(\vartheta)\}_{m \geq 0} \) is a uniform Cauchy sequence in \( L^{r_{I(j)}}(\mathbb{P}_{\vartheta^*}) \) that converges uniformly with respect to \( \vartheta \in N^* \mathbb{P}_{\vartheta^*}\)-a.s. and in \( L^{r_{I(j)}}(\mathbb{P}_{\vartheta^*}) \) to a random variable that does not depend on \( x \). Then, in view of Proposition 4 and Lemma 3, \( \{\nabla^j \ell_{k,m,x}(\vartheta)\}_{m \geq 0} \) and \( \{\nabla^j \ell_{k,m}(\vartheta)\}_{m \geq 0} \) converge to \( \nabla^j \ell_{k,\infty}(\vartheta) \) uniformly with respect to \( \vartheta \in N^* \mathbb{P}_{\vartheta^*}\)-a.s. and in \( L^{r_{I}}(\mathbb{P}_{\vartheta^*}) \) as the following proposition shows. Define \( \rho := 1 - \sigma_-/\sigma_+ \).
Proposition 5. Under Assumptions 1, 2, and 4, for \( j = 1, \ldots, 6 \), there exist random variables \( K_j, M_j \in L^r(\mathbb{P}_{\theta^*}) \) such that, for all \( 1 \leq k \leq n \) and \( m' \geq m \geq 0 \),

\[
\begin{align*}
(a) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta) - \nabla^j \ell_{k,m}(\theta)| \leq K_j (k + m)^6 \rho^{(k+m-1)/12}, \quad \mathbb{P}_{\theta^*}\text{-a.s.}, \\
(b) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta) - \nabla^j l_{k,m',x}(\theta)| \leq K_j (k + m)^6 \rho^{(k+m-1)/53}, \quad \mathbb{P}_{\theta^*}\text{-a.s.}, \\
(c) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta)| + \sup_{\theta} |\nabla^j \ell_{k,m}(\theta)| + \sup_{\theta} |\nabla^j l_{k,\infty}(\theta)| \leq M_j, \quad \mathbb{P}_{\theta^*}\text{-a.s.},
\end{align*}
\]

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

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

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

\[
\begin{align*}
(a) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta) - \nabla^j \ell_{k,m}(\theta)| \leq K_j (k + m)^6 \rho_{*}^{k+m-1}, \quad \mathbb{P}_{\theta^*}\text{-a.s.}, \\
(b) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta) - \nabla^j l_{k,m',x}(\theta)| \leq K_j (k + m)^6 \rho_{*}^{k+m-1}, \quad \mathbb{P}_{\theta^*}\text{-a.s.}, \\
(c) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,0,x}(\theta) - \nabla^j l_{k,\infty}(\theta)| \leq K_j k^6 \rho_{*}^{k-1}, \quad \mathbb{P}_{\theta^*}\text{-a.s.}, \\
(d) & \quad \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{N}^*} |\nabla^j l_{k,m,x}(\theta)| + \sup_{\theta} |\nabla^j \ell_{k,m}(\theta)| + \sup_{\theta} |\nabla^j l_{k,\infty}(\theta)| \leq M_j, \quad \mathbb{P}_{\theta^*}\text{-a.s.},
\end{align*}
\]

When we apply Propositions 2 and 3 and Corollary 1 to regime switching models, \( l_{k,0,x}(\theta) \) corresponds to \( l_{\theta k x_0} \) on the left hand side of (5), and \( s_{\pi k} \) and \( r_{\theta k} \) in (5) are functions of \( \nabla^j l_{k,m}(\theta) \)'s. The term \( u_{\theta k x_0} \) in (5) that depends on \( x_0 \) satisfies Assumption 3(e) because of Proposition 6(a)(c).

Proposition 6 and dominated convergence theorem for conditional expectations (Durrett, 2010, Theorem 5.5.9) imply that \( \mathbb{E}_{\theta^*}[\nabla^j l_{k,\infty}(\theta)|Y^{k-1}] = 0 \) for all \( \theta \in \mathcal{N}^* \) and \( \{\nabla^j l_{k,\infty}(\theta)\}_{k=-\infty}^{\infty} \) is a stationary, ergodic and square integrable martingale difference sequence so that we can apply a martingale central limit theorem. \( \nabla^j l_{k,\infty}(\theta) \) for \( j = 1, \ldots, 5 \) satisfies Assumption 3(g) from Theorem 2 of Hansen (1996) because \( \nabla^j l_{k,\infty}(\theta) \) is Lipschitz continuous with Lipschitz coefficient \( \nabla^{j+1} l_{k,\infty}(\theta) \in L^{\max\{2,\dim(\theta)\}}(\mathbb{P}_{\theta^*}) \) from Proposition 6.

6 Testing homogeneity

Before developing the LRT of \( M_0 \) components, we analyze a simpler case of testing the null hypothesis \( H_0 : M = 1 \) against \( H_A : M = 2 \) when the data is from \( H_0 \). We assume that the parameter
space for $\vartheta_{2,x} = (p_{11},p_{22})'$ satisfies Assumption 2(a) with restriction $p_{11},p_{22} \in [\epsilon,1-\epsilon]$ for a small $\epsilon \in (0,1/2)$, and let $\Theta_{2x}$ denote $\Theta_2$ with this restriction. This assumption is necessary because the LRTS is unbounded under the null hypothesis when $p_{11}$ or $p_{22}$ tends to 1 (Gassiat and Keribin, 2000). Denote the true parameter in a one-regime model by $\vartheta^*_1 := ((\theta^*)',(\gamma^*)')'$. The two-regime model gives rise to the true density $p_{\vartheta^*_1}(Y^n_1|\bar{Y}_0,x_0)$ if the parameter $\vartheta_2 = (\theta_1,\theta_2,\gamma,p_{11},p_{22})'$ lies in a subset of the parameter space

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

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

Let $\ell_n(\vartheta_2,\xi_2) := \log \left( \sum_{x_0=1}^2 p_{\vartheta_2}(Y^n_1|\bar{Y}_0,W^n_1,x_0)\xi_2(x_0) \right)$ denote the two-regime log-likelihood for a given initial distribution $\xi_2(x_0) \in \Xi_2$, and let $\hat{\vartheta}_2 := \arg \max_{\vartheta_2 \in \Theta_{2x}} \ell_n(\vartheta_2,\xi_2)$ denote the maximum likelihood estimator (MLE) of $\vartheta_2$ given $\xi_2$. Because $\xi_2$ does not matter asymptotically, we treat $\xi_2$ fixed for simplicity and suppress $\xi_2$ from $\hat{\vartheta}_2$. Let $\hat{\vartheta}_1$ denote the one-regime MLE that maximizes the one-regime log-likelihood function $\ell_{0,n}(\vartheta_1) := \sum_{k=1}^n \log f(Y_k|\bar{Y}_{k-1},W_k;\gamma,\theta)$ 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 $p_{\vartheta^*_1}(Y^n_1|\bar{Y}_{-m},W^n_{-m})$ and $p_{\vartheta_2}(Y^n_1|\bar{Y}_{-m},W^n_{-m})$ is 0 if and only if $\vartheta_2 \in \Gamma^*$. This assumption is similar to Assumption (A5') in DMR.\(^4\)

**Assumption 5.** (a) $\Theta_1$ and $\Theta_2$ are compact. (b) For all $(x,x') \in X$ and all $(\pi,y',w) \in \mathcal{Y}^x \times \mathcal{Y} \times \mathcal{W}$, the function $(\theta,\gamma) \mapsto f(y'/\bar{y}_0,w;\gamma,\theta)$ is continuous. (c) If $\vartheta_1 \neq \vartheta^*_1$, then $\mathbb{P}_{\vartheta^*_1} \left( f(Y_1|\bar{Y}_0,W_1;\gamma,\theta) \neq f(Y_1|\bar{Y}_0,W_1;\gamma^*,\theta^*) \right) > 0$. (d) $\mathbb{E}_{\vartheta^*_1} \left[ \log p_{\vartheta_2}(Y_1|\bar{Y}_{-m},W^n_{-m}) \right] = \mathbb{E}_{\vartheta^*_1} \left[ \log p_{\vartheta^*_1}(Y_1|\bar{Y}_{-m},W^n_{-m}) \right]$ for all $m \geq 0$ if and only if $\vartheta_2 \in \Gamma^*$.

The following proposition shows that the MLE of $\vartheta^*_1$ 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 $\inf_{\vartheta_2 \in \Gamma^*} \| \hat{\vartheta}_2 - \vartheta_2 \|_p \xrightarrow{p} 0$.

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 $\vartheta_2$ as $\vartheta_2 = (\psi,\pi)$, where $\pi$ is the part of $\vartheta$ that is not identified under the null hypothesis, and the elements of $\psi$ will be delineated later. In the current setting, $\vartheta_{2,x} = (p_{11},p_{22})'$ is not identified under the null. Define $\rho := \text{corr}_{\vartheta_{2,x}}(X_k,X_{k+1}) = p_{11} + p_{22} - 1$ and $\alpha := \mathbb{P}_{\vartheta_{2,x}}(X_k = 1) = (1 - p_{22})/(2 - p_{11} - p_{22})$ (see Lemma 13). The parameter spaces for $\rho$ and $\alpha$ under restriction $p_{11},p_{22} \in [\epsilon,1-\epsilon]$ are given by $\Theta_{\text{red}} := [0,1] \cup [-\rho,0]$ and $\Theta_{\alpha} := [\alpha,1-\alpha]$, respectively. Because the mapping from $(p_{11},p_{22})$ to $(\rho,\alpha)$ is one-to-one, we

\(^4\)DMR derive their Assumption (A5') from their "minimal assumption" (Assumption (A5)). We do not repeat DMR here because extending their proof (in particular, the proof of their Lemmas 5 and 6) to models with $W_k$ is not very straightforward.
reparameterize \( \pi \) as \( \pi := (\varrho, \alpha) \in \Theta_{\pi \kappa} := \Theta_{\varrho \kappa} \times \Theta_{\alpha \kappa} \), and let \( p_{\psi \pi}(\cdot | \cdot) := p_{\theta 2}(\cdot | \cdot) \). Henceforth, we suppress \( W_1^n \) for notational brevity and write, for example, \( p_{\psi \pi}(Y_1^n | Y_0, W_1^n, x_0) \) as \( p_{\psi \pi}(Y_1^n | Y_0, x_0) \) and \( p_{\psi \pi}(y, x, k | \bar{y}_{k-1}, w, x_{k-1}) \) as \( p_{\psi \pi}(y, x, k | \bar{y}_{k-1}, x_{k-1}) \) unless confusion might arise. We apply Corollary 1 to \( \ell_n(\psi, \pi, \xi_2) \) by finding a representation of \( (t_{\vartheta}, s_{\pi \kappa}, r_{\vartheta k}, u_{\vartheta k x_0}) \) in (5) in terms of \( \vartheta \), \( p_{\psi \pi}(\cdot | \cdot) \), and derivatives of \( p_{\psi \pi}(\cdot | \cdot) \) and then showing that \( (t_{\vartheta}, s_{\pi \kappa}, r_{\vartheta k}, u_{\vartheta k x_0}) \) satisfy Assumption 3. Because of the degeneracy of Fisher information matrix, \( s_{\pi \kappa} \) involves higher-order derivatives, and \( t_{\vartheta} \) consists of functions of polynomials of (reparameterized) \( \vartheta \).

The remainder of this section derives \( s_{\pi \kappa} \) as a function of \( \nabla^j\bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}) / \bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}) \) with \( \bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}) \) defined in (9). This approximation is valid because Proposition 6 implies that \( \nabla^j\bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}, x_0) / \bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}, x_0) - \nabla^j\bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}) / \bar{p}_{\psi \pi}(Y_k | \bar{Y}_0^{k-1}) \) 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 \( \bar{Y}_{-\infty}^\infty \) and \( X_{-\infty}^\infty \) are independent when \( \psi = \psi^* \),

\[
\mathbb{P}_{\psi^* \pi}(X_{-\infty}^\infty | Y_{-\infty}^\infty) = \mathbb{P}_{\psi^* \pi}(X_{-\infty}^\infty).
\]

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

Consider the following reparameterization. Let

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

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

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

$$
\nabla \gamma g_{\psi^*}(y_k|x_k) = \nabla f(y_k; \gamma^*, \theta^*), \quad \nabla \nu g_{\psi^*}(y_k|x_k) = \nabla \theta f(y_k; \gamma^*, \theta^*),
$$

$$
\nabla \lambda g_{\psi^*}(y_k|x_k) = (q_k - \alpha) \nabla \theta f(y_k; \gamma^*, \theta^*),
$$

$$
\nabla \lambda \nu g_{\psi^*}(y_k|x_k) = (q_k - \alpha) \nabla \theta (\gamma', \theta') f(y_k; \gamma^*, \theta^*),
$$

$$
\nabla \lambda \lambda^2 g_{\psi^*}(y_k|x_k) = (q_k - \alpha)^2 \nabla \theta \theta f(y_k; \gamma^*, \theta^*).
$$

Similarly, it follows from Lemma 1, (13), (14), and (15), and the definition of $\nabla \lambda \lambda^2 g_{\psi^*}(y_k|x_k)$, and the definition of $\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})$ in (9), we obtain

$$
\frac{\nabla \psi \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})}{\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})} = \nabla \psi \log \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1}) = \nabla \psi \log g_{\psi^*}(y_k|\overline{y}_0^{k-1}) = \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \psi \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right] - \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \psi \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right].
$$

Applying (13), (14), and (15) to the right hand side gives

$$
\nabla \eta \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})/\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1}) = \nabla_{(\gamma', \theta')} \log f_{\psi^*}^*,
$$

$$
\nabla \lambda \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})/\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1}) = 0.
$$

Similarly, it follows from Lemma 1, (13), (14), and (15), and $g_{\psi^*}^* = f_{\psi^*}^*$ that

$$
\nabla \lambda \nu g_{\psi^*}(y_k|\overline{y}_0^{k-1})/\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1}) = \nabla \lambda \nu g_{\psi^*}(y_k|\overline{y}_0^{k-1}) = 0,
$$

$$
\nabla \lambda \lambda^2 g_{\psi^*}(y_k|\overline{y}_0^{k-1})/\overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})
$$

$$
= \nabla \lambda \lambda^2 \log \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})
$$

$$
= \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \lambda \lambda^2 \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right] - \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \lambda \lambda^2 \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right]
$$

$$
+ \sum_{t=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \lambda \lambda^2 \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right] - \sum_{t=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla \lambda \lambda^2 \log g_{\nu_t^*} \left| \overline{y}_0^{k-1} \right. \right]
$$

$$
= \alpha(1 - \alpha) \left[ \frac{\nabla \theta \theta f_{\psi^*}^*}{f_{\psi^*}^*} + \sum_{t=1}^{k} \theta^{k-t} \left( \frac{\nabla \theta f_{\psi^*}^*}{f_{\psi^*}^*} + \frac{\nabla \theta \theta f_{\psi^*}^*}{f_{\psi^*}^*} \right) \right].
$$

Note that $\varrho$ is bounded away from $-1$ and $1$ in $\Theta_{\varrho^e}$. Because the first-order derivative with respect to $\lambda$ is identically equal to zero in (15), the information on $\lambda$ is provided by the second-order derivative with respect to $\lambda$ in (17). Consequently, the unique elements of $\nabla \eta \log \overline{p}_{\psi^*}(y_k|\overline{y}_0^{k-1})$ and
Proposition 8. Suppose Assumptions 1, 2, 4, 5 and 6 hold. Then, under the null hypothesis of Assumption 6. Here, \( \nabla \) and \( s \) in Corollary 1. Because this score is approximated by a stationary martingale difference sequence and the remainder term satisfies Assumption 4 from Lemmas 6, we can apply Corollary 1 to the likelihood ratio to derive the asymptotic distribution of the LRTS.

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

\[
v(\lambda) := (\lambda_1^2, \ldots, \lambda_q^2, \lambda_1 \lambda_2, \ldots, \lambda_1 \lambda_q, \lambda_2 \lambda_3, \ldots, \lambda_2 \lambda_q, \ldots, \lambda_{q-1} \lambda_q)', \quad V(s) := (s_{11}/2, \ldots, s_{qq}/2, s_{12}, \ldots, s_{1q}, s_{23}, \ldots, s_{2q}, \ldots, s_{q,q-1})'.
\]

(18)

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

\[
t(\psi, \pi) := \begin{pmatrix} \eta - \eta' \\ t(\lambda, \pi) \end{pmatrix}, \quad s_{\ell k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda\ell k} \end{pmatrix}, \quad \text{where } t(\lambda, \pi) := \alpha(1 - \alpha)v(\lambda), \quad s_{\eta k} := \begin{pmatrix} \nabla_\theta f_\ell^* / f_k^* \\ \nabla_\theta f_k^* / f_k^* \end{pmatrix},
\]

(19)

and \( s_{\lambda\ell k} := V(s_{\lambda\ell k}) \) with \( s_{\lambda\ell k} := \nabla_\lambda \log \mathcal{P}_{\psi, \pi}(Y_k | Y_0^{k-1}) / [\mathcal{P}_{\psi, \pi}(Y_k | Y_0^{k-1}) \alpha(1 - \alpha)] \), namely,

\[
s_{\lambda\ell k} = \frac{\nabla_\theta f_\ell^* \nabla_\theta f_\ell^*}{f_k^*} + \sum_{t=1}^{k-1} \left( \frac{\nabla_\theta f_\ell^* \nabla_\theta f_\ell^*}{f_k^*} + \frac{\nabla_\theta f_k^* \nabla_\theta f_k^*}{f_k^*} \right).
\]

(20)

Here, \( s_{\ell k} \) in (19) depends on \( \varrho \) but not on \( \alpha \) and corresponds to \( s_{\eta k} \) 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 \( \mathcal{N}_\varepsilon := \{ \vartheta_2 \in \Theta_{nc} : |t(\psi, \pi)| < \varepsilon \} \). Let \( A_{nc}(\xi) := \{ \vartheta \in \mathcal{N}_\varepsilon : \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) \geq 0 \} \cup \mathcal{N}_{\varepsilon / \sqrt{n}} \), where we suppress the subscript 2 from \( \xi_2 \). We use this definition of \( A_{nc}(\xi) \) through Sections 6.1-6.3.

Assumption 6. \( 0 < \inf_{\vartheta \in \Theta_{nc}} \lambda_{\min}(I_{\vartheta}) \leq \sup_{\vartheta \in \Theta_{nc}} \lambda_{\max}(I_{\vartheta}) < \infty \) for \( I_{\vartheta} = \lim_{k \to \infty} E_{\vartheta^*}(s_{\ell k} s_{\ell k}^*) \), where \( s_{\ell k} \) is given in (19).

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

\[
\sup_{\xi \in \Xi} \sup_{\vartheta \in A_{nc}(\xi)} \left| \ell_n(\psi, \pi, \xi) - \ell_n(\psi^*, \pi, \xi) - \sqrt{n}t(\psi, \pi)' \nu_n(s_{\ell k}) + nt(\psi, \pi)' I_{\vartheta} t(\psi, \pi) / 2 \right| = o_p(1).
\]

(21)

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

\[
I_{\eta} := E_{\vartheta^*}(s_{\eta k} s_{\eta k}^*), \quad I_{\lambda_{\theta_1 k}} := \lim_{k \to \infty} E_{\vartheta^*}(s_{\lambda_{\theta_1 k} k} s_{\lambda_{\theta_2 k}}), \quad I_{\lambda_{\eta k}} := \lim_{k \to \infty} E_{\vartheta^*}(s_{\lambda_{\ell k} k} s_{\lambda_{\ell k} k}^*), \quad I_{\lambda_{\ell k}} := \lim_{k \to \infty} E_{\vartheta^*}(s_{\lambda_{\ell k} k} s_{\lambda_{\ell k} k}^*),
\]

\[
I_{\eta_{\lambda k}} := I_{\lambda_{\theta_1 k}}, \quad I_{\lambda_{\eta_{\theta_1 k}}} := I_{\lambda_{\theta_1 k}} - I_{\lambda_{\eta k}} I_{\eta}^{-1} I_{\eta_{\lambda k}}, \quad I_{\lambda_{\eta k}} := I_{\lambda_{\eta k}}, \quad Z_{\lambda_{\theta}} := (I_{\lambda_{\theta_{1 k}}})^{-1} G_{\lambda_{\theta_{1 k}}}.
\]

(22)
where $G_{\alpha,\eta}$ is a $q_\lambda$-vector mean zero Gaussian process indexed by $\varrho$ with $\text{cov} (G_{\lambda,\eta_1}, G_{\lambda,\eta_2}) = I_{\lambda,\eta_1,\eta_2}$. Define the set of admissible values for $\sqrt{n}(1 - \alpha)\nu(\lambda)$ 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

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

(23)

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

**Proposition 9.** Suppose Assumptions 1, 2, 4, 5 and 6 hold. Then, under the null hypothesis of $M = 1, 2 [\ell_n(\hat{\theta}_2, \xi_2) - \ell_0(\hat{\theta}_1)] \xrightarrow{d} \sup_{\varrho \in \Theta_{\varrho}} \left( \tilde{I}_{\lambda,\varrho} I_{\lambda,\eta} \tilde{I}_{\lambda,\varrho} \right)$. 

[Let’s use $\xrightarrow{d}$ to denote the convergence in distribution without referring to stochastic process. This is the notation Andrew (2001) uses.] In proposition 9, the LRTS and its asymptotic distribution depends on the choice of $\varrho$. 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 $\mu_j$ and variance $\sigma_j^2$. We split $\theta_j$ into $\theta_j = (\zeta_j, \sigma_j^2)' = (\mu_j, \beta_j, \sigma_j^2)'$, and write the density for the $j$-th regime as

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

(24)

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

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

(25)

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

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

(26)
Let \( \psi := (\eta', \lambda')' \in \Theta_\psi = \Theta_\eta \times \Theta_\lambda \), where \( \eta := (\gamma', \nu', \nu')' \) and \( \lambda := (\lambda', \lambda')' \). 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 \( \Theta_\psi \). Henceforth, we suppress \( \bar{y}_{k-1} \) from \( g_\psi(y_k|\bar{y}_{k-1}, x_k) \). We proceed to derive the derivatives of \( g_\psi(y_k|x_k) \) with respect to \( \psi \). \( \nabla_\psi g_\psi(y_k|x_k) \), \( \nabla_{\lambda\lambda'} g_\psi(y_k|x_k) \), and \( \nabla_{\lambda\lambda''} g_\psi(y_k|x_k) \) are the same as those given in (13) except for \( \nabla_{\lambda2} g_\psi(y_k|x_k) \) and that those with respect to \( \lambda_\mu \) are multiplied by \( 2^j \). Higher-order derivatives of \( g_\psi(y_k|x_k) \) with respect to \( \lambda_\mu \) are derived by following Kasahara and Shimotsu (2015). From Lemma 5 and the fact that the normal density \( f(\mu, \sigma^2) \) satisfies

\[
\nabla_{\mu^2} f(\mu, \sigma^2) = 2\nabla_{\sigma^2} f(\mu, \sigma^2), \quad \nabla_{\mu^3} f(\mu, \sigma^2) = 2\nabla_{\mu_\sigma^2} f(\mu, \sigma^2), \quad \text{and} \\
\nabla_{\mu^4} f(\mu, \sigma^2) = 2\nabla_{\sigma^2} f(\mu, \sigma^2) = 4\nabla_{\sigma^2} f(\mu, \sigma^2),
\]

we have

\[
\nabla_{\lambda_i^*} g_k^* = d_{ik} \nabla_{\mu} f(Y_k|\gamma^*, \theta^*), \quad i = 1, \ldots, 4,
\]

where

\[
d_{0k} := 1, \quad d_{1k} := q_k - \alpha, \quad d_{2k} := (q_k - \alpha)(C_2 - \alpha), \quad d_{3k} := 2(q_k - \alpha)^2(1 - \alpha - q_k), \\
d_{4k} := -2(q_k - \alpha)^4 + 3(q_k - \alpha)^2(\alpha - C_2)^2.
\]

It follows from \( \mathbb{E}_{\theta^*}[g_k|\bar{Y}_{-\infty}^k] = \alpha \), Lemma 13(a), and elementary calculation that

\[
\begin{align*}
\mathbb{E}_{\theta^*}[d_{ik}|\bar{Y}_{-\infty}^k] &= 0, \quad \mathbb{E}_{\theta^*}[\nabla_{\lambda_i^*} g_k^*|\bar{Y}_{-\infty}^k] = 0, \quad i = 1, 2, 3, \\
\mathbb{E}_{\theta^*}[d_{4k}|\bar{Y}_{-\infty}^k] &= \alpha (1 - \alpha) b(\alpha), \\
\mathbb{E}_{\theta^*}[\nabla_{\lambda_i^*} g_k^*|\bar{Y}_{-\infty}^k] &= \alpha (1 - \alpha) b(\alpha) \nabla_{\mu^*} f(Y_k; \gamma^*, \theta^*) \\
&= \alpha (1 - \alpha) b(\alpha) 4\nabla_{\sigma^2 \mu^*} f(Y_k; \gamma^*, \theta^*) = b(\alpha) \mathbb{E}_{\theta^*}[\nabla_{\lambda_i^*} g_k^*|\bar{Y}_{-\infty}^k]
\end{align*}
\]

with \( b(\alpha) := -(2/3)(\alpha^2 - \alpha + 1) < 0 \). Hence, \( \mathbb{E}_{\theta^*}[\nabla_{\lambda_i^*} g_k^*|\bar{Y}_{-\infty}^k] \) and \( \mathbb{E}_{\theta^*}[\nabla_{\lambda_i^*} g_k^*|\bar{Y}_{-\infty}^k] \) are linearly dependent.

We proceed to derive \( \nabla_j p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1})/p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1}) \). Repeating the calculation leading to (15)-(17) and using (29) gives the following: first, (15) and (16) still hold; second, the elements of \( \nabla_{\lambda\lambda'} p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1})/p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1}) \) except for the (1, 1)th element are given by (17) after adjusting that the derivative with respect to \( \lambda_\sigma \) must be multiplied by 2 (e.g., \( \mathbb{E}_{\theta^*}[\nabla_{\lambda_\sigma} g_k^*|\bar{Y}_{-\infty}^k] = 2\nabla_{\sigma^2 \mu^*} f_k^* \) and \( \mathbb{E}_{\theta^*}[\nabla_{\lambda_\sigma} g_k^*|\bar{Y}_{-\infty}^k] = 2\nabla_{\sigma^2 \mu^*} f_k^* \)); third,

\[
\nabla_{\lambda_i^2} p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1})/p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1}) = \alpha(1 - \alpha) \sum_{t=1}^{k-1} q^{k-t} \left( 2\nabla_{\mu_i f_t^*} f_k^* / f_k^* \right).
\]

When \( q \neq 0 \), \( \nabla_{\lambda_i^2} p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1})/p_{\psi^*}(Y_k|\bar{Y}_{0}^{k-1}) \) is a non-degenerate random variable as in the non-
normal case. When \( \varrho = 0 \), however, \( \nabla \lambda^2 \bar{P}_{\psi, \pi}(Y_k | Y_0^{k-1}) / \bar{P}_{\psi, \pi}(Y_k | Y_0^{k-1}) \) becomes identically equal to 0, and indeed the first non-zero derivative with respect to \( \lambda_\mu \) 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 \( \lambda^4_\mu \) (and other terms) when \( \varrho \neq 0 \) and a quadratic function of \( \lambda^4_\mu \) 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 \( \varrho \lambda^2_\mu, \lambda^4_\mu, \) and other terms.

For \( m \geq 0 \), define \( \zeta_{k,m}(\varrho) := \sum_{t=-m+1}^{k-1} \varrho^{k-t} 2^{t-k} (2n \lambda^4_\mu / f_{k,t}^* / f_{t}^*) \). Then, we can write (30) as

\[
\frac{\nabla \lambda^2 \bar{P}_{\psi, \pi}(Y_k | Y_0^{k-1})}{\alpha(1-\alpha) \bar{P}_{\psi, \pi}(Y_k | Y_0^{k-1})} = \sum_{t=1}^{k-1} \varrho^{k-t} \left( \frac{2 \nabla \lambda^4_\mu}{f_{k,t}^*} \right) = \varrho \zeta_{k,0}(\varrho).
\]

Note that \( \zeta_{k,m}(\varrho) \) satisfies \( \mathbb{E}_{\varrho^*} \zeta_{k,m}(\varrho) | Y_0^{k-1} \) = 0 and is non-degenerate even when \( \varrho = 0 \).

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

\[
t(\psi, \pi) := \left( \frac{\eta - \eta^*}{t_\lambda(\lambda, \pi)} \right),
\]

where

\[
t_\lambda(\lambda, \pi) := \begin{pmatrix}
\alpha(1-\alpha) \varrho \lambda^2_\mu \\
\alpha(1-\alpha) \lambda_\mu \lambda_\sigma \\
\alpha(1-\alpha) \lambda_\beta \lambda_\mu \\
\alpha(1-\alpha) \lambda_\beta \lambda_\sigma \\
\alpha(1-\alpha) v(\lambda_\beta)
\end{pmatrix},
\]

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

\[
\begin{pmatrix}
s_{\lambda_\beta, \psi \epsilon k} & s_{\lambda_\mu, \psi \epsilon k} & s_{\lambda_\sigma, \psi \epsilon k} \\
s_{\lambda_\beta, \psi \epsilon k} & s_{\lambda_\beta, \psi \epsilon k} & s_{\lambda_\sigma, \psi \epsilon k} \\
s_{\lambda_\sigma, \psi \epsilon k} & s_{\lambda_\sigma, \psi \epsilon k} & s_{\lambda_\sigma, \psi \epsilon k}
\end{pmatrix}
= \frac{\nabla \theta \epsilon f_{k,t}^*}{f_{k,t}^*} + \sum_{t=1}^{k-1} \varrho^{k-t} \left( \frac{\nabla \theta \epsilon f_{k,t}^*}{f_{k,t}^*} + \frac{\nabla \theta \epsilon f_{k,t}^*}{f_{k,t}^*} \right).
\]

19
Define the generalized score as

\[ s_{\ell k} := \begin{pmatrix} s_{\eta k} \\ s_{\lambda k} \end{pmatrix}, \quad \text{where} \quad s_{\eta k} := \left( \nabla f_k^*/f_k^* \right) \quad \text{and} \quad s_{\lambda k} := \left( \begin{array}{c} \nabla f_k^*/f_k^* \\ \nabla s_{\lambda k} \\ \nabla V(s_{\lambda k}) \end{array} \right). \tag{35} \]

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

**Assumption 7.** (a) \( 0 < \inf_{\theta \in \Theta_p} \lambda_{\min}(I_{\lambda}) \leq \sup_{\theta \in \Theta_p} \lambda_{\max}(I_{\lambda}) < \infty \) for \( I_{\lambda} = \lim_{t \to \infty} E_{\lambda}(s_{\ell k} s_{\ell k}') \), where \( s_{\ell k} \) is given in (35). (b) \( \sigma_1^*, \sigma_2^* \geq \epsilon_{\sigma} \).

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

\[ \sup_{\xi \in \Xi} \sup_{\theta \in A_n(\xi)} |\ell_n(\psi, \pi, \xi) - \ell_n(\psi, \pi, \xi) - \sqrt{n}t(\psi, \pi)'\nu_n(s_{\ell k}) + nt(\psi, \pi)'I_{\lambda}t(\psi, \pi)/2| = o_p(1). \tag{36} \]

Let \( \Lambda_{\lambda n} \) be the set of possible values of \( \sqrt{n}t(\lambda, \pi) \) defined in (33). The asymptotic null distribution of \( 2[\ell_n(\hat{\theta}_2, \xi_2) - \ell_{0,n}(\hat{\theta}_1)] \) is characterized by the supremum of \( 2t'_{\lambda}G_{\lambda, \eta} - t'_{\lambda}I_{\lambda, \eta n0}t_{\lambda} \), where \( G_{\lambda, \eta} \) and \( I_{\lambda, \eta n0} \) are defined analogously to those in (22) but with \( s_{\ell k} \) defined in (35), and the supremum is taken with respect to \( t_{\lambda} \) and \( \theta \in \Theta_p \) under the constraint implied by the limit of \( \Lambda_{\lambda n} \) as \( n \to \infty \). This constraint is given by \( \Lambda_{1}^{1} \) and \( \Lambda_{2}^{2} \), where \( q_{\beta} := \dim(\beta) \), \( q_{\lambda} := 3 + 2q_{\beta} + q_{\beta}(q_{\beta} + 1)/2 \), and

\[ \Lambda_{1}^{1} := \left\{ t_{\lambda} = (t_{\ell}^{\mu_2}, t_{\lambda, \sigma}, t_{\sigma, 2}, t_{\beta, \mu}, t_{\mu, \beta}, t_{v(\beta)})' \in \mathbb{R}^{q_{\lambda}} : \right. \]

\[ \left. (t_{\ell}^{\mu_2}, t_{\lambda, \sigma}, t_{\sigma, 2}, t_{\beta, \mu}, t_{\mu, \beta})' \in \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{q_{\beta}}, t_{\lambda, \sigma} = 0, t_{v(\beta)} = 0 \right\}; \tag{37} \]

\[ \Lambda_{2}^{2} := \left\{ t_{\lambda} = (t_{\ell}^{\mu_2}, t_{\lambda, \sigma}, t_{\sigma, 2}, t_{\beta, \mu}, t_{\mu, \beta})' \in \mathbb{R}^{q_{\lambda}} : \right. \]

\[ \left. t_{\ell}^{\mu_2} = \rho \lambda_{\mu}^2, t_{\mu, \sigma} = \lambda_{\mu} \lambda_{\sigma}, t_{\sigma, 2} = \lambda_{\sigma}^2, t_{\beta, \mu} = \lambda_{\beta} \lambda_{\mu}, t_{\mu, \beta} = \lambda_{\beta} \lambda_{\sigma}, t_{v(\beta)} = v_{\beta}(\lambda_{\beta}) \right\} \quad \text{for some} \quad \lambda \in \mathbb{R}^{q_{\beta}} \].

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

Define \( Z_{\lambda0} \) and \( I_{\lambda, \eta0} \) as in (22) but with \( s_{\ell k} \) defined in (35). Let \( Z_{\lambda0} \) and \( I_{\lambda, \eta0} \) denote \( Z_{\lambda0} \) and
\(I_{\lambda,\eta_0}\) evaluated at \(\varrho = 0\). Define \(\bar{I}_{\lambda}^1\) and \(\bar{I}_{\lambda,\varrho}^2\) by

\[
\begin{align*}
 r_{\lambda}(\bar{I}_{\lambda}^1) &= \inf_{t_\lambda \in \Lambda_{\lambda 1}^1} r_{\lambda}(t_\lambda), \quad r_{\lambda}(t_\lambda) := (t_\lambda - Z_{\lambda 0})' I_{\lambda,\eta_0}(t_\lambda - Z_{\lambda 0}) \\
r_{\lambda,\varrho}(\bar{I}_{\lambda,\varrho}^2) &= \inf_{t_\lambda \in \Lambda_{\lambda,\varrho}^2} r_{\lambda,\varrho}(t_\lambda), \quad r_{\lambda,\varrho}(t_\lambda) := (t_\lambda - Z_{\lambda \varrho})' I_{\lambda,\eta_0}(t_\lambda - Z_{\lambda \varrho}).
\end{align*}
\tag{38}
\]

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

**Proposition 11.** Suppose that assumptions in Proposition 10 hold. Then, under the null hypothesis of \(M = 1\),
\[
2[\ell_n(\hat{\vartheta}_2, \xi_2) - \ell_{0,n}(\hat{\vartheta}_1)] \xrightarrow{d} \sup_{\varrho \in \Theta_0^*} \max \{ \mathbb{I}\{\varrho = 0\}(\bar{I}_{\lambda}^1)' I_{\lambda,\eta_0}(\bar{I}_{\lambda}^1), (\bar{I}_{\lambda,\varrho}^2)' I_{\lambda,\eta_0}^{\varrho}(\bar{I}_{\lambda,\varrho}^2) \}.
\]

**Remark 1.** Qu and Fan (2017) derived the asymptotic distribution of the LRTS under the restriction that \(\varrho \geq \epsilon > 0\). While the assumption that \(\varrho > 0\) may be realistic for many Markov regime switching models with two regimes, the corresponding assumption might not hold for regime switching models with more than three regimes in empirical applications.

**Remark 2.** It is straightforward to extend our analysis to the average exponential-type LR 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 \(\mu_j\) but with common variance \(\sigma^2\). We split \(\gamma\) and \(\theta_j\) into \(\gamma = (\tilde{\gamma}, \sigma^2)\) and \(\theta_j = (\mu_j, \beta_j)\), and write the density for the \(j\)-th regime as

\[
f(y_k|\bar{Y}_{k-1}; \gamma, \theta_j) = f(y_k|\bar{Y}_{k-1}; \tilde{\gamma}, \beta_j, \sigma^2) = \frac{1}{\sigma}\phi \left( \frac{y_k - \mu_j - \omega(\bar{Y}_{k-1}; \tilde{\gamma}, \beta_j)}{\sigma} \right). \tag{39}\]

for some function \(\omega\). Consider the following reparameterization:

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

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

\[
g_\psi(y_k|\bar{Y}_{k-1}, x_k) = f(y_k|\bar{Y}_{k-1}; \tilde{\gamma}, \nu_\theta + (q_k - \alpha)\lambda, \nu_\sigma - \alpha(1-\alpha)\lambda^2_\mu). \tag{41}\]

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

\[
\begin{align*}
\nabla_{\lambda_\mu} g_k^* &= d_{1k} \nabla_{\gamma\eta} f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*) \quad \text{for } i = 0, 1, \ldots, \\
\nabla_{\lambda_\mu} g_k^* &= d_{1k} \nabla_{\mu\eta} f(Y_k|\overline{Y}_{k-1};\gamma^*,\theta^*) \quad \text{for } i = 0, 1, \ldots, 4,
\end{align*}
\]

(42)

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 \( E_{\omega}[q_k|\overline{Y}_{\infty}] = \alpha, \) Lemma 13(a), and elementary calculation that

\[
\begin{align*}
E_\omega[\nabla_{\lambda_\mu} g_k|\overline{Y}_0^k] &= E_\omega[d_{ik}|\overline{Y}_0^k] = 0, \quad i = 1, 2,  \\
E_\omega[d_{2k}|\overline{Y}_0^k] &= (1 - \alpha)(1 - 2\alpha), \quad E_\omega[d_{4k}|\overline{Y}_0^k] = (1 - \alpha)(1 - 6\alpha + 6\alpha^2). 
\end{align*}
\]

(43)

Repeating the calculation leading to (15)–(17) and using (43) gives the following: first, (15) and (16) still hold; second, the elements of \( \nabla_{\lambda_\mu}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) \) are given by (17) except for the \((1,1)\)th element; third, \( \nabla_{\lambda_\mu}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) \) is given by (30). Further, Lemma 7 in the Appendix shows that, when \( \varrho = 0, \nabla_{\lambda_\mu}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = (1 - \alpha)(1 - 2\alpha)\nabla_{\mu^*} f_k^*/f_k^* \) and \( \nabla_{\lambda_\mu}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = (1 - \alpha)(1 - 6\alpha + 6\alpha^2)\nabla_{\mu^*} f_k^*/f_k^* \). Because \( \nabla_{\lambda_\mu}\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1})/\overline{p}_{\psi^*\pi}(Y_k|\overline{Y}_0^{k-1}) = 0 \) when \( \alpha = 1/2 \) and \( \varrho = 0 \), we expand \( \ell_{\alpha}(\psi, \pi, \xi) \) four times and express it in terms of \( \theta_\lambda^2, (1 - 2\alpha)\lambda_\mu^3, \lambda_\mu^4, \) and other terms to establish a uniform approximation.

Collect the relevant parameters as

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

(44)

Define the generalized score as

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

(45)

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

$\Lambda_\lambda^1 := \{t_\lambda = (t_{\theta_1}, t_{\theta_2}, t_{\mu_1}, t_{\mu_2}, t_{\mu_3}, t_{\mu_4}, t_{\nu_1})', t_{\nu_2} = 0\},$

$\Lambda_{\lambda_0}^{-2} := \{t_\lambda = (t_{\theta_1}, t_{\theta_2}, t_{\mu_1}, t_{\mu_2}, t_{\mu_3}, t_{\mu_4}, t_{\nu_1})': t_{\theta_1} = \theta_1^2, t_{\mu_1} = 0, t_{\beta_1} = \lambda_0 \lambda_1, t_{\nu_2} = \nu_2(\lambda_0)\}$

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

**Assumption 8.** $0 < \inf_{\phi \in \Theta_{\phi}} \lambda_{\min}(I_{\phi}) \leq \sup_{\phi \in \Theta_{\phi}} \lambda_{\max}(I_{\phi}) < \infty$ for $I_{\phi} = \lim_{k \to \infty} E_{\phi} \cdot (s_{\phi k} s_{\phi k}')$, where $s_{\phi k}$ is given in (45).

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

**Proposition 13.** Suppose that assumptions in Proposition 12 hold. Then, under the null hypothesis of $M = 1, 2(\ell_1 (\hat{\theta}_1, \hat{\theta}_2) - \ell_0 (\hat{\theta}_1)) \delta_0 \sup_{\phi \in \Theta_{\phi}} \max \{I_{\phi} = 0\} (\hat{t}_1, \hat{t}_2) I_{\lambda, \eta_0} (\hat{t}_1, \hat{t}_2) I_{\lambda, \eta_0} (\hat{t}_1, \hat{t}_2), \}$, where $\hat{t}_1$ and $\hat{t}_2$ are defined as in (38) but in terms of $(Z_{\lambda, \eta}, I_{\lambda, \eta_0}, Z_{\lambda_0}, I_{\lambda, \eta_0})$ constructed with $s_{\phi k}$ defined in (45) and $\Lambda_{\lambda}^1$ and $\Lambda_{\lambda_0}^{-2}$ defined in (46).

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

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

Let $\theta_{M_0}^* = ((\theta_{M_0,x}^*)', (\theta_{M_0,y}^*)')'$ denote the parameter of the $M_0$-regime model, where $\theta_{M_0,x}^*$ contains $p_{ij}^* = q_{\theta_{M_0,x}}^* (i, j) > 0$ for $i = 1, \ldots, M_0$ and $j = 1, \ldots, M_0 - 1$, and $\theta_{M_0,y}^* = ((\theta_1^*, \ldots, (\theta_{M_0}^*)'))'$. We assume $\max \sum_{j=1}^{M_0-1} p_{ij}^* < 1$, and we assume $\theta_1^* < \ldots < \theta_{M_0}^*$ for identification. The true $M_0$-regime conditional density function of $Y_{1n}$ given $Y_0$ and $x_0$ is

$$p_{\theta_{M_0}^*}(Y_{1n} | Y_0, x_0) = \sum_{x_0} \prod_{k=1}^{n} p_{\theta_{M_0}^*}(Y_k, x_k | Y_{k-1}, x_{k-1}),$$

(47)

where $p_{\theta_{M_0}^*}(y_k, x_k | Y_{k-1}, x_{k-1}) = g_{\theta_{M_0,y}^*}(y_k | Y_{k-1}, x_k) q_{\theta_{M_0,x}^*}(x_k | Y_{k-1}, x_k)$ with $g_{\theta_{M_0,y}^*}(y_k | Y_{k-1}, x_k) = \sum_{j=1}^{M_0} \mathbb{I}\{x_k = j\} f(y_k | Y_{k-1}, \gamma, \theta_{M_0,y}^*).$

Let the conditional density of $Y_{1n}$ of an $(M_0 + 1)$-regime model be

$$p_{\theta_{M_0+1}^*}(Y_{1n} | Y_0, x_0) = \sum_{x_0} \prod_{k=1}^{n} p_{\theta_{M_0+1}^*}(Y_k, x_k | Y_{k-1}, x_{k-1}),$$

(48)

where $p_{\theta_{M_0+1}^*}(y_k, x_k | Y_{k-1}, x_{k-1})$ is defined similarly to $p_{\theta_{M_0}^*}(y_k, x_k | Y_{k-1}, x_{k-1})$ with $\theta_{M_0+1,x}^* := \{p_{ij}\}_{i=1}^{M_0+1,j=1,...,M_0}$ and $\theta_{M_0+1,y}^* := (\theta_1^*, \ldots, \theta_{M_0+1}^*, \gamma)'$. We assume that $\min_{i,j} p_{ij} \geq \epsilon$ for
some $\epsilon \in (0, 1/2)$, and let $\Theta_{M_0+1,\epsilon}$ denote $\Theta_{M_0+1}$ with this restriction.

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

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

Define the set of values of $\theta_{M_0+1}$ that yields the true density $(47)$ under $p_{\theta_{M_0+1}}$ as $\Upsilon^* := \{ m_{M_0+1} \in \Theta_{M_0+1,\epsilon} : p_{\theta_{M_0+1}}(Y_n^m| Y_0^m, x_0) = p_{\theta_{M_0+1}}(Y_n^m| Y_0^m, x_0) \ P_{\theta_{M_0+1}} -a.s. \}$. Under $H_{0m}$, the $(M_0+1)$-regime model $(48)$ generates the true $M_0$-regime density $(47)$ if $\theta_m = \theta_{m+1} = \theta_{M_0+1}^* \text{ and the transition matrix of } X_k$ reduces to that of the true $M_0$-regime model. Define $J_m := \{ m, m+1 \}$ and $J_m := \{ 1, \ldots, M_0 + 1 \} \ \backslash \ J_m$, and let $p_j$ and $p_j^*$ denote $p_{\theta_{M_0+1}}(X_k = j)$ and $p_{\theta_{M_0+1}}(X_k = j)$, respectively. Define the subset of $\Upsilon^*$ that corresponds to $H_{0m}$ as

$$\Upsilon^*_m := \{ \theta_{M_0+1} \in \Theta_{M_0+1,\epsilon} : \theta_j = \theta_{j+1}^* \text{ for } 1 \leq j < m; \ \theta_m = \theta_{m+1} = \theta_{M_0+1}^*; \ \theta_j = \theta_{j+1}^* \text{ for } h + 1 \leq j \leq M_0 \}; \ \gamma = \gamma^*; \ \theta_{M_0+1,x} \in \Theta_{\theta_{M_0+1}} \times \Pi_m^*,$$

where, with $\wedge$ and $\vee$ denoting “and” and “or”,

$$\Theta_{\theta_{M_0+1}} := \{ \{ p_{ij} \}_{i,j=1}^{M_0+1} \{ p_{ij} \}_{i,j=1}^{M_0+1} : p_{ij} = p_{ij}^* \text{ for } 1 \leq j < h \text{ and } p_{ij} = p_{ij}^* \text{ for } m + 1 < j \leq M_0 \},$$

$$\Pi_m^* := \{ \{ p_{ij} \}_{i,j=1}^{M_0+1} \{ p_{ij} \}_{i,j=1}^{M_0+1} : p_{im} + p_{m+1,im} = p_{im}^* \text{ for } \forall i \leq M \},$$

then $\Upsilon^* = \bigcup_{m=1}^{M_0} \Upsilon^*_m$ holds. Lemma 18 in the Appendix shows that $\Upsilon^*_m$ gives the parameter set that corresponds to $H_{0m}$.

For $M = M_0, M_0 + 1$, let $\ell_n(\theta_M, \xi_M) := \log \left( \sum_{x=0}^{M} p_{\theta_M}(Y_n^m| Y_0^m, x_0) \xi_M(x_0) \right)$ denote the $M$-regime log-likelihood for a given initial distribution $\xi_M(x_0) \in \Xi_M$. We treat $\xi_M(x_0)$ as fixed. Let $\hat{\theta}_{M_0} := \arg \max_{\theta_{M_0}} \ell_n(\theta_{M_0}, \xi_M)$ and $\hat{\theta}_{M_0+1} := \arg \max_{\theta_{M_0+1}} \ell_n(\theta_{M_0+1}, \xi_{M_0+1})$. The following proposition shows that the MLE is consistent in the sense that the distance between $\hat{\theta}_{M_0+1}$ and $\Upsilon^*$ 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) $\Theta_{M_0}$ and $\Theta_{M_0+1}$ are compact. (b) For all $(x, x') \in \mathcal{X}$ and all $(y, y', w) \in \mathcal{Y} \times \mathcal{W}$, the function $(\theta, \gamma) \mapsto f(y'| y_0, w; \gamma, \theta)$ is continuous. (c) $E_{\theta_{M_0}} \log(p_{\theta_{M_0}}(Y_1| Y_0^0, W_{-m}^0)) = E_{\theta_{M_0}} \log(p_{\theta_{M_0}}(Y_1| Y_0^0, W_{-m}^0))$ for all $m \geq 0$ if and only if $\theta_{M_0} = \theta_{M_0}^*$; (d) $E_{\theta_{M_0}} \log(p_{\theta_{M_0}}(Y_1| Y_0^0, W_{-m}^0)) = E_{\theta_{M_0}} \log(p_{\theta_{M_0+1}}(Y_1| Y_0^0, W_{-m}^0))$ for all $m \geq 0$ if and only if $\theta_{M_0+1} = \theta_{M_0+1}^*$; (e) $\ell_{ij} > 2\epsilon$ for all $i \neq j \in \{ 1, \ldots, M_0 \}$.

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

---

5 Strictly speaking, when $m = M_0$, we need to redefine $\theta_{M_0+1,x}$ so that $\theta_{M_0+1,x}$ contains $p_1, \ldots, p_{M_0+1}$ and $p_1$ is determined by $p_{11} = 1 - \sum_{j=2}^{M_0} p_{1j}$. We suppress this technical detail.
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 \( h \). Split \( \vartheta_{M_0+1,x} \) as \( \vartheta_{M_0+1,x} = (\vartheta_{x,m}^1, (\vartheta_{x,m}^c)^\prime) \), where \( \vartheta_{x,m} := \{p_{ij}\}_{i\in J_m, j\in J_m\setminus\{M_0+1\}} \) is identified under \( H_{0m} \), and \( \vartheta_{x,m}^c := \{p_{ij}\}_{i\in J_m, j\in J_m\setminus\{M_0+1\}} \) is not point identified under \( H_{0m} \). We reparameterize \( (p_{mm}, p_{m,m+1}, p_{m+1,m}, p_{m,m+1}) \in \vartheta_{x,m}^c \) to \( (p_{mm}, p_{mm,J}, p_{m+1,J}, p_{m+1,m+1}, J) \), where \( p_{mm} := p_{mm} + p_{m+1,m+1} \), \( p_{mm,J} := p_{mm,J} \), \( p_{m+1,J} := p_{m+1,m+1} \), and \( p_{m+1,m+1,J} := p_{m+1,m+1}/p_{m+1,J} \). If \( X_1^k \in J_m^k \), then \( X_1^k \) follows a two-state Markov chain on \( J_m \) with transition matrix \( P_J = \begin{pmatrix} 1-p_{mm,J} & 1-p_{m+1,m+1,J} \\ p_{mm,J} & p_{m+1,m+1,J} \end{pmatrix} \). We further reparameterize \( p_{mm,J} \) and \( p_{m+1,m+1,J} \) to \( \alpha_m := \mathbb{P}_{\vartheta_{M_0}}(X_k = h | X_k \in J_m) \) and \( \varrho_m := \text{corr}_{\vartheta_{M_0}}(X_{k-1}, X_k | (X_{k-1}, X_k) \in J_m^2) \) and collect reparameterized \( \vartheta_{x,m}^c \) into \( \pi_m := (\alpha'_m, \varrho'_m)^\prime \), where \( \varphi_m \) collects the elements of \( \vartheta_{x,m}^c \) that are not point identified under \( H_{0m} \) and that do not affect the transition probability of \( X_1^k \) when \( X_1^k \in J_m^k \).

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

\[
\alpha_m = \mathbb{E}_{\vartheta_{M_0}}(q_{km} | X_k \in J_m, \Upsilon_{-\infty}^\infty) \quad \text{and} \quad \varrho_m = \text{corr}_{\vartheta_{M_0}}(q_{k-1,m}, q_{km} | (X_{k-1}, X_k) \in J_m^2, \Upsilon_{-\infty}^\infty). \tag{49}
\]

### 7.1 Non-normal distribution

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

\[
\begin{pmatrix} \nu_m \\ \nu_m + (1 - \alpha_m) \lambda_m \end{pmatrix} = \begin{pmatrix} \nu_m \\ \nu_m - \alpha_m \lambda_m \end{pmatrix}.
\]

Collect the reparameterized identified parameters into one vector \( \psi_m := (\eta_m, \lambda_m^\prime)^\prime \), where \( \eta_m = (\gamma', \{\theta_j^1\}_{j=1}^{M_0+1}, \nu_m' \}, \{\theta_j'\}_{j=m+2}^{M_0+1}, \vartheta_{x,m}^c), \) so that the reparameterized \((M_0+1)\)-regime log-likelihood function is \( \ell_m(\psi_m, \pi_m, \xi_{M_0+1}) \). Let \( \psi_m = (\eta_m^*, \lambda_m^*) = ((\vartheta_{M_0}^*)^\prime, \theta^*) \) denote the value of \( \psi_m \) under \( H_{0m}. \)

Define the reparameterized conditional density of \( y_k \) as

\[
g_{\psi_m}^m(y_k | \bar{y}_{k-1}; x_k) := \mathbb{I}(x_k \in J_m) f(y_k | \bar{y}_{k-1}; \gamma, \nu_m + (q_{km} - \alpha_m) \lambda_m) + \sum_{j \in J_m} q_{kj} f(y_k | \bar{y}_{k-1}; \gamma, \theta_j).
\]
Let \( f^*_{mk} \) denote \( f(Y_k | \bar{Y}_{k-1}; \gamma^*, \theta^*_{mk}) \). It follows from (49) and the law of iterated expectations that

\[
\mathbb{E}_{\theta^*_{M_0}} \left[ \frac{I(X_k \in J_m)(q_{km} - \alpha_m)}{g^m_{\psi_m^*}(Y_k | \bar{Y}_{k-1}, X_k)} \right] = 0,
\]

where the second equality holds because \( g^m_{\psi_m^*}(Y_k | \bar{Y}_{k-1}, X_k) = f^*_{mk} \) if \( X_k \in J_m \), and last equality holds because, conditional on \( \{X^t_{2i} \in J^{2i-t_i-1}_m, \bar{Y}_{-\infty} \} \), \( X^t_{2i} \) is a two-state stationary Markov process with transition probability \( P_j \).

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

\[
\nabla_{q^*_{\psi_{m}^*}} (Y_k | \bar{Y}_{0}^{k-1}) / \mathbb{P}_{q^*_{\psi_{m}^*}} (Y_k | \bar{Y}_{0}^{k-1}) = \begin{cases} \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla_{(\gamma^*, \theta^*_{M_0}, x)^t} \log g^t_{\theta^*_{M_0}} \right] \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) - \sum_{t=1}^{k} \mathbb{E}_{\theta^*} \left[ \nabla_{(\gamma^*, \theta^*_{M_0}, x)^t} \log g^t_{\theta^*_{M_0}} \right] \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) \end{cases}
\]

(51)

\[
\nabla_{\lambda_m} \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) / \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) = 0, \quad \nabla_{\lambda_m} \eta_m \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) / \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) = 0, \quad \text{and}
\]

(52)

\[
\nabla_{\lambda_m, \lambda_m} \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) / \mathbb{P}_{\psi_{m}^*} (Y_k | \bar{Y}_{0}^{k-1}) = \alpha_m (1 - \alpha_m) \nabla_{(\gamma^*, \theta^*_{M_0}, x)^t} f^*_{mk} \mathbb{P}_{\theta^*_{M_0}} (X_k \in J_m | \bar{Y}_{0}^{k}) + \alpha_m (1 - \alpha_m) \sum_{t=1}^{k} \mathbb{E}_{\theta^*_{M_0}} \left[ \frac{\nabla_{(\gamma^*, \theta^*_{M_0}, x)^t} \log g^t_{\theta^*_{M_0}}}{f^*_{mk}} \right] \mathbb{P}_{\theta^*_{M_0}} (X_k \in J_m | \bar{Y}_{0}^{k})
\]

(53)

Define \( \hat{\rho} := (q_1, \ldots, q_{M_0})', \) define \( t_\lambda (\lambda_m, \pi_m) \) as \( t_\lambda (\lambda, \pi) \) in (19) by replacing \( (\lambda, \pi) \) with \( (\lambda_m, \pi_m) \),
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 $\Upsilon^*_M$ similar to the one in Proposition 8: under $H_0 : M = M_0$, for all $c > 0$ and for $m = 1, \ldots, M_0$,

$$\sup_{\xi \in \varTheta_{M_0+1} \in \Lambda^m_{nc}(\xi)} \left| \ell_n(\psi_m, \pi_m, \xi) - \ell_n(\psi^*, \pi_m, \xi) - \sqrt{nt(\psi_m, \pi_m)} \nu_n(s_{\varrho_m}k^c) + nt(\psi_m, \pi_m) \bar{I}_{\varrho_m} t(\psi_m, \pi_m)/2 \right| = o_p(1),$$

where $s_{\varrho_m}k^c := (s^m_{\varrho, k^c}, s^m_{\varrho, k^c})'$, $\bar{I}_{\varrho_m} = \lim_{k \to \infty} \mathbb{E}_{\varrho_{M_0}}(s_{\varrho_m}k^c s^m_{\varrho_m}k^c)$, and $\Lambda^m_{nc}(\xi) := \{ \varrho_{M_0+1} \in \mathcal{N}_c^m : \ell_n(\psi_m, \pi_m, \xi) - \ell_n(\psi^*, \pi_m, \xi) \geq 0 \} \cap \mathcal{N}_c^m$, $\mathcal{N}_c^m := \{ \varrho_{M_0+1} \in \Theta_{M_0+1} : |t(\psi_m, \pi_m)| < \varepsilon \} \cap \mathcal{N}_c^m$, where $\mathcal{N}_c^m$ is an arbitrary small neighborhood of $\Upsilon^*_M$. 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 $\varrho_m$ under restriction $p_{ij} \geq \varepsilon$ by $\varTheta_{\varrho_m \varepsilon}$, and let $\bar{\Theta}_{\varrho \varepsilon} := \varTheta_{\varrho_1 \varepsilon} \times \ldots \times \varTheta_{\varrho_{M_0} \varepsilon}$.

**Assumption 10.** $0 < \inf_{\bar{\varrho} \in \bar{\Theta}_{\varrho \varepsilon}} \lambda_{\min}(\bar{\bar{I}}_{\bar{\varrho}}) \leq \sup_{\bar{\varrho} \in \bar{\Theta}_{\varrho \varepsilon}} \lambda_{\max}(\bar{\bar{I}}_{\bar{\varrho}}) < \infty$ for $\bar{\bar{I}}_{\bar{\varrho}} := \lim_{k \to \infty} \mathbb{E}_{\varrho_{M_0}}(s^m_{\varrho k} s^m_{\varrho k})$. 


where $\tilde{s}_{mk}$ is given in (54).

**Proposition 15.** Suppose Assumptions 1, 2, 4, 9, and 10 hold. Then, under $H_0 : M = M_0$, $2[\ell_n(\hat{\theta}_{M_0+1}; \xi_{M_0+1}) - \ell_n(\hat{\theta}_{M_0}, \xi_{M_0})] \leq \max_{n=1, \ldots, M_0} \{\sup_{\theta_n \in \Theta_n} \left(\left(\overline{I}_{\lambda_n}^{\hat{\theta}_{m}}\right)^{\hat{\lambda}_n \eta_{\hat{\theta}_{m}}}/\lambda_{\hat{\theta}_{m}}\right)\}.$

### 7.2 Heteroscedastic normal distribution

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

$$
\begin{pmatrix}
\zeta_m \\
\zeta_m + \sigma_m^2 \\
\sigma_m^2 \\
\sigma_m^2 + 1
\end{pmatrix}
= \begin{pmatrix}
\nu_{\zeta_m} + (1 - \alpha_m)\lambda_{\zeta_m} \\
\nu_{\zeta_m} - \alpha_m \lambda_{\zeta_m} \\
\nu_{\sigma_m} + (1 - \alpha_m)(2\lambda_{\sigma_m} + C_1\lambda_{\mu_m}^2) \\
\nu_{\sigma_m} - \alpha_m(2\lambda_{\sigma_m} + C_1\lambda_{\mu_m}^2)
\end{pmatrix},
$$

where $\nu_{\zeta_m} = (\nu_\mu, \nu_\theta')', \lambda_{\zeta_m} = (\lambda_{\mu_m}, \lambda_{\mu_{m+1}})'$, $C_1 := -(1/3)(1 + \alpha_m)$, and $C_2 := (1/3)(2 - \alpha_m)$. As in Section 7.1, we collect the reparameterized identified parameters into $\psi_m := (\eta', \lambda', \theta')$, where $\eta_m = (\gamma_j, \{\theta_j\}_{j=1}^{m-1}, \nu_{\zeta_m}, \nu_{\sigma_m}, \{\theta_j\}_{j=m+2}^{M-1}, \nu_{\zeta_m}, \nu_{\sigma_m})'$ and $\lambda_m := (\lambda_{\mu_m}, \lambda_{\mu_{m+1}})'. Similar to (26), define the reparameterized conditional density of $y_k$ as

$$g_{\psi_m}^m(y_k|\overline{y}_{k-1}, \gamma, \theta) = \sum_{j \in J_m} q_{kj} f(y_k|\overline{y}_{k-1}, \gamma, \theta_j) + \mathbb{I}\{x_k \in J_m\} f(y_k|\overline{y}_{k-1}, \gamma, \nu_{\zeta_m} + (q_{km} - \alpha_m)\lambda_{\zeta_m}, \nu_{\sigma_m} + (q_{km} - \alpha_m)(2\lambda_{\sigma_m} + (C_2 - q_{km})\lambda_{\mu_m}^2)).$$

Let $g_k^{m, \ast}, \nabla g_k^{m, \ast},$ and $\nabla f^{s, \ast}$ denote $g_{\psi_m}^m(Y_k|\overline{y}_{k-1}, X_k), \nabla g_{\psi_m}^m(Y_k|\overline{y}_{k-1}, X_k),$ and $\nabla f(Y_k|\overline{y}_{k-1}; \gamma, \theta_m).$ It follows from (28) and a derivation similar to (50) that

$$\mathbb{E}_{\theta_m} \nabla_{\lambda^{m, \ast}} g_k^{m, \ast} \frac{g_k^{m, \ast}}{g_k} \nabla_{\lambda^k} = 0, \quad i = 1, 2, 3,$$

$$\mathbb{E}_{\theta_m} \nabla_{\lambda^{m, \ast}} g_k^{m, \ast} \frac{g_k^{m, \ast}}{g_k} \nabla_{\lambda^k} = \alpha_m(1 - \alpha_m)b(\alpha_m)(\nabla_{\mu^s} f_{\mu k}/f_{\mu k})\mathbb{P}_{\theta_m}(X_k \in J_m|\overline{y}_{k-1})$$

$$= b(\alpha_m)\mathbb{E}_{\theta_m} \nabla_{\lambda^{m, \ast}} g_k^{m, \ast} \frac{g_k^{m, \ast}}{g_k} \nabla_{\lambda^k}$$

which corresponds to (29) in testing homogeneity. Repeating the calculation leading to (51)–(53) and using (57) gives the following: first, (51) and (52) still hold; second, the elements of $\nabla_{\lambda m} \mathbb{P}_{\psi^{\ast}}(Y_k|\overline{y}_{k-1}^0)/\mathbb{P}_{\psi^{\ast}}(Y_k|\overline{y}_{k-1}^0)$ except for the (1, 1)th element are given by (53) while adjusting the derivative with respect to $\lambda_{\sigma_m}$ by multiplying by 2; third,

$$\frac{\nabla_{\lambda^{m, \ast}} \mathbb{P}_{\psi^{\ast}}(Y_k|\overline{y}_{k-1}^0)}{\mathbb{P}_{\psi^{\ast}}(Y_k|\overline{y}_{k-1}^0)} = \alpha_m(1 - \alpha_m) \sum_{t=1}^{k-1} g_{k-t} \frac{\nabla_{\mu^s} f_{\mu k}/f_{\mu k}}{f_{\mu k}} \mathbb{P}_{\theta_m}(X_t, X_k \in J_m|\overline{y}_{k-1}^0).$$

For $m \geq 0$, define $\zeta_{s,k,m}^{k}(\gamma_m) := \sum_{t=-m+1}^{k-1} \zeta_{s,k-t}^{k-1}(\gamma_m) = \sum_{t=-m+1}^{k-1} \zeta_{s,k-t}^{k-1}(\gamma_m)$.
Define $s^m_{\lambda \theta m k}$ as in (55) but using density (24), and denote each element of $s^m_{\lambda \theta m k}$ as

$$s^m_{\lambda \theta m k} := \begin{pmatrix} \ast & s^m_{\lambda \theta m k} & s^m_{\lambda \sigma \theta m k} \\ s^m_{\lambda \mu \theta m k} & s^m_{\lambda \beta \theta m k} & s^m_{\lambda \sigma \theta m k} \\ s^m_{\lambda \sigma \theta m k} & s^m_{\lambda \sigma \theta m k} & s^m_{\lambda \sigma \theta m k} \end{pmatrix}.$$ 

Define $\tilde{s}^m_{\lambda \theta m k}$ as in (54) with redefining $s^m_{\lambda \theta m k}$ in (54) as

$$s^m_{\lambda \theta m k} := \left( \begin{array}{c} \tilde{c}^m_{k,0}(\theta_m) / 2 \\ 2s^m_{\lambda \mu \theta m k} \\ 2s^m_{\lambda \sigma \theta m k} \end{array} \right)' 2(s^m_{\lambda \theta m k})' V(s^m_{\lambda \theta m k})'.$$ (58)

Define $\tilde{I}^m_{\lambda \eta \theta m}$ and $Z^m_{\lambda \theta m}$ as in (56) with $s^m_{\lambda \theta m k}$ defined in (58). Let $Z^m_{\lambda \theta m}$ and $\tilde{I}^m_{\lambda \eta \theta m}$ denote $Z^m_{\lambda \theta m}$ and $\tilde{I}^m_{\lambda \eta \theta m}$ evaluated at $\theta_m = 0$. Define $\Lambda^1_{\lambda}$ as in (37), and define $\Lambda^2_{\lambda \eta \theta m}$ as in (37) with replacing $\varphi$ with $\tilde{\varphi}_m$. Similar to (38), define $\tilde{t}^m_{\lambda}$ and $\tilde{t}^m_{\lambda \eta \theta m}$ by $r_{\lambda}(\tilde{t}^m_{\lambda}) = \inf_{t_{\lambda} \in \Lambda^1_{\lambda}} r_{\lambda}(t_{\lambda})$ and $r_{\lambda \eta \theta m}(\tilde{t}^m_{\lambda \eta \theta m}) = \inf_{t_{\lambda} \in \Lambda^2_{\lambda \eta \theta m}} r_{\lambda \eta \theta m}(t_{\lambda})$, where $r_{\lambda}(t_{\lambda}) := (t_{\lambda} - Z^m_{\lambda \theta m})/\tilde{I}^m_{\lambda \eta \theta m}(t_{\lambda} - Z^m_{\lambda \theta m})$, and $r_{\lambda \eta \theta m}(t_{\lambda}) := (t_{\lambda} - Z^m_{\lambda \theta m})/\tilde{I}^m_{\lambda \eta \theta m}(t_{\lambda} - Z^m_{\lambda \theta m})$.

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 $\tilde{s}^m_{\lambda \theta m k}$ is given in (58).

**Proposition 16.** Suppose Assumptions 1, 2, 4, 9, and 11 hold and the component density for the $j$-th regime is given by (24). Then, under $H_0 : m = M_0$, $\sqrt{\ell_n(\hat{\theta}_{M_0}, \xi_{M_0}) - \ell_n(\hat{\theta}_{M_0}, \xi_{M_0}) \sim_d \max_{m = 1, \ldots, M_0} \sup_{\theta_m \in \Theta_{\theta m}} \max \{ I\{ \theta_m = 0 \}, (\tilde{t}^m_{\lambda} / \tilde{I}^m_{\lambda \eta \theta m}), (\tilde{t}^m_{\lambda \eta \theta m} / \tilde{I}^m_{\lambda \eta \theta m}) \}$. 

### 7.3 Homoscedastic normal distribution

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

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

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

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

$$s^m_{\lambda \theta m k} := \begin{pmatrix} \ast & s^m_{\lambda \theta m k} & s^m_{\lambda \sigma \theta m k} \\ s^m_{\lambda \mu \theta m k} & s^m_{\lambda \beta \theta m k} & s^m_{\lambda \sigma \theta m k} \\ s^m_{\lambda \sigma \theta m k} & s^m_{\lambda \sigma \theta m k} & s^m_{\lambda \sigma \theta m k} \end{pmatrix}.$$
Define $\tilde{q}_k$ as in (54) with redefining $s_{\lambda,\varrho,m,k}^n$ in (54) as

$$s_{\lambda,\varrho,m,k}^n := \left( \zeta_{k,0}(\varrho_m)/2 \right) s_{\lambda,k}^{m-1} / 3! \quad s_{\lambda,k}^{m-1} / 4! \quad (s_{\lambda,\varrho,m,k}^n)' \quad V(s_{\lambda,\varrho,m,k}^n)'', \quad (59)$$

where $s_{\lambda,k}^{m} := \nabla_{m'} f(Y_k | \overline{Y}_{k-1}; \gamma^*, \vartheta_m)/f(Y_k | \overline{Y}_{k-1}; \gamma^*, \vartheta_m)$ for $i = 3, 4$.

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

**Assumption 12.** Assumption 10 holds when $\tilde{q}_k$ is given in (59).

**Proposition 17.** Suppose Assumptions 1, 2, 4, 9, and 12 hold and the component density for the $j$-th regime is given by (39). Then, under $H_0 : m = M_0$, $2[\ell_n(\hat{\varrho}_{M_0+1}, \xi_{M_0+1}) - \ell_n(\hat{\varrho}_{M_0}, \xi_{M_0})] \xrightarrow{d} \max_{m=1,...,M_0} \sup_{\varrho_m \in \Theta_{\varrho_m}} \max\{\ell(\varrho_m) - \ell_1(\overline{Y}_n^{m_1}) \lambda_{m_1}^{m_1} \lambda_{m_2}^{m_2}, (\lambda_{m_1},\lambda_{m_2}) \}}$, where $\hat{\varrho}_1$ and $\hat{\varrho}_2$ are defined as in Proposition 16 but in terms of $(Z_{\lambda,\varrho}^m, T_{\lambda,\varrho}^m, Z_{\lambda,0}^m, T_{\lambda,0}^m)$ constructed with $s_{\lambda,\varrho,m,k}^n$ given in (59) and $\Lambda_{\lambda}$ and $\Lambda_{\lambda,\varrho}^2$ defined as in (46) but replacing $\rho$ with $\varrho_m$.

### 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 $\varpi \in \Theta_{\varpi,\varrho}$, we define a local parameter $h = \sqrt{n}t(\psi, \varpi)$ so that

$$h = \left( h_\eta \right) = \left( \sqrt{n}(\eta - \eta^*) \right) \left( \sqrt{n}t(\lambda, \pi) \right)$$

for $h_\lambda \in \sqrt{n}t(\Theta_{\lambda, \varrho})$ that does not depend on $n$, where $t(\lambda, \pi)$ differs across different models and is given by (20), (33), and (44). Given $h = (h_\eta, h_\lambda)'$ and $\varpi \in \Theta_{\varpi,\varrho}$, we consider the sequence of contiguous local alternatives $\varrho_n = (\psi_n', \pi_n') = (\eta_n', \lambda_n', \pi_n')' \in \Theta_\eta \times \Theta_\lambda \times \Theta_\varpi$ such that

$$\eta_n = \eta^* + h_\eta/\sqrt{n}, \quad h_\lambda = \sqrt{n}t(\lambda_n, \pi_n) + o(1), \quad \text{and} \quad |\pi_n - \pi| = o(1). \quad (60)$$

Let $P_{\varrho}$ be the probability measure on $\{Y_k\}_{k=1}^n$ conditional on the value of $\overline{Y}_0$ and $W_1$ given the initial distribution $\xi$ under $\varrho$ that are obtained from the probability measure $P_\varrho$ via appropriate marginalization and conditioning. Then, the log likelihood ratio is given by

$$\log \frac{dP_{\varrho_n}}{dP_{\varrho}} = \ell_n(\psi_n', \pi_n', \xi) - \ell_n(\psi^*, \pi, \xi) = \log \frac{\sum x_k^n \xi(x_0) \prod_{k=1}^n f_k(\theta_n, \lambda_n) q_{\pi_n}(x_{k-1}, x_k)}{\prod_{k=1}^n f_k(\psi^n, 0)},$$

where $f_k(\theta_n, \lambda_n)$ is defined by the right hand side of (12), (26), and (41) for the models of non-normal distribution, heteroskedastic normal distribution, and homoskedastic normal distribution, respectively.

The following result is useful to derive the asymptotic distribution of our LRT under $P_{\varrho_n}$.
Proposition 18. Suppose that the assumptions of Propositions 8, 10, and 12 hold. Then, for the models of non-normal, heteroskedastic normal, and homoskedastic normal distributions, (a) $\mathbb{P}_{\theta, n}$ is mutually contiguous with respect to $\mathbb{P}_{\theta, n}^0$, and (b) $\log \frac{d\mathbb{P}_{n}}{d\mathbb{P}_{n}^0} = h'\nu_n(s_{\phi, n}) - \frac{1}{2} h'\mathcal{I}_0 h + o_{s_{\phi, n}}^n(1)$ with $\nu_n(s_{\phi, n}) \xrightarrow{d} N(\mathcal{I}_0 h, \mathcal{I}_0)$ under $\mathbb{P}_{\theta, n}^0$.

The 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 that of LRTS under the null hypothesis by appropriately shifting the mean of the Gaussian process.

8.1 Non-normal distribution

For non-normal distribution, the sequence of local alternatives is given by $\lambda_n = \tilde{\lambda}/n^{1/4}$ because $h_\lambda = \sqrt{n}\alpha(1 - \alpha)v(\lambda_n)$ holds by choosing $\lambda = \pm \sqrt{\alpha(1 - \alpha)}h_\lambda$ for any $h_\lambda > 0$ and $\pi \in \Theta_\alpha$.

The following proposition derives the asymptotic distribution of LRTS for non-normal distribution under $H_{1n} : (\pi_n, \eta_n) = (\tilde{\pi}, \tilde{\eta}^*)$ and $\lambda_n = \tilde{\lambda}/n^{1/4}$.

Proposition 19. Suppose that the assumptions of Proposition 9 hold. Then, under $H_{1n} : (\pi_n, \eta_n) = (\tilde{\pi}, \tilde{\eta}^*)$ and $\lambda_n = \tilde{\lambda}/n^{1/4}$ for $\tilde{\pi} \in \Theta_\alpha$ and $\tilde{\eta} \neq 0$, we have $2[\ell_n(\tilde{\theta}, \xi_2) - \ell_0(\tilde{\theta})] \xrightarrow{d} \sup_{\pi \in \Theta_\alpha} \tilde{P}_{\theta, \bar{\lambda}_0} \mathcal{I}_{\lambda_\theta, \eta}\bar{\lambda}_0$ given $\bar{\lambda}_0$ defined by (23) with $Z_{\lambda_\theta} = (\mathcal{I}_{\lambda, \eta})^{-1}\mathcal{G}_{\lambda, \eta}$, where $E_{\theta_n}[\mathcal{G}_{\lambda, \eta}] = \mathcal{I}_{\lambda, \eta} h_\lambda$ with $h_\lambda := \bar{\alpha}(1 - \bar{\alpha})v(\lambda)$ and $\text{cov}_{\theta_n}(\mathcal{G}_{\lambda, \eta_1}, \mathcal{G}_{\lambda, \eta_2}) = \mathcal{I}_{\lambda, \eta_1, \eta_2}$ with $\mathcal{I}_{\lambda, \eta_1}$ and $\mathcal{I}_{\lambda, \eta_2}$ given by (22) for $s_{\phi, n}$ defined in (19).

8.2 Heteroskedastic normal distribution

For the model with heteroskedastic normal distribution, the sequences of contiguous local alternatives characterized by (60) include the local alternatives of order $n^{-1/8}$.

Proposition 20. Suppose that the assumptions of Proposition 11 hold for the model (24). For $\kappa \in [0, 1/4]$, let

$$H_{1n} : (\tilde{\gamma}_n, \alpha_n, \eta_n) = (\tilde{\gamma}/n^\kappa, \tilde{\alpha}, \eta^*)$$

and $$(\lambda_{\mu n}, \lambda_{\sigma n}, \lambda_{\beta n}) = (\tilde{\lambda}_\mu/n^{1/4 - \kappa/2}, \tilde{\lambda}_\sigma/n^{1/4 + \kappa/2}, \tilde{\lambda}_\beta/n^{1/4 + \kappa/2}),$$

where $$(\tilde{\gamma}_n, \alpha_n) \in \Theta_\pi$$ with $\tilde{\gamma} \neq 0$ and $\tilde{\lambda} := (\tilde{\lambda}_\mu, \tilde{\lambda}_\sigma, \tilde{\lambda}_\beta)' \neq (0, 0, 0)'$. Then, under $H_{1n}$, we have $2[\ell_n(\tilde{\theta}, \xi_2) - \ell_0(\tilde{\theta})] \xrightarrow{d} \sup_{\pi \in \Theta_\alpha} \text{max} \{I(\tilde{\phi} = 0)\} \mathcal{I}_{\lambda_\theta, \eta_1, \eta_2} \mathcal{I}_{\lambda_\theta, \eta_1, \eta_2}$ with $\mathcal{I}_{\lambda_\theta, \eta_1}$ and $\mathcal{I}_{\lambda_\theta, \eta_2}$ defined by (38) with $Z_{\lambda_\theta} = (\mathcal{I}_{\lambda, \eta})^{-1}\mathcal{G}_{\lambda, \eta}$, where $E_{\theta_n}[\mathcal{G}_{\lambda, \eta}] = \mathcal{I}_{\lambda, \eta} h_\lambda$ with

$$h_\lambda = \begin{cases} 
\bar{\alpha}(1 - \bar{\alpha}) \times (\tilde{\gamma}^2_{\mu, \lambda_\mu} \tilde{\lambda}_\mu \tilde{\lambda}_\sigma, b(\tilde{\alpha})\tilde{\lambda}_\mu^{12}/12, \tilde{\lambda}_\beta \tilde{\lambda}_\mu, 0, 0)' & \text{if } \kappa \in (1/6, 1/4], \\
\bar{\alpha}(1 - \bar{\alpha}) \times (\tilde{\gamma}^2_{\mu, \lambda_\mu} \tilde{\lambda}_\mu \tilde{\lambda}_\sigma, \tilde{\lambda}_\sigma^2, \tilde{\lambda}_\beta \tilde{\lambda}_\mu, 0, 0)' & \text{if } \kappa \in (0, 1/6), \\
\bar{\alpha}(1 - \bar{\alpha}) \times (\tilde{\gamma}^2_{\mu, \lambda_\mu} \tilde{\lambda}_\mu \tilde{\lambda}_\sigma, \tilde{\lambda}_\sigma^2, \tilde{\lambda}_\beta \tilde{\lambda}_\mu, v(\tilde{\lambda}_\beta)', \tilde{\lambda}_\beta)' & \text{if } \kappa = 0,
\end{cases}$$

and $\text{cov}_{\theta_n}(\mathcal{G}_{\lambda, \eta_1}, \mathcal{G}_{\lambda, \eta_2}) = \mathcal{I}_{\lambda, \eta_1, \eta_2}$ with $\mathcal{I}_{\lambda, \eta_1}$ and $\mathcal{I}_{\lambda, \eta_2}$ given by (22) for $s_{\phi, n}$ defined in (35).
When $\kappa > 0$, the local alternatives are in the neighborhood of $\rho = 0$ and their orders for $\lambda_\mu$ are slower than $n^{-1/4}$. When $\kappa = 1/4$, the local alternatives for $\lambda_\mu$ are in the order of $n^{-1/8}$. Our test has non-trivial power against these local alternatives in the neighborhood of $\rho = 0$ because the test statistic $(\hat{t}_\lambda^1)^\top I_{\lambda, \rho}^\top \hat{t}_\lambda^1$ has power against the local alternatives in $H_{1n}$ for $\kappa \in (0, 1/4]$. In contrast, the existing test of Carrasco et al. (2014) does not have power against the local alternatives in the neighborhood of $\rho = 0$. See Section 5 of Carrasco et al. (2014). The test proposed by Qu and Fan (2017) assumes that $\rho$ is bounded away from zero and hence their test also lacks power against the local alternatives in the neighborhood of $\rho = 0$.

8.3 Homoskedastic normal distribution

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

Proposition 21. Suppose that the assumptions of Proposition 12 hold for the model (39). Let

$$H_{1n}: (\varrho_n, \alpha_n, \eta_n, \lambda_{\mu n}, \lambda_{\beta n}) = \begin{cases} \left( \varrho/n^\kappa, 1/2 + \Delta \varrho/n^{3\kappa/2-1/4}, \eta^{\star}, \lambda_{\mu}/n^{1/4-\kappa/2}, \lambda_{\beta}/n^{1/4+\kappa/2} \right) & \text{for } \kappa \in (1/6, 1/4], \\ \left( \varrho/n^\kappa, \alpha^{\star}, \eta^{\star}, \lambda_{\mu}/n^{1/4-\kappa/2}, \lambda_{\beta}/n^{1/4+\kappa/2} \right) & \text{for } \kappa \in [0, 1/6], \end{cases}$$

where $(\varrho_n, \alpha_n) \in \Theta_{\pi \epsilon}$ with $\varrho \neq 0$, $\Delta \varrho \neq 0$, and $\lambda := (\lambda_{\mu}, \lambda_{\beta}) \neq (0, 0)^{\prime}$. Then, under $H_{1n}$, we have $2[\ell_n(\hat{\varrho}_2, \hat{\xi}_2) - \ell_{0,n}(\hat{\varrho}_1)] \overset{d}{\rightarrow} \sup_{\varrho \in \Theta_{\pi \epsilon}} \max \{1\{\varrho = 0\}(\hat{t}_\lambda^2)^\top I_{\lambda, \rho}^\top \hat{t}_\lambda^2, (\hat{t}_\lambda^2)^\top I_{\lambda, \rho}^\top \hat{t}_\lambda^2 \} + \hat{t}_\lambda^2$ and $I_{\lambda, \rho}^2$ defined as in (38) but in terms of $(Z_{\lambda_\rho}, I_{\lambda, \rho}^1, Z_{\lambda, 0}, I_{\lambda, \rho})$ constructed with $s_{ek}$ defined in (45) and $\Lambda_\lambda^2$ and $\Lambda_{\lambda_\rho}^2$ defined in (46), where $E_{\varrho_n}[G_{\lambda, \rho}] = I_{\lambda, \rho}^1 h_{\lambda}$ with

$$h_{\lambda} = \begin{cases} (1/4) \times (\hat{\varrho}^2_{\lambda_\mu} + \Delta \hat{\varrho}^2_{\lambda_\mu}, -(1/2)\hat{\lambda}^4_{\mu}, \hat{\lambda}_{\mu}, \hat{\lambda}_{\beta}, 0)^\prime & \text{if } \kappa \in (1/6, 1/4], \\ \lambda (1 - \alpha) \times (\hat{\varrho}^2_{\lambda_\mu}, (1 - 2\alpha)\hat{\lambda}^3_{\mu}, 0, \hat{\lambda}_{\mu}, \hat{\lambda}_{\beta}, 0)^\prime & \text{if } \kappa = 1/6, \\ \alpha (1 - \alpha) \times (\hat{\varrho}^2_{\lambda_\mu}, 0, 0, \hat{\lambda}_{\mu}, \hat{\lambda}_{\beta}, 0)^\prime & \text{if } \kappa \in (0, 1/6], \\ \alpha (1 - \alpha) \times (\hat{\varrho}^2_{\lambda_\mu}, 0, 0, \hat{\lambda}_{\mu}, \hat{\lambda}_{\beta}, v(\hat{\lambda}_{\beta})^\prime)^\prime & \text{if } \kappa = 0, \end{cases}$$

and $\text{cov}_{\varrho_n}(G_{\lambda, \rho_1}, G_{\lambda, \rho_2}) = I_{\lambda, \rho_1, \rho_2}$ with $I_{\lambda, \rho}$ and $I_{\lambda, \rho_1, \rho_2}$ given by (22) for $s_{ek}$ defined in (45).

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 $\hat{\varrho}_{M_0}$ and $\hat{\varrho}_{M_0+1}$ as $\hat{\varrho}_{M} := \arg \max_{\varrho \in \Theta_{M}, \xi_M} \ell_n(\varrho_M, \xi_M)$ for some choice of $\xi_M$ for $M = M_0, M_0 + 1$. Compute $LR_n = 2[\ell_n(\hat{\varrho}_{M_0+1}, \xi_{M_0+1}) - \ell_n(\hat{\varrho}_{M_0}, \xi_{M_0})]$.

2. Given $\hat{\varrho}_{M_0}$ and $\xi_{M_0}$, generate $B$ independent samples $\{y_{b_1}, ..., y_{b_B}\}$ under $H_0$ with $\varrho_{M_0} = \hat{\varrho}_{M_0}$ conditional on the observed values of $Y_0 = \bar{y}_0$ and $W_1^n = w_1^n$. 

32
3. For each simulated sample \( \{ y_k^b \}_{k=1}^n \) with \( (y_0, w_0) \), estimate \( \hat{\nu}_M^b \) and \( \hat{\nu}_{M+1}^b \) as in Step 1 and let \( LR_n^b = 2[\ell_n(\hat{\nu}_M^b; \xi_M) - \ell_n(\hat{\nu}_{M+1}^b; \xi_{M+1})] \) for \( b = 1, \ldots, B \).

4. Let \( c_{\alpha, B} \) be the \((1 - \alpha)\) quantile of \( LR_n^b \), \( b = 1, \ldots, B \), and define the bootstrap p-value by
   \[
   \frac{1}{B} \sum_{b=1}^B I\{LR_n^b > LR_n\}.
   \]

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 15, 16, and 17 hold. Then, for the models of non-normal, heteroskedastic normal, and homoskedastic normal distributions, 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 \( H_{1_n} \) in Propositions 15, 16, and 17.

10 Appendix

10.1 Definition of \( \Phi^{(5)}_{\theta_T}(F) \) and \( \Phi^{(6)}_{\theta_T}(F) \)

Define
\[
\Phi^{(5)}_{\theta_T}(F) := \frac{1}{|\sigma(I(5))|} \sum_{(\ell_1, \ldots, \ell_5) \in \sigma(I(5))} \left( E_0^\alpha \left[ \phi_{\theta_{\ell_1}} \phi_{\theta_{\ell_2}} \phi_{\theta_{\ell_3}} \phi_{\theta_{\ell_4}} \phi_{\theta_{\ell_5}} \right] | F \right)
- \sum_{(a,b,c),\{d,e\} \in \sigma_5} E_0^\alpha \left[ \phi_{\theta_a} \phi_{\theta_b} \phi_{\theta_c} \phi_{\theta_d} \phi_{\theta_e} \right] | F \right) E_0^\alpha \left[ \phi_{\theta_d} \phi_{\theta_e} \right] | F \right),
\]
\[
\Phi^{(6)}_{\theta_T}(F) := E_0^\alpha \left[ \phi_{\theta_a} \phi_{\theta_b} \phi_{\theta_c} \phi_{\theta_d} \phi_{\theta_e} \phi_{\theta_f} | F \right]
- \sum_{(a,b,c),\{d,e,f\} \in \sigma_{61}} E_0^\alpha \left[ \phi_{\theta_a} \phi_{\theta_b} \phi_{\theta_c} \phi_{\theta_d} \phi_{\theta_e} \phi_{\theta_f} | F \right]
+ 2 \sum_{(a,b),\{c,d,e\} \in \sigma_{62}} E_0^\alpha \left[ \phi_{\theta_a} \phi_{\theta_b} | F \right] E_0^\alpha \left[ \phi_{\theta_c} \phi_{\theta_d} | F \right] E_0^\alpha \left[ \phi_{\theta_e} \phi_{\theta_f} | F \right] ,
\]

where
\[
\sigma_5 := \text{the set of } \binom{5}{3} = 10 \text{ partitions of } \{1, 2, 3, 4, 5\} \text{ of the form } \{a, b, c\}, \{d, e\},
\]
\[
\sigma_{61} := \text{the set of } \binom{6}{3} = 15 \text{ partitions of } \{1, 2, 3, 4, 5, 6\} \text{ of the form } \{a, b, c, d\}, \{e, f\},
\]
\[
\sigma_{62} := \text{the set of } \binom{6}{2}/2 = 10 \text{ partitions of } \{1, 2, 3, 4, 5, 6\} \text{ of the form } \{a, b\}, \{c, d\}, \{e, f\},
\]
\[
\sigma_{63} := \text{the set of } \binom{6}{2}/6 = 15 \text{ partitions of } \{1, 2, 3, 4, 5, 6\} \text{ of the form } \{a, b\}, \{c, d\}, \{e, f\}.
\]

10.2 Proof of Propositions and Corollaries

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

Proof of Proposition 1. The proof is essentially identical to the proof of Lemma 2 in DMR. Therefore, the details are omitted. The only difference from DMR is (i) we do not impose Assumption (A2) of DMR, but this does not affect the proof because Assumption (A2) is not used in the proof of Lemma 2 in DMR, and (ii) we have $W^n_1$, but our Lemma 10 extends Corollary 1 of DMR to accommodate $W_k$'s. Consequently, the argument of the proof of DMR goes through.

Proof of Proposition 2. Let $m_{\vartheta k} := t'_\vartheta s_{\pi k} + r_{\vartheta k}$, so that $l_{\vartheta k x_0} - 1 = m_{\vartheta k} + u_{\vartheta k x_0}$. Observe that

$$\max_{1 \leq k \leq n} \sup_{\vartheta \in N_{c}/\sqrt{n}} |m_{\vartheta k}| = \max_{1 \leq k \leq n} \sup_{\vartheta \in N_{c}/\sqrt{n}} |t'_\vartheta s_{\pi k} + r_{\vartheta k}| = o_p(1),$$

from Assumption 3(a)(c) and Lemma 9. Define $h_{\vartheta k x_0} := \sqrt{l_{\vartheta k x_0}} - 1$. We first show

$$\sup_{\vartheta \in N_{c}/\sqrt{n}} \left|nP_n(h_{\vartheta k x_0}^2) - nt'_\vartheta I_\vartheta t_\vartheta / 4\right| = o_p(1).$$

To show (64), write $4P_n(h_{\vartheta k x_0}^2)$ as

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

It follows from Assumption 3(a)(b)(c)(e) and $(E|XY|)^2 \leq E|X|^2 E|Y|^2$ that, uniformly for $\vartheta \in N_\varepsilon$,

$$P_n(l_{\vartheta k x_0} - 1)^2 = t'_\vartheta P_n(s_{\pi k} s'_{\pi k}) t_\vartheta + 2t'_\vartheta P_n(s_{\pi k} (r_{\vartheta k} + u_{\vartheta k x_0})) + P_n(r_{\vartheta k} + u_{\vartheta k x_0})^2$$

$$= t'_\vartheta I_\vartheta t_\vartheta + o_p(|t_\vartheta|^2) + O_p(n^{-1}|t_\vartheta||\psi - \psi^*|) + O_p(n^{-1}|\psi - \psi^*|^2).$$

Then, (64) holds because the second term on the right of (65) is bounded by, from (63), $P_n(m_{\vartheta k}^2) = t'_\vartheta I_\vartheta t_\vartheta + o_p(|t_\vartheta|^2)$, and Assumption 3(e),

$$C \sup_{\vartheta \in N_\varepsilon/\sqrt{n}} P_n \left[|m_{\vartheta k}|^3 + 3m_{\vartheta k}^2 |u_{\vartheta k x_0}| + 3|m_{\vartheta k}| u_{\vartheta k x_0}^2\right] + C \sup_{\vartheta \in N_\varepsilon/\sqrt{n}} P_n(|u_{\vartheta k x_0}|^3)$$

$$\leq o_p(1) \sup_{\vartheta \in N_\varepsilon/\sqrt{n}} P_n \left[m_{\vartheta k}^2 + |m_{\vartheta k}| |u_{\vartheta k x_0}| + u_{\vartheta k x_0}^2\right] + C \sup_{\vartheta \in N_\varepsilon/\sqrt{n}} P_n(|u_{\vartheta k x_0}|^3) = o_p(n^{-1}).$$

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

$$h_{\vartheta k x_0} = (l_{\vartheta k x_0} - 1)/2 - h_{\vartheta k x_0}^2 = (t'_\vartheta s_{\pi k} + r_{\vartheta k} + u_{\vartheta k x_0})/2 - h_{\vartheta k x_0}^2/2.$$

It follows from (64), (67), and Assumption 3(d)(e) that $nP_n(h_{\vartheta k x_0}) = \sqrt{nt'_\vartheta f_n(s_{\pi k})/2 - nt_\vartheta I_\vartheta t'_\vartheta / 8 +$
\(\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) = 2\sum_{k=1}^{n} \log(1 + h_{\theta kx_0}) = nP_n(2h_{\theta kx_0} - [1 + o_p(1)]h_{\theta kx_0}^2)
\]
\[= \sqrt{n}l_\theta \nu_n(s_{\pi k}) - t_\theta I_\pi t_\theta/4 + nP_n(h_{\theta kx_0}^2) + o_p(1)
\]
\[= \sqrt{n}l_\theta \nu_n(s_{\pi k}) - t_\theta I_\pi t_\theta/2 + o_p(1),
\]
giving the stated result. \(\square\)

**Proof of Proposition 3.** For part (a), applying the inequality \(\log(1 + x) \leq x\) to the log-likelihood ratio function and using (67) give

\[\ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) = 2\sum_{k=1}^{n} \log(1 + h_{\theta kx_0}) \leq 2nP_n(h_{\theta kx_0}) = \sqrt{n}\nu_n(l_{\theta kx_0} - 1) - nP_n(h_{\theta kx_0}^2).
\]

From the first equality in (65), for some \(\delta > 0,
\]
\[P_n(h_{\theta kx_0}^2) \geq \left(\frac{l_{\theta kx_0} - 1}{l_{\theta kx_0} + 1}\right)^2
\]
\[\geq \frac{1}{\delta + 1} P_n(\mathbb{I}\{l_{\theta kx_0} \leq \delta\}(l_{\theta kx_0} - 1)^2)
\]
\[\geq \frac{1}{\delta + 1} [P_n((l_{\theta kx_0} - 1)^2) - P_n(\mathbb{I}\{l_{\theta kx_0} > \delta\}(l_{\theta kx_0} - 1)^2)].
\]

From Assumption 3(a)/(b)/(c)/(e), we have \(B := \sup_{\theta \in \mathcal{N}_c} |l_{\theta kx_0} - 1| \in L^2(\mathbb{P}_{\theta^*})\), and hence
\[
\lim_{\delta \to \infty} \sup_{\theta \in \mathcal{N}_c} P_n(\mathbb{I}\{l_{\theta kx_0} > \delta\}(l_{\theta kx_0} - 1)^2) \leq \lim_{\delta \to \infty} P_n(\mathbb{I}\{B + 1 > \delta\}B^2) = 0 \text{ almost surely.}
\]
Therefore, by choosing \(\delta\) sufficiently large, it follows from (66) and Assumption 3(f) that, uniformly for \(\theta \in \mathcal{N}_c,
\]
\[P_n(h_{\theta kx_0}^2) \geq \kappa l_\theta I_\pi t_\theta + o_p(|t_\theta|^2) + O_p(n^{-1}|t_\theta||\psi - \psi^*|) + O_p(n^{-1}|\psi - \psi^*|^2),
\]
for \(\kappa = (2(\delta + 1))^{-1} > 0\).

Because \(\sqrt{n}\nu_n(l_{\theta kx_0} - 1) = \sqrt{n}l_\theta[Z_n(\theta) + O_p(1)] + O_p(1)\) from Assumption 3(d)/(e)/(g), it follows from (68) and (69) that
\[0 \leq \ell_n(\psi, \pi, x_0) - \ell_n(\psi^*, \pi, x_0) \leq \sqrt{n}l_\theta[Z_n(\theta) + O_p(1)] - \kappa nt_\theta I_\pi t_\theta + o_p(n|t_\theta|^2) + O_p(1).
\]

The rest of the proof follows the proof of Theorem 1 of Andrews (1999). Let \(T_n := T_n^{1/2}\sqrt{n}t_\theta\). In view of Assumption 3(f)/(g), we can write (70) as
\[0 \leq ||T_n||O_p(1) - \kappa ||T_n||^2 + o_p(||T_n||^2) + O_p(1).
\]
Rearranging this equation gives \(||T_n||^2 \leq 2||T_n||\varsigma_n + O_p(1)\) with \(\varsigma_n = O_p(1)\). Then, \(||T_n|| - \varsigma_n\) \(\leq \varsigma_n^2 + O_p(1)\) = \(O_p(1)\), and taking square roots gives \(||T_n|| \leq O_p(1)\), giving part (a). Part (b) follows
from part (a) and Proposition 2.

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

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

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

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

$$
\nabla^j \ell_{k,m,x}(\vartheta) - \nabla^j \ell_{k,m}(\vartheta) = \Psi^j_{k,m,x}(\vartheta) \left( \frac{p_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k}, X_{-m} = x)}{p_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k}, X_{-m} = x)} - \frac{\bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})}{\bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})} \right)
$$

$$
+ \frac{\bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})}{\bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})} \left( \Psi^j_{k,m,x}(\vartheta) - \Psi^j_{k,m}(\vartheta) \right),
$$

where

$$
\Psi^j_{k,m,x}(\vartheta) := \frac{\nabla^j p_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k}, X_{-m} = x)}{p_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k}, X_{-m} = x)}, \quad \Psi^j_{k,m}(\vartheta) := \frac{\nabla^j \bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})}{\bar{p}_\vartheta(Y_{-m+1}^{-k} \mid Y_{-m}^{-k})}.
$$

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

$$
(A) \quad \Psi^j_{k,m,x}(\vartheta) \in L^{q_0}(\mathbb{P}, \vartheta^\ast), \quad \text{and} \quad (B) \quad |\Psi^j_{k,m,x}(\vartheta) - \Psi^j_{k,m}(\vartheta)| \leq A_j(k + m)^6 \rho^{k+m-1}. \quad (71)
$$
Note that, from (89) we have, suppressing \((\vartheta)\) and superscript 1 from \(\nabla_1 \ell_{k,m,x}\),

\[
\Psi_{k,m,x}^1 = \nabla \ell_{k,m,x}, \quad \Psi_{k,m,x}^2 = \nabla^2 \ell_{k,m,x} + (\nabla \ell_{k,m,x})^2,
\]

\[
\Psi_{k,m,x}^3 = \nabla^3 \ell_{k,m,x} + 3 \nabla^2 \ell_{k,m,x} \nabla \ell_{k,m,x} + (\nabla \ell_{k,m,x})^3,
\]

\[
\Psi_{k,m,x}^4 = \nabla^4 \ell_{k,m,x} + 4 \nabla^3 \ell_{k,m,x} \nabla \ell_{k,m,x} + 3 (\nabla^2 \ell_{k,m,x})^2 + 6 \nabla^2 \ell_{k,m,x} (\nabla \ell_{k,m,x})^2 + (\nabla \ell_{k,m,x})^4,
\]

\[
\Psi_{k,m,x}^5 = \nabla^5 \ell_{k,m,x} + 5 \nabla^4 \ell_{k,m,x} \nabla \ell_{k,m,x} + 10 \nabla^3 \ell_{k,m,x} \nabla^2 \ell_{k,m,x} + 10 \nabla^3 \ell_{k,m,x} (\nabla \ell_{k,m,x})^2 + 15 (\nabla^2 \ell_{k,m,x})^2 \nabla \ell_{k,m,x} + 10 \nabla^2 \ell_{k,m,x} (\nabla \ell_{k,m,x})^3 + (\nabla \ell_{k,m,x})^5,
\]

\[
\Psi_{k,m,x}^6 = \nabla^6 \ell_{k,m,x} + 6 \nabla^5 \ell_{k,m,x} \nabla \ell_{k,m,x} + 15 \nabla^4 \ell_{k,m,x} \nabla^2 \ell_{k,m,x} + 15 \nabla^4 \ell_{k,m,x} (\nabla \ell_{k,m,x})^2 + 10 \nabla^3 \ell_{k,m,x}^2 + 60 \nabla^3 \ell_{k,m,x} \nabla \ell_{k,m,x} \nabla \ell_{k,m,x} + 20 \nabla^3 \ell_{k,m,x} (\nabla \ell_{k,m,x})^3 + 15 (\nabla^2 \ell_{k,m,x})^3 + 45 (\nabla^2 \ell_{k,m,x})^2 (\nabla \ell_{k,m,x})^2 + 15 \nabla^2 \ell_{k,m,x} (\nabla \ell_{k,m,x})^4 + (\nabla \ell_{k,m,x})^6,
\]

and \(\Psi_{j,k,m,x}^i\) is written analogously with \(\nabla_{j,k,m,x}^i\) replacing \(\nabla \ell_{k,m,x}\).

Consequently, (A) of (71) follows from Proposition 5(c), the Cauchy-Schwartz inequality, and the Hölder’s inequality such as \(\sup_{W} \ell_{k,m,x} \leq \mathbb{E} W_{\vartheta} \in \mathcal{P}_{\vartheta} \sup_{W} \ell_{k,m,x} \rightarrow \mathbb{E} W_{\vartheta} \in \mathcal{P}_{\vartheta} \sup_{W} \ell_{k,m,x} \rightarrow \infty\) when \(a \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\) and \(b \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\), \(E|a|^{3\varrho/2}/2(\mathcal{P}_{\vartheta})<\infty \) when \(a \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\) and \(b \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\), and \(E|a|^{3\varrho/2}/2(\mathcal{P}_{\vartheta})<\infty \) when \(a \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\), \(b \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\), \(c \in L^{3\varrho/2}(\mathcal{P}_{\vartheta})\). (B) of (71) follows from Proposition 5(a)(c), the relation \(ab-\cd a=(a-b)-c(a-d), a^n-b^n=(a-b)\sum_{i=0}^{n-1}(a^{n-1-i}b^i)\), and the Hölder’s inequality.

Part (c) follows from parts (a) and (b). For part (d), \(\{\nabla_j l_{k,m,x}(\vartheta)\}_{m \geq 0}\) is uniformly in \(L^{\max(2,\dim(\vartheta))}(\mathcal{P}_{\vartheta})\) from writing \(\nabla_j l_{k,m,x}(\vartheta) = \sup_{Y_k \in \mathcal{P}_{\vartheta}} (\nabla Y_k \nabla_{-m-} X_{-m})\) \(\mathbb{P}_{\vartheta} (Y_k \mid Y_{-m-1} X_{-m} = x)\) \(\Psi_{j,k,m,x}^i(\vartheta)\), and using (71) and Lemma 4. From a similar argument, \(\{\nabla_j l_{k,m,x}(\vartheta)\}_{m \geq 0}\) is also uniformly in \(L^{\max(2,\dim(\vartheta))}(\mathcal{P}_{\vartheta})\). Therefore, \(\nabla_j l_{k,m,x}(\vartheta)\) is uniformly in \(L^{\max(2,\dim(\vartheta))}(\mathcal{P}_{\vartheta})\) from part (c) and the completeness of \(L^2(\mathcal{P}_{\vartheta})\) and part (d) is proven.

\(\square\)

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

We proceed to prove the consistency of \(\hat{\vartheta}_2\). Define, similarly to pp. 2265–2266 in DMR, \(\Delta_{k,m,x}(\vartheta_2):=\log p_{\vartheta_2}(Y_{k-1} X_{-m}, W_{m+1} X_{-m} = x), \Delta_{k,m}(\vartheta_2):=\log p_{\vartheta_2}(Y_{k-1} X_{-m}, W_{m+1})\),

\(\Delta_{k,\infty}(\vartheta_2):=\lim_{m \to \infty} \Delta_{k,m}(\vartheta_2)\), and \(\ell(\vartheta_2):=\mathbb{E}_{\vartheta_1}[\Delta_{k,\infty}(\vartheta_2)]\). 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 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 \(\mathbb{E} \hat{\vartheta}_1 \log p_{\vartheta_2}(Y_1 \mid \bar{Y}_0, W_{m})\) converges to \(\ell(\vartheta_2)\) uniformly in \(\vartheta_2\) as \(m \to \infty\) from Lemma 3.
Proof of Proposition 8. We prove the stated result by applying Corollary 1 to $l_{\partial kx_0} - 1$ with $l_{\partial kx_0}$ defined in (4). Because the first and second derivatives of $l_{\partial kx_0}$ play the role of the score, we expand $l_{\partial kx_0} - 1$ with respect to $\psi$ up to the third order. Let $d = \dim(\psi)$. Recall that the $p$-th order Taylor expansion of $f(x)$ with $x \in \mathbb{R}^d$ around $x = x^\ast$ is given by

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

where $\sum_{(u_1, \ldots, u_d)}$ denotes the sum over the sets of nonnegative integers $(u_1, \ldots, u_d)$ such that $u_1 + u_2 + \cdots + u_d = j$. For $m \geq 0$ and $j = 1, 2, \ldots$, let $A^j_{k,m,x} (\psi, \pi)$ denote the vector that collects the corresponding elements of $\psi - \psi^\ast$, i.e., the terms of the form $(\psi_1 - \psi_1^\ast)^{u_1}(\psi_2 - \psi_2^\ast)^{u_2} \cdots (\psi_d - \psi_d^\ast)^{u_d}$. Define $A^j_{k,m,x} (\psi, \pi)$ similarly to $A^j_{k,m,x} (\psi, \pi)$ by replacing $p\psi(Y_k|Y_m^{m-1})$ in (72) with $\bar{p}\psi(Y_k|Y_m^{m-1})$. With this notation, expanding $l_{\partial kx_0} - 1$ three times around $\psi^\ast$ while fixing $\pi$ gives, with $\bar{\psi} \in [\psi, \psi^\ast]$,}

$$
l_{\partial kx_0} - 1 = (\Delta^1 \psi)'A^1_{k,0,x_0} (\psi^\ast, \pi) + (\Delta^2 \psi)'A^2_{k,0,x_0} (\psi^\ast, \pi) + (\Delta^3 \psi)'A^3_{k,0,x_0} (\bar{\psi}, \pi)
= (\Delta^1 \psi)'A^1_{k,0} (\psi^\ast, \pi) + (\Delta^2 \psi)'A^2_{k,0} (\psi^\ast, \pi) + (\Delta^3 \psi)'A^3_{k,0} (\bar{\psi}, \pi) + u_{kx_0}(\psi, \pi),$$

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

Define, for $m \geq 0$,

$$s_{k,m}(\pi) := \left( \begin{array}{c}
\nabla_{\psi}p_{\psi^\ast}(Y_k|Y_m^{m-1}) \\
\bar{p}_{\psi^\ast}(Y_k|Y_m^{m-1}) \\
\nabla_{\bar{\psi}}(\lambda)p_{\psi^\ast}(Y_k|Y_m^{m-1}) \\
(\alpha(1-\alpha)p_{\psi^\ast}(Y_k|Y_0^{m-1}))
\end{array} \right),$$

where $\nabla_{\psi} := (\nabla_{\lambda_1 \lambda_1}/2, \ldots, \nabla_{\lambda_q \lambda_q}/2, \nabla_{\lambda_1 \lambda_2}, \ldots, \nabla_{\lambda_{q-1} \lambda_q})'$. Noting that $\nabla_{\lambda}p_{\psi^\ast}(Y_k|Y_0^{m-1}) = 0$
and \( \nabla \lambda^f \tilde{p}_{\lambda^*}(Y_k | \tilde{Y}_0^{k-1}) = 0 \) from (15), we may rewrite (73) as

\[
l_k \theta_{x_0} - 1 = t(\psi, \pi)' s_{k,0}(\pi) + r_{k,0}(\psi, \pi) + u_{k,x_0}(\psi, \pi),
\]

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

For \( m \geq 0 \), define \( \nu_{k,m}(\theta) := (s_{k,m}(\pi)', \tilde{\Lambda}_{k,m}(\pi)', \Lambda_{k,m}^3(\psi, \pi)'), \) and define \( v_{k,\infty}(\theta) := \lim_{m \to \infty} v_{k,m}(\theta) \).

Choose \( \epsilon > 0 \) sufficiently small that \( \mathcal{N}_\epsilon \) is a subset of \( \mathcal{N}^* \) in Assumption 4. Then, \( v_{k,\infty}(\theta) \) is a \( \mathbb{P}_{\theta^*} \)-stationary \( L^2(\mathbb{P}_{\theta^*}) \) martingale difference sequence for all \( \theta \in \mathcal{N}_\epsilon \) from Proposition 6. In order to apply Corollary 1 to \( l_{\theta_{x_0}} \), we first show

\[
sup_{\theta \in \mathcal{N}_\epsilon} \left| P_n[v_{k,0}(\theta) \nu_{k,0}(\theta)'] - \mathbb{E}_{\theta^*} [v_{k,\infty}(\theta) \nu_{k,\infty}(\theta)'] \right| = o_p(1), \tag{75}
\]

\[
\nu_n(v_{k,0}(\theta)) \Rightarrow W(\theta), \tag{76}
\]

where \( W(\theta) \) is a mean-zero continuous Gaussian process with \( \mathbb{E}_{\theta^*} [W(\theta_1) W(\theta_2)'] = \mathbb{E}_{\theta^*} [v_{k,\infty}(\theta_1) v_{k,\infty}(\theta_2)'] \).

(75) holds because \( P_n[v_{k,0}(\theta) \nu_{k,0}(\theta)' - v_{k,\infty}(\theta) \nu_{k,\infty}(\theta)'] = o_p(1) \) from Proposition 6, and \( v_{k,\infty}(\theta) \nu_{k,\infty}(\theta)' \) satisfies a uniform law of large numbers (Lemma 2.4 and footnote 18 of Newey and McFadden (1994)) because \( v_{k,\infty}(\theta) \) is continuous and \( \mathbb{E}_{\theta^*} \sup_{\theta \in \mathcal{N}} |v_{k,\infty}(\theta) \nu_{k,\infty}(\theta)'| < \infty \) from Proposition 6.

(76) holds because \( \sup_{\theta \in \mathcal{N}_\epsilon} \nu_n(v_{k,0}(\theta) - v_{k,\infty}(\theta)) = o_p(1) \) from Proposition 6 and \( \nu_n(v_{k,\infty}(\theta)) \Rightarrow W(\theta) \) from Theorem 10.2 of Pollard (1990) because (i) the space of \( \theta \) is totally bounded, (ii) the finite dimensional distributions of \( \nu_n(v_{k,\infty}(\cdot)) \) converge to those of \( W(\cdot) \) from a multivariate martingale CLT, and (iii) \( \{\nu_n(v_{k,\infty}(\cdot)) : n \geq 1\} \) is stochastically equicontinuous from Theorem 2 of Hansen (1996) because \( v_{k,\infty}(\theta) \) is Lipschitz continuous in \( \theta \) and both \( v_{k,\infty}(\theta) \) and the Lipschitz coefficient are in \( L^q(\mathbb{P}_{\theta^*}) \) with \( q > \dim(\theta) \) from Proposition 6.

We proceed to show that the terms on the right hand side of (74) satisfies Assumption 3(a)–(g).

Observe that \( t(\psi, \pi) = 0 \) if and only if \( \psi = \psi^* \). First, \( s_{k,0}(\pi) \) satisfies Assumption 3(a)(b)(f)(g) by Proposition 6, (75), (76), and Assumption 6. Second, \( r_{k,0}(\psi, \pi) \) satisfies assumptions (c)(d) from Proposition 6 and (76). Third, \( u_{k,x_0}(\psi, \pi) \) satisfies assumption (e) from Proposition 6(c). Therefore, the stated result follows from Corollary 1(b). \( \square \)

**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 \( \hat{\psi}_\pi := \arg \max_{\psi \in \Theta_0} t_n(\psi, \pi, \xi) \) denote the MLE of \( \psi \), and split \( t(\hat{\psi}_\pi, \pi) \) as \( t(\hat{\psi}_\pi, \pi) = (t'_\eta, t'_\lambda)' \), where we suppress the dependence of \( \hat{t}_\eta \) and \( \hat{t}_\lambda \) on \( \pi \). Define \( G_{\hat{t}_n} := \nu_n(s_{\hat{t}_\lambda}) \).

\[
G_{\hat{t}_n} = \begin{bmatrix} g_{\hat{t}_n} \\ G_{\lambda, \eta} \end{bmatrix}, \quad G_{\lambda, \eta} := G_{\lambda, \eta} - \mathcal{I}_{\lambda, \eta} \mathcal{I}_{\eta}^{-1} G_{\lambda, \eta}, \quad Z_{\lambda, \eta} := \mathcal{I}_{\lambda, \eta}^{-1} G_{\lambda, \eta}, \quad t_{\eta, \lambda} := t_{\eta} + \mathcal{I}_{\eta}^{-1} \mathcal{I}_{\eta, \lambda} t_{\lambda}.
\]
Then, we can write (21) as

$$\sup_{\xi \in \Xi} \sup_{d \in A_n(\xi)} \left| 2\left[\ell_0(\psi, \pi, \xi) - \ell_0(\psi^*, \pi, \xi)\right] - A_n(\sqrt{n}t_{n, \lambda_0}) - B_{gn}(\sqrt{n}t) \right| = o_p(1), \quad (77)$$

where

$$A_n(t_{n, \lambda_0}) = 2t'_{n, \lambda_0}G_{\pi m} - t'_{n, \lambda_0}I_{\lambda_0},$$

$$B_{gn}(t) = 2t'_nG_{\pi \eta} - t'_nI_{\lambda_0} = Z'_{\lambda_0}I_{\lambda_0}(t - Z_{\lambda_0}). \quad (78)$$

Observe that

$$2\left[\ell_0(\hat{\psi}, \pi, \xi) - \ell_0(\hat{\psi}_0)\right] = \max_{t_{n0} \in [0,1]} 2\left[\ell_0(\hat{\psi}_0) - \ell_0(\hat{\psi}_0)\right] + o_p(1) \quad (79)$$

Define \( \hat{t}_\lambda \) by

$$B_{gn}(\sqrt{n}t_{\lambda}) = \max_{t_{\lambda} \in [0,1]} 2\left[\ell_n(\hat{\psi}, \pi, \xi) - \ell_0(\hat{\psi}_0)\right] = B_{gn}(\sqrt{n}t_{\lambda}) + o_p(1).$$

Then, we have

$$2\left[\ell_n(\hat{\psi}, \pi, \xi) - \ell_0(\hat{\psi}_0)\right] = B_{gn}(\sqrt{n}t_{\lambda}) + o_p(1) \quad (80)$$

uniformly in \( \pi \in \Theta_{\pi, *} \) because (i) \( B_{gn}(\sqrt{n}t_{\lambda}) \geq 2\left[\ell_n(\hat{\psi}, \pi, \xi) - \ell_0(\hat{\psi}_0)\right] + o_p(1) \) from the definition of \( \hat{t}_\lambda \) and (79), and (ii) \( 2\left[\ell_n(\hat{\psi}, \pi, \xi) - \ell_0(\hat{\psi}_0)\right] \geq B_{gn}(\sqrt{n}t_{\lambda}) + o_p(1) \) from the definition of \( \hat{\psi}, (77) \), and \( \hat{t}_\lambda = O_p(n^{-1/2}) \).

Finally, the asymptotic distribution of \( \sup B_{gn}(\sqrt{n}t_{\lambda}) \) follows from applying Theorem 1(c) of Andrews (2001) to \( B_{gn}(\sqrt{n}t_{\lambda}) \). First, Assumption 2 of Andrews (2001) holds trivially for \( B_{gn}(\sqrt{n}t_{\lambda}) \). Second, Assumption 3 of Andrews (2001) is satisfied by (76) and Assumption 6. Assumption 4 of Andrews (2001) is satisfied by Proposition 8. Assumption 5* of Andrews (2001) holds with \( B_T = n^{1/2} \) because \( \alpha(1 - \alpha)v(\Theta_{\lambda}) \) is locally equal to the cone \( v(\mathbb{R}^q) \) given that \( \alpha(1 - \alpha) > 0 \) for all \( \alpha \in \Theta_{\alpha, \epsilon} \). Therefore, \( \sup_{\psi \in \Theta_{\pi, \epsilon}} B_{gn}(\sqrt{n}t_{\lambda}) \rightarrow^d \sup_{\psi \in \Theta_{\pi, \epsilon}} (\hat{t}'_\lambda I_{\lambda_0} \hat{\psi}_0) \) follows from Theorem 1(c) of Andrews (2001).

**Proof of Proposition 10.** The proof is similar to that of Proposition 8. Define \( \Lambda^{\psi}_{k,m}(\psi, \pi) \) and \( \Lambda^{\psi}_{k,0}(\psi, \pi) \) as in the proof of Proposition 8. Expanding \( l_{k,\delta x_0} - 1 \) five times around \( \psi^* \) similarly to (73) while fixing \( \pi \) gives, with \( \overline{\psi} \in [\psi, \psi^*] \),

$$l_{k,\delta x_0} - 1 = \sum_{j=1}^{4}(\Delta^j\psi)'\Lambda^{\psi}_{k,0}(\psi^*, \pi) + (\Delta^5\psi)'\Lambda^{\psi}_{k,0}(\overline{\psi}, \pi) + u_{kx_0}(\psi, \pi), \quad (80)$$

where \( u_{kx_0}(\psi, \pi) : = \sum_{j=1}^{4}(\Delta^j\psi)'[\Lambda^{\psi}_{k,0}(\psi^*, \pi) - \Lambda^{\psi}_{k,0}(\psi^*, \pi)] + (\Delta^5\psi)'[\Lambda^{\psi}_{k,0}(\overline{\psi}, \pi) - \Lambda^{\psi}_{k,0}(\overline{\psi}, \pi)]. \)
Define $p_{\psi \pi k,m} := p_{\psi \pi}(Y_{k|Y_{-m}^k})$, and define, for $m \geq 0$,

$$s_{k,m}(\pi) := \begin{pmatrix} \nabla_\eta p_{\psi^* \pi k,m}/p_{\psi^* \pi k,m} \\ \zeta_{k,m}(\varrho)/2 \\ \nabla_\lambda_4 \lambda_4 p_{\psi^* \pi k,m}/(1 - \alpha) p_{\psi^* \pi k,m} \\ \nabla_\lambda_4 \lambda_4 p_{\psi^* \pi k,m}/2(1 - \alpha) p_{\psi^* \pi k,m} \\ \nabla_\lambda_4 \lambda_4 p_{\psi^* \pi k,m}/(1 - \alpha) p_{\psi^* \pi k,m} \\ \nabla_\lambda_4 \lambda_4 p_{\psi^* \pi k,m}/(1 - \alpha) p_{\psi^* \pi k,m} \\ \nabla_\lambda_4 \lambda_4 p_{\psi^* \pi k,m}/(1 - \alpha) p_{\psi^* \pi k,m} \end{pmatrix},$$

where $\nabla_{\psi(\lambda_\varrho)} := (\nabla_{\lambda_{\varrho} \lambda_{\varrho} 1}/2, \ldots, \nabla_{\lambda_{\beta_q} \lambda_{\beta_q} 1}/2, \nabla_{\lambda_{\beta_1} \lambda_{\beta_2} 1}, \ldots, \nabla_{\lambda_{\beta_{q-1} \beta_q}})$. Noting that $\nabla_\lambda p_{\psi^* \pi}(Y_{k|Y_{0}^{k-1}}) = 0$ and $\nabla_\lambda p_{\psi^* \pi}(Y_{k|Y_{0}^{k-1}}) = 0$ from (15) and (16), we may rewrite (80) as, with $t(\psi, \pi)$ defined in (32),

$$l_{\theta k_0} - 1 = t(\psi, \pi)s_{k,0}(\pi) + r_{k,0}(\pi) + u_{k_0}(\psi, \pi),$$

where $r_{k,0}(\pi) := \tau(\psi)^t \tilde{A}_{k,0}(\pi) + (\Delta^5 \psi)^t \Lambda_{k,0}(\psi, \pi) + \alpha_{k,0}^{\pi} p_{\psi^* \pi k,0} - b(\alpha) \nabla_\lambda p_{\psi^* \pi k,0}/4! \alpha_{k,0}(1 - \alpha) p_{\psi^* \pi k,0},$ $\tau(\psi)$ is the vector that collects the elements of $\{\Delta^5 \psi\}_{j=2}^4$ that are not in $t(\psi, \pi)$, and $\tilde{A}_{k,0}(\psi, \pi)$ denotes the vector of the corresponding elements of $\{A_{k,0}^j(\psi^*, \pi)\}_{j=2}^4$.

The stated result follows from Corollary 1 if the terms on the right hand side of (81) satisfy Assumption 3. Similarly to the proof of Proposition 9, define $v_{k,m}(\varrho) := (s_{k,m}(\pi)^t, \tilde{A}_{k,0}(\pi)^t, \Lambda_{k,0}(\psi, \pi)^t)$. Because $\nabla_\lambda p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}}) = 0$ from (30), we can rewrite (31) using the mean value theorem as, for $\varrho \in [0, \varrho]$,

$$\zeta_{k,0}(\varrho) = \frac{\nabla_\lambda \lambda_2 p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}}) - \nabla_\lambda \lambda_2 p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}})}{\alpha(1 - \alpha) p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}})} = \frac{\nabla_\varrho \nabla_\lambda \lambda_2 p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}})}{\alpha(1 - \alpha) p_{\psi^* \pi_k}(Y_{k|Y_{0}^{k-1}})}.$$  

The right hand side is well-defined when $\varrho \to 0$ and satisfies Proposition 6. Therefore, $v_{k,\infty}(\varrho) := \lim_{m \to \infty} v_{k,m}(\varrho)$ is well-defined, and $v_{k,0}(\varrho)$ and $v_{k,\infty}(\varrho)$ satisfy (75)–(76) from repeating the argument in the proof of Proposition 9.

We proceed to show that the terms on the right hand side of (81) satisfy Assumption 3. Observe that $t(\psi, \pi) = 0$ if and only if $\psi = \psi^*$. $s_{k,0}(\pi)$ and $u_{k_0}(\psi, \pi)$ satisfy Assumption 3 from the argument in the proof of Proposition 8 with replacing Assumption 6 with Assumption 7. We show that each component of $r_{k,0}(\pi)$ satisfies Assumption 3(c)(d). First, $(\Delta^5 \psi)^t \tilde{A}_{k,0}(\psi, \pi)$ satisfies Assumption 3(c)(d) from Proposition 6, (76) and $\lambda_{\alpha}^t = (12\lambda_2/b(\alpha))\lambda_2^2 + (12\lambda_2/b(\alpha))\lambda_2^2/12 - 12(\lambda_2/b(\alpha))\lambda_2^2 = O(|\psi||t(\psi, \pi)|)$. Second, $\lambda_{\alpha}^t [\nabla_\lambda \lambda_2 p_{\psi^* \pi_k,0} - b(\alpha) \nabla_\lambda \lambda_2 p_{\psi^* \pi_k,0}/4! \alpha(1 - \alpha) p_{\psi^* \pi_k,0}$ satisfies Assumption 3(c)(d) from Lemma 6(b). Third, for $t(\psi)^t \tilde{A}_{k,0}(\pi)$, observe that $\nabla_\lambda p_{\psi^* \pi_k,0} = 0$ for any $j \geq 1$ in view of (26)–(29). Therefore, $t(\psi)^t \tilde{A}_{k,0}(\pi)$ is written as, with $\hat{\eta} := \eta - \eta^* $,

$$\hat{\eta}^t (\nabla_\eta p_{\psi^* \pi_k,0}/p_{\psi^* \pi_k,0}) + R_{3k0} + R_{4k0},$$

(82)
where, for $j = 3, 4,$

$$R_{jk\varrho} := \frac{1}{\hat{P}_{\psi^\star k,0}(u_1,\ldots,u_d)_{(u_1,\ldots,u_d)\in D(j)}} \sum_{u_1\ldots u_d} \frac{\partial^j \hat{P}_{\psi^\star k,0}(\psi_1^\star u_1^1 \cdots (\psi_d^\star u_d^d)}{\partial \psi_1^1 \cdots \partial \psi_d^d},$$  \tag{83}

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

$$D(3) := \{(u_1,\ldots,u_d) : u_1 + \cdots + u_d = 3\},$$
$$D(4) := \{(u_1,\ldots,u_d) : u_1 + \cdots + u_d = 4, (\psi_1^\star u_1^1 \cdots (\psi_d^\star u_d^d) \neq \lambda_4^4\}.$$

The first term in (82) clearly satisfies Assumption 3(c)(d). The terms in $R_{3k\varrho}$ belong to one of the following three groups: (i) the term associated with $\lambda^3_{\sigma}$, (ii) the term associated with $\lambda^3_{\mu}$; (iii) the other terms. Then, Assumption 3(c)(d) is satisfied by the term (i) because $\lambda^3_{\varrho} = \lambda_{\varrho}[\lambda^2_{\varrho} + b(\beta)\lambda^3_{\mu}]/12$; (ii) the term from Lemma 6(a) because $t(\psi,\pi)$ includes $\varrho \lambda^3_{\varrho}$, and the terms in (iii) because they either contain $\dot{\psi}$ or a term of the form $\lambda^3_{\mu} \lambda^3_{\sigma} \lambda^3_{\beta}$ with $i + j + k = 3$ and $i, j \neq 3$. Similarly, the terms in $R_{4k\varrho}$ satisfy Assumption3(c)(d) because they either contain $\dot{\psi}$ or a term of the form $\lambda^3_{\mu} \lambda^3_{\varrho} \lambda^3_{\beta}$ 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 $(\psi, \varrho)$ for a given $\alpha$. Consider the sets $\Theta^1_\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\psi| \geq n^{-1/8}(\log n)^{-1}\}$ and $\Theta^2_\lambda := \{\lambda \in \Theta_\lambda : |\lambda_\psi| < n^{-1/8}(\log n)^{-1}\}$, so that $\Theta_\lambda = \Theta^1_\lambda \cup \Theta^2_\lambda$. For $j = 1, 2$, define $(\hat{\psi}_j, \hat{\varrho}_j) := \arg\max_{(\psi,\varrho)\in \Theta_{\psi}\times \Theta_{\varrho}, \lambda\in \Theta^j_\lambda} \ell_n(\psi, \varrho, \alpha, \xi).$ Then, uniformly in $\alpha$,

$$\ell_n(\hat{\psi}_\alpha, \hat{\varrho}_\alpha, \alpha, \xi) = \max \left\{ \ell_n(\hat{\psi}_1, \hat{\varrho}_1, \alpha, \xi), \ell_n(\hat{\psi}_2, \hat{\varrho}_2, \alpha, \xi) \right\}.$$ 

Henceforth, we suppress the dependence of $\hat{\psi}_\alpha, \hat{\varrho}_\alpha,$ etc. on $\alpha$.

Define $B_{\varrho n}(t_\lambda(\lambda, \varrho, \alpha))$ as in (78) in the proof of Proposition 9 but using $t(\psi, \pi)$ and $s_{\varrho k}$ defined in (32) and (35) and replacing $t_\lambda$ in (78) with $t_\lambda(\lambda, \varrho, \alpha)$. Observe that the proof of Proposition 9 goes through up to (79) with the current notation and that $G_{\varrho n}$ and $F_\varrho$ are continuous in $\varrho$. Further, $\dot{\varrho}^1 = O_p(n^{-1/4}(\log n)^2)$ because $\dot{\varrho}^1(\hat{\lambda}_1^1)^2 = O_p(n^{-1/2})$ from Proposition 10(a) and $|\hat{\lambda}_1^1| \geq n^{-1/8}(\log n)^{-1}$. Consequently, $B_{\varrho n}(\sqrt{n}t_\lambda(\hat{\lambda}_1, \hat{\varrho}_1, \alpha)) = B_{\varrho n}(\sqrt{n}t_\lambda(\hat{\lambda}_1, \hat{\varrho}_1, \alpha)) + o_p(1)$, and, uniformly in $\alpha$,

$$2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_n(\varrho_0)] = \max \{B_{\varrho n}(\sqrt{n}t_\lambda(\hat{\lambda}_1, \hat{\varrho}_1, \alpha)), B_{\varrho n}(\sqrt{n}t_\lambda(\hat{\lambda}_2, \hat{\varrho}_2, \alpha))\} + o_p(1).$$  \tag{84}

We proceed to construct parameter spaces $\tilde{\Lambda}_{\varrho \varrho}$ and $\hat{\Lambda}_{\alpha \varrho}$ that are locally equal to the cones $\Lambda^3_\lambda$ and $\Lambda^3_{\lambda \varrho}$ defined in (37). Define $c(\alpha) := \alpha(1 - \alpha),$ and denote the elements of $t_\lambda(\hat{\lambda}_1, \hat{\varrho}_1, \alpha)$
corresponding to (33) by

\[ t_\lambda(\hat{\lambda}^1, \hat{\varrho}^1, \alpha) = \begin{pmatrix} \tilde{t}_{\varrho \mu}^2 \\ \tilde{t}_{\varrho \sigma}^2 \\ \tilde{t}_{\varrho}^2 \\ \tilde{t}_{\lambda \rho}^2 \\ \tilde{t}_{\lambda}^2 \\ \tilde{t}_{\mu}^2 \\ \tilde{t}_{\sigma}^2 \end{pmatrix} := \begin{pmatrix} c(\alpha)\hat{\varrho}^1(\hat{\lambda}^1_\mu)^2 \\ c(\alpha)\hat{\lambda}_\mu \hat{\lambda}_\sigma \\ c(\alpha)\hat{\lambda}_\lambda \hat{\lambda}_\sigma \\ c(\alpha)\hat{\lambda}_\lambda \hat{\lambda}_\rho \\ c(\alpha)\hat{\lambda}_\lambda \hat{\lambda}_\rho \\ c(\alpha)\nu(\hat{\lambda}_\beta^2) \end{pmatrix} \].

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

\[ \tilde{t}_{\varrho \mu}^2 = o_p(n^{-1/2}), \quad \tilde{t}_{\varrho \sigma}^2 = o_p(n^{-1/2}), \quad \tilde{t}_{\lambda \rho}^2 = c(\alpha)b(\alpha)(\hat{\lambda}_\mu^1)^4/12 + o_p(n^{-1/2}), \quad \tilde{t}_{\lambda}^2 = c(\alpha)(\hat{\lambda}_\beta^2)^2 + o_p(n^{-1/2}). \quad (85) \]

In view of this, let \( t_\lambda(\lambda, \varrho, \alpha) := (t_{\varrho \mu}, t_{\varrho \sigma}, t_{\lambda \rho}, t_{\lambda}, t_{\varrho}, t_{\lambda \rho}, t_{\lambda})' \in \mathbb{R}^{q_1} \), and consider the following sets:

\[ \tilde{\Lambda}_{\lambda}^{1} := \{t_\lambda(\lambda, \varrho, \alpha) : t_{\varrho \mu} = c(\alpha)\varrho \lambda_\mu^2, t_{\varrho \sigma} = c(\alpha)\lambda_\mu \lambda_\sigma, t_{\lambda \rho} = c(\alpha)\lambda_\mu / 12, t_{\lambda} = c(\alpha)\nu(\lambda) \in \Theta_\lambda \times \Theta_{\varrho} \}, \]

\[ \tilde{\Lambda}_{\alpha \varrho}^{2} := \{t_\lambda(\lambda, \varrho, \alpha) : t_{\varrho \mu} = c(\alpha)\varrho \lambda_\mu^2, t_{\varrho \sigma} = c(\alpha)\lambda_\mu \lambda_\sigma, t_{\lambda \rho} = c(\alpha)\lambda_\mu / 12, t_{\lambda} = c(\alpha)\nu(\lambda) \in \Theta_\alpha \times \Theta_{\varrho} \}. \]

\( \tilde{\Lambda}_{\lambda}^{1} \) is indexed by \( \alpha \) but does not depend on \( \varrho \) because \( B_{\alpha \varrho}(\cdot) \) in (84) does not depend on \( \varrho \), whereas \( \tilde{\Lambda}_{\alpha \varrho}^{2} \) is indexed by both \( \alpha \) and \( \varrho \) because \( \tilde{\varrho}^2 \) depends on \( \varrho^2 \). Define \( (\tilde{\lambda}_{\alpha \varrho}^{1}(\tilde{\lambda}_{\alpha}^{1}, \tilde{\varrho}_{\alpha}^{1}, \alpha)) = \max_{t_\lambda(\lambda, \varrho, \alpha) \in \tilde{\Lambda}_{\lambda}^{1}} B_{\alpha \varrho}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha)) \) and \( (\tilde{\lambda}_{\alpha \varrho}^{2}(\tilde{\lambda}_{\alpha}^{2}, \tilde{\varrho}_{\alpha}^{2}, \alpha)) = \max_{t_\lambda(\lambda, \varrho, \alpha) \in \tilde{\Lambda}_{\alpha \varrho}^{2}} B_{\alpha \varrho}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha)) \).

Define \( W_\alpha := \max_{t_\lambda(\lambda, \varrho, \alpha) \in \tilde{\Lambda}_{\lambda}^{1}} B_{\alpha \varrho}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha)) \), then we have

\[ 2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{\alpha \varrho}(\hat{\theta}_0)] = W_\alpha + o_p(1), \quad (86) \]

uniformly in \( \alpha \in \Theta_\alpha \) because (i) \( W_\alpha \geq 2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{\alpha \varrho}(\hat{\theta}_0)] + o_p(1) \) in view of the definition of \( (\tilde{\lambda}_{\alpha}^{1}, \tilde{\varrho}_{\alpha}^{1}, \tilde{\lambda}_{\alpha \varrho}^{2}) \), (84), and (85), and (ii) with \( \tilde{\lambda}_{\alpha \varrho}^{1} := \arg \max_{\eta} \ell_n(\eta, \tilde{\lambda}_{\alpha}^{1}, \tilde{\varrho}_{\alpha}^{1}, \alpha, \xi) \) and \( \tilde{\lambda}_{\alpha \varrho}^{2} := \arg \max_{\eta} \ell_n(\eta, \tilde{\lambda}_{\alpha \varrho}^{2}, \tilde{\varrho}_{\alpha \varrho}, \alpha, \xi) \), we have \( 2[\ell_n(\hat{\psi}, \hat{\varrho}, \alpha, \xi) - \ell_{\alpha \varrho}(\hat{\theta}_0)] \geq \max\{2[\ell_n(\tilde{\lambda}_{\alpha \varrho}^{1}, \tilde{\varrho}_{\alpha \varrho}, \alpha, \xi)] - 2\ell_{\alpha \varrho}(\hat{\theta}_0) + o_p(1)\} = W_\alpha + o_p(1) \) from the definition of \( (\hat{\psi}, \hat{\varrho}) \).

The asymptotic distribution of the LRTS follows from applying Theorem 1(c) of Andrews (2001) to \( (B_{\alpha \varrho}(\sqrt{n}t_\lambda(\lambda, \varrho, \alpha)), B_{\varrho}(\sqrt{n}t_\lambda(\lambda, \alpha, \xi))) \). First, Assumption 2 of Andrews (2001) holds trivially for \( B_{\varrho}(\sqrt{n}t_\lambda(\lambda, \alpha, \xi)) \). Second, Assumption 3 of Andrews (2001) is satisfied by (76)
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 $\Lambda_1^1$ is locally (in a neighborhood of $\rho = 0, \lambda = 0$) equal to the cone $\Lambda_1^0$ and $\Lambda_2^0$ is locally equal to the cone $\Lambda_2^0$ uniformly in $\rho \in \Theta_\rho$. Consequently, $W_n(\alpha) \xrightarrow{d} \sup_{\rho \in \Theta_\rho} \max \{\mathbb{I}\{\rho = 0\}/\mathcal{I}_{\Lambda_0^0}(\lambda^2_\Lambda), (\mathcal{I}_{\Lambda_\rho^0}/\mathcal{I}_{\Lambda_\rho^0})\} \text{ uniformly in } \alpha$ from Theorem 1(c) of Andrews (2001), and the stated result follows from (86).

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

$$l_{k\theta x_0} - 1 = t(\psi, \pi)^s_{k,0}(\pi) + r_{k,0}(\pi) + u_{kx_0}(\psi, \pi), \quad (87)$$

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

$$s_{k,m}(\pi) := \begin{pmatrix}
\nabla_\eta \bar{p}_{\psi^*_{k,m} / \bar{p}_{\psi^*_{k,m}}} \\
(1/2, (1/2 - 1/2) f_k^\ast) \\
\nabla_\mu f_k^\ast / 3\alpha (1 - \alpha) f_k^\ast \\
\nabla_\mu f_k^\ast / 3\alpha (1 - \alpha) f_k^\ast \\
\nabla_\lambda \lambda_\mu \bar{p}_{\psi^*_{k,m}} / (1 - \alpha) \bar{p}_{\psi^*_{k,m}} \\
\nabla_\psi(\lambda_\beta) \bar{p}_{\psi^*_{k,m}} / (1 - \alpha) \bar{p}_{\psi^*_{k,m}} \\
\n\end{pmatrix},$$

$$r_{k,0}(\pi) := \tau(\psi) ^{\hat{A}}_{k,0}(\pi) + (\Delta^5 \psi) ^{\hat{A}}_{k,0}(\bar{\psi}, \pi), \text{ and } \{t(\psi), \bar{A}_{k,0}(\pi), \Delta^5 \psi, \Lambda^5_{k,0}(\bar{\psi}, \pi), \bar{p}_{\psi^*_{k,m}}, \bar{\nabla}_{v(\lambda_\beta), u_{kx_0}(\psi, \pi)}\}$$

are defined similarly to those in the proof of Proposition 10.

The stated result is proven if the terms on the right hand side of (87) satisfy Assumption 3. $t(\psi, \pi) = 0$ if and only if $\psi = \psi^*$. $s_{k,0}(\psi, \pi)$ and $u_{kx_0}(\psi, \pi)$ satisfy Assumption 3 by the same argument as the proof of Proposition 10. For $r_{k,0}(\pi)$, first, $(\Delta^5 \psi) ^{\hat{A}}_{k,0}(\bar{\psi}, \pi)$ satisfies Assumption 3(c)(d) from a similar argument to the proof of Proposition 10. Second, similar to the proof of Proposition 10, write $\tau(\psi) ^{\hat{A}}_{k,0}(\pi)$ as $\bar{\eta} ^{\hat{A}}_{\psi^*_{k,0}}(\bar{\psi}, \pi)$, and $\hat{R}_{jk\theta}$ is defined as $\hat{R}_{jk\theta}$ in (83) with $D(j)$ replaced with $D(j) := \{(u_1, \ldots, u_d) : u_1 + \cdots + u_d = j, (\psi_1 - \psi_1^*)^\mathbf{u} \cdots (\psi_d - \psi_d^*)^\mathbf{u}^d \neq \lambda_{\mu}\}$ for $j = 3, 4$. The term $\bar{\eta} ^{\hat{A}}_{\psi^*_{k,0}}(\bar{\psi}, \pi)$ clearly satisfies Assumption 3(c)(d). The terms in $\hat{R}_{jk\theta}$ satisfy Assumption 3(c)(d) because they contain either $\hat{\eta}$ or $\lambda_{\mu}^2 \lambda_{\beta}$ or $\mu \lambda_{\beta}^2$ or $\lambda_{\beta}^2$. The terms in $\hat{R}_{jk\theta}$ satisfy assumptions (c)(d) because they either contain $\hat{\eta}$ or a term of the form $\lambda_{\mu}^2 \lambda_{\beta}^{4-i}$ with $1 \leq i \leq 3$. Therefore, $r_{k,0}(\pi)$ satisfies Assumption 3(c)(d), and the stated result is proven.

**Proof of Proposition 13.** The proof is similar to the proof of Proposition 11. Let $(\hat{\psi}, \hat{\rho}, \hat{\alpha}) := \arg \max_{(\psi, \rho, \alpha) \in \Theta_\psi \times \Theta_\rho \times \Theta_\alpha} \ell_n(\psi, \rho, \alpha, \xi)$ denote the MLE of $(\psi, \rho, \alpha)$. Consider the sets $\Theta_1^\lambda := \{\lambda \in \Theta_\lambda : |\lambda_{\mu}| \geq n^{-1/6}(\log n)^{-1}\}$ and $\Theta_2^\lambda := \{\lambda \in \Theta_\lambda : |\lambda_{\mu}| < n^{-1/6}(\log n)^{-1}\}$, so that $\Theta_\lambda = \Theta_1^\lambda \cup \Theta_2^\lambda$. For $j = 1, 2$, define $(\hat{\psi}, \hat{\rho}, \hat{\alpha}) := \arg \max_{(\psi, \rho, \alpha) \in \Theta_\psi \times \Theta_\rho \times \Theta_\lambda} \ell_n(\psi, \rho, \alpha, \xi)$, so that $\ell_n(\hat{\psi}, \hat{\rho}, \hat{\alpha}) = \max_{(\psi, \rho, \alpha) \in \Theta_\psi \times \Theta_\rho \times \Theta_\lambda} \ell_n(\psi, \rho, \alpha, \xi)$. Define $B_{\eta n}(\ell_n(\lambda, \rho, \alpha))$ as in (78) in the proof of Proposition 9 but using $t(\psi, \pi)$ and $s_{k,0}$ defined in
Note that \( \hat{\lambda}^1 \) and \( \hat{\lambda}^2 \) and replacing \( t_\lambda \) in (78) with \( t_\lambda(\lambda, \varrho, \alpha) \). Observe that \( \hat{\varrho}^1 = O_p(n^{-1/6}(\log n)^2) \) because \( \hat{\varrho}^1(\hat{\lambda}_\mu)^2 = O_p(n^{-1/2}) \) from Proposition 12(a) and \( |\hat{\lambda}_\mu| \geq n^{-1/6}(\log n)^{-1} \). Using the argument of the proof of Proposition 11 leading to (84), we obtain

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

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

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

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

\[
\hat{t}_{\nu(\beta)}^1 = o_p(n^{-1/2}), \hat{t}_{\nu(\beta)}^2 = o_p(n^{-1/2}), \hat{t}_{\nu(\beta)}^3 = o_p(n^{-1/2}).
\]

In view of this, let \( t_\lambda(\lambda, \varrho, \alpha) := (t_{\mu^2}, t_{\mu^3}, t_{\mu^4}, t_{\beta\mu}, t_{\nu(\beta)})' \in \mathbb{R}^{6\alpha} \), and consider the following sets:

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

\[
\begin{align*}
\tilde{\Lambda}^2_{\lambda,\varrho} &:= \{t_\lambda(\lambda, \varrho, \alpha) : t_{\mu^2} = c(\alpha)\varrho\lambda^2_\mu, t_{\mu^3} = t_{\mu^4} = t_{\beta\mu} = t_{\nu(\beta)} = 0, \\
t_{\beta\mu} = c(\alpha)\lambda_\beta\lambda_\mu, t_{\nu(\beta)} = c(\alpha)\nu(\beta) \text{ for some } \lambda \in \Theta_\lambda\}
\end{align*}
\]

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

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

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

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 (21) or (36). Therefore, the stated result follows from applying the proof of Propositions 9, 11, and 13 to \( \ell_n(\psi_m, \pi_m, \xi_{M_0+1}) - \ell_n(\hat{\psi}_{M_0}, \xi_{M_0}) \) for each \( h \) and combining the results to derive the joint asymptotic distribution of \( \{ \ell_n(\psi_m, \pi_m, \xi_{M_0+1}) - \ell_n(\hat{\psi}_{M_0}, \xi_{M_0}) \}_{m=1}^{M_0} \). \( \square \)

**Proof of Proposition 18.** The proof of Proposition 8, 10, and 12 shows that the assumption of Corollary 1 holds for these three models under \( P_{n\theta} \). Because \( \vartheta_n = (\vartheta'_{m, \lambda}, \pi_n') \in \mathcal{N}_{c/\sqrt{n}} \) for all \( n \) by choosing \( c > |h| \), we may apply Corollary 1(a) to show that
\[
\left| \log \frac{dP_{n\vartheta}}{dP_{n\vartheta'}} - h'\nu_n(s_{\vartheta_{nk}}) + \frac{1}{2} h'\mathcal{I}_{\vartheta} h \right| \leq \sup_{\vartheta \in \mathcal{N}_{c/\sqrt{n}}} \left| \log \frac{dP_{n\vartheta}}{dP_{n\vartheta'}} - h'\nu_n(s_{\vartheta_{nk}}) + \frac{1}{2} h'\mathcal{I}_{\vartheta} h \right| = o_{P_{n\vartheta}^*}(1),
\]
where \( s_{\vartheta_{nk}} \) is given by (19), (35), and (45) for the models of non-normal distribution, heteroskedastic normal distribution, and homoskedastic normal distribution, respectively. In conjunction with the uniform continuity of \( G_{\vartheta} \) and \( \mathcal{I}_{\vartheta} \) with respect to \( \vartheta \),
\[
\log \frac{dP_{n\vartheta}}{dP_{n\vartheta'}} = h'\nu_n(s_{\vartheta_{nk}}) - \frac{1}{2} h'\mathcal{I}_{\vartheta} h + o_{P_{n\vartheta}^*}(1), \tag{88}
\]
with \( \nu_n(s_{\vartheta_{nk}}) \Rightarrow G_{\vartheta} \), where \( G_{\vartheta} \) is a mean zero Gaussian process with \( \text{cov}(G_{\vartheta_1}, G_{\vartheta_2}) = \mathcal{I}_{\vartheta_1\vartheta_2} := \lim_{k \to \infty} E_{\vartheta_k}(s_{\vartheta_{1k}} s_{\vartheta_{2k}}^\prime) \) under \( P_{n\vartheta}^* \). It follows from (88) that \( \frac{dP_{n\vartheta}}{dP_{n\vartheta'}} \) converges in distribution under \( P_{n\vartheta}^* \) to \( \exp(N(\mu, \sigma^2)) \) with \( \mu = -(1/2)h'\mathcal{I}_{\vartheta} h \) and \( \sigma^2 = h'\mathcal{I}_{\vartheta} h \) so that \( E(\exp(N(\mu, \sigma^2))) = 1 \). Therefore, 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 (88) imply that
\[
\left( \begin{array}{c}
\nu_n(s_{\vartheta_{nk}}) \\
\log \frac{dP_{n\vartheta}}{dP_{n\vartheta'}}
\end{array} \right) \xrightarrow{d} N \left( \begin{array}{c}
0 \\
\frac{1}{2} h'\mathcal{I}_{\vartheta} h
\end{array} \right), \tag{\text{I}_{\vartheta}, \text{I}_{\vartheta}^h \right) \right) \right) \text{ under } P_{n\vartheta}^*.
\]
\( \square \)

**Proof of Proposition 19.** We apply the argument in the proof of Proposition 9 under \( P_{n\vartheta}^* \). By Proposition 8, in view of Theorem 12.3.2 of Lehmann and Romano (2005), (21) holds under \( P_{n\vartheta}^* \), and, therefore, (77) holds under \( P_{n\vartheta}^* \). Given the fixed value of \( \vartheta_n \), Proposition 6 holds by replacing \( P_{\vartheta} \) with \( P_{\vartheta_n} \). Repeating the argument in the proof of Proposition 8 after (76) in conjunction with Proposition 18(b) and Proposition 6 under \( P_{n\vartheta_n} \), we have \( \nu_n(s_{\vartheta_{nk}}) \Rightarrow G_{\vartheta_n} \) under \( P_{n\vartheta_n}^* \), where \( G_{\vartheta_n} \) is a Gaussian process indexed by \( \vartheta \) with \( E_{\vartheta_n}(G_{\vartheta_n}) = \mathcal{I}_{\vartheta_n} h \) and \( \text{cov}_{\vartheta_n}(G_{\vartheta_1}, G_{\vartheta_2}) = \text{cov}_{\vartheta_n}(G_{\vartheta_1}, G_{\vartheta_2}) = \mathcal{I}_{\vartheta_1\vartheta_2} \). This implies that, under \( P_{n\vartheta_n}^* \), \( G_{\lambda, \eta_0} \) is a \( q \)-vector Gaussian process indexed by \( \vartheta \) with \( E_{\vartheta_n}[G_{\lambda, \eta_0}] = \mathcal{I}_{\lambda_\eta_0} \mathcal{I}_{\lambda_\eta_0}^{-1} E_{\vartheta_n}[G_{\eta_0}] = \mathcal{I}_{\lambda_\eta_0} h_{\lambda} \) and \( \text{cov}_{\vartheta_n}(G_{\lambda, \eta_0, 1}, G_{\lambda, \eta_0, 2}) = \text{cov}_{\vartheta_n}(G_{\lambda, \eta_0, 1}, G_{\lambda, \eta_0, 2}) = \mathcal{I}_{\lambda_\eta_0} \).

46
Therefore, repeating the argument in the proof of Proposition 9 under $\mathbb{P}_{\hat{\vartheta}_n}$ gives the stated result.

**Proof of Propositions 20 and 21.** The proof follows the argument in the proof of Propositions 11 and 13. Confirm that $h_\lambda = \sqrt{n}t_\lambda(\lambda_n, \pi_n) + o(1)$ holds under $H_{1n}$ for $\kappa \in [0, 1/4]$ and, therefore, $\mathbb{P}_{\hat{\vartheta}_n}$ and $\mathbb{P}_{\hat{\vartheta}_n^*}$ are mutually contiguous by Proposition 18(a). Then, Propositions 10 and 12 hold in $\mathbb{P}_{\hat{\vartheta}_n}$-probability in view of Theorem 12.3.2 of Lehmann and Romano (2005). The stated result follows from repeating the argument in the proof of Propositions 11 and 13 under $\mathbb{P}_{\hat{\vartheta}_n}$ while noting that $G_{\lambda, \eta_0n} \Rightarrow G_{\lambda, \eta_0}$, where $G_{\lambda, \eta_0}$ is a Gaussian process indexed by $\varrho$ with $\mathbb{E}_{\hat{\vartheta}_n}[G_{\lambda, \eta_0}] = \mathcal{I}_{\lambda, \eta_0}h_{\lambda}$ and $\text{cov}_{\hat{\vartheta}_n}(G_{\lambda, \eta_01}, G_{\lambda, \eta_02}) = \mathcal{I}_{\lambda, \eta_01, \eta_2}$ by Proposition 18(b) and Theorem 10.2 of Pollard (1990).

**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 models of heteroskedastic and homoskedastic normal distribution is similar. The proof follows the argument in the proof of Theorem 15.4.2 in Lehmann and Romano (2005) while referring to the argument in the proof of Theorem 3.2 of Carrasco et al. (2014). By Lemma 8, the distribution of the LRTS under $\mathbb{P}_{\hat{\vartheta}_n}$ is common across all sequences $\eta_n$ such that $\sqrt{n}(\eta_n - \eta^*) \to h_\eta$. Denote the MLE of the $M_0$-regime model parameter by $\hat{\eta}_n$. For the MLE under $H_0$, $\sqrt{n}(\hat{\eta}_n - \eta^*)$ converges in distribution 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 $\tilde{\eta}_n$ defined on a common probability space and a finite constant $\tilde{h}_\eta$ 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, $\mathbb{P}_{\hat{\vartheta}_n}/\mathbb{P}_{\tilde{\vartheta}_n}$ is $\mathcal{C}_q$ with probability one, and the stated result under $H_0$ follows from Lemma 8. For the MLE under $H_{1n}$, note that $\mathbb{P}_{\hat{\vartheta}_n}$ under $H_{1n}$ and $\mathbb{P}_{\hat{\vartheta}_n}$ under $H_0$ are mutually contiguous by Proposition 18(a). Then, by Theorem 12.3.2 of Lehmann and Romano (2005), $\sqrt{n}(\eta_n - \eta^*)$ converges in distribution under $H_{1n}$ because $\sqrt{n}(\tilde{\eta}_n - \eta^*)$ converges in distribution under $H_0$. Repeating the argument as in the case for $H_0$, we have the stated result under $H_{1n}$ from Lemma 8.

10.3 Auxiliary results

10.3.1 Missing information principle

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

47
Lemma 1. For any random variables $X$ and $Y$ with density $P(Y; X; \theta)$ and $P(Y; \theta)$,

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

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 $E[\cdot|Y]$ on the right hand side replaced with $P(Y, X|Z; \theta)$ and $E[\cdot|Y, Z]$. 

Proof of Lemma 1. The stated result follows from a direct calculation and relations such as
\( \nabla_{\theta}^2 P(Y; \vartheta) / P(Y; \vartheta) = \mathbb{E}[\nabla_{\theta}^2 P(Y, X; \vartheta)/P(Y, X; \vartheta)|Y] \) and

\[
\nabla \log f = \nabla f / f, \quad \nabla^2 \log f = \nabla^2 f / f - (\nabla \log f)^2, \\
\nabla^3 \log f = \nabla^3 f / f - 3\nabla^2 f \nabla f / f^2 + 2(\nabla f / f)^3, \\
\nabla^4 \log f = \nabla^4 f / f - 4\nabla^3 f \nabla^2 f / f^2 - 3(\nabla^2 f / f)^2 + 12\nabla^2 f (\nabla f / f)^3 - 6(\nabla f / f)^4, \\
\nabla^5 \log f = \nabla^5 f / f - 5\nabla^4 f \nabla^3 f / f^2 - 10\nabla^3 f \nabla^2 f / f^2 + 20\nabla^2 f (\nabla f / f)^3 / f^3 \\
+ 30(\nabla f / f)^2 \nabla f / f^3 - 60\nabla^2 f (\nabla f / f)^3 / f^4 + 24(\nabla f / f)^5, \\
\nabla^6 \log f = \nabla^6 f / f - 6\nabla^5 f \nabla^4 f / f^2 - 15\nabla^4 f \nabla^3 f / f^2 + 30\nabla^3 f \nabla^2 f / f^3 - 10(\nabla^2 f / f)^2 / f^4 \\
+ 120\nabla^3 f \nabla^2 f \nabla f / f^3 + 120(\nabla^2 f / f)^3 / f^4 + 30(\nabla f / f)^3 / f^3 \\
- 270(\nabla f / f)^2 \nabla f / f^4 + 360\nabla^2 f (\nabla f / f)^3 / f^5 - 120(\nabla f / f)^6 / f^6, \\
\] (89)

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

\[
\nabla^3 \ell(Y) \\
= \frac{\nabla^3 P(Y)}{P(Y)} - 3 \frac{\nabla^2 P(Y) \nabla P(Y)}{P(Y)} + 2 \left( \frac{\nabla P(Y)}{P(Y)} \right)^3 \\
= \mathbb{E} \left[ \frac{\nabla^3 P(Y, X)}{P(Y, X)} | Y \right] - 3 \mathbb{E} \left[ \frac{\nabla^2 P(Y, X)}{P(Y, X)} | Y \right] \mathbb{E} \left[ \frac{\nabla P(Y, X)}{P(Y, X)} | Y \right] + 2 \left\{ \mathbb{E} \left[ \frac{\nabla P(Y, X)}{P(Y, X)} | Y \right] \right\}^3 \\
= \mathbb{E} \left[ \nabla^3 \ell(Y, X) + 3 \nabla^2 \ell(Y, X) \nabla \ell(Y, X) + (\nabla \ell(Y, X))^3 | Y \right] \\
- 3 \mathbb{E} \left[ \nabla^2 \ell(Y, X) + (\nabla \ell(Y, X))^2 | Y \right] \mathbb{E} \left[ \nabla \ell(Y, X) | Y \right] + 2 \left\{ \mathbb{E} \left[ \nabla \ell(Y, X) | Y \right] \right\}^3,
\]

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

10.3.2 Auxiliary Lemmas

Henceforth, the conditioning variable \( W_{m+1} \) is suppressed from the conditioning sets and conditional densities unless confusions might arise.

The following Lemma provides bounds on \( \Phi_{\sigma \tau} \mathcal{F} \) defined in (7) and (61) and is used in the proof of Lemma 3. For \( j = 2, \ldots, 6 \), define \( \| \phi_{\sigma \tau} \mathcal{F} \|_\infty \) as defined in the proof of Lemma 3.
Lemma 2. Under Assumptions 1, 2, and 4, there exists a finite nonstochastic constant $C$ that does not depend on $\rho$ 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 $x \in X$, and $j = 2, \ldots, 6$,

(a) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n]| \leq C \rho^{(t_2-t_1-1)\vee(t_3-t_2-1)\vee\cdots\vee(t_j-t_{j-1}-1)}||\phi_{T(j)}^I||_{\infty}$,

(b) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n, X_{-m} = x]| \leq C \rho^{(t_2-t_1-1)\vee(t_3-t_2-1)\vee\cdots\vee(t_j-t_{j-1}-1)}||\phi_{T(j)}^I||_{\infty}$,

(c) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n, X_{-m} = x] - \Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n]| \leq C \rho^{(m+t_1-1)}||\phi_{T(j)}^I||_{\infty}$,

(d) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n, X_{-m} = x] - \Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m'}^n, X_{-m'} = x]| \leq C \rho^{(m+t_1-1)}||\phi_{T(j)}^I||_{\infty}$,

(e) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n] - \Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^{n-1}]| \leq C \rho^{(n-1-t_j)}||\phi_{T(j)}^I||_{\infty}$,

(f) $|\Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^n, X_{-m} = x] - \Phi_{\vartheta_T(j)}^{I(j)}[\tilde{Y}_{-m}^{n-1}, X_{-m} = x]| \leq C \rho^{(n-1-t_j)}||\phi_{T(j)}^I||_{\infty}$.

Proof of Lemma 2. Recall $\text{sup}_{\vartheta \in \mathcal{N}^*} \text{sup}_{x,x'} |\phi^I(\vartheta, Y_1, x, \tilde{Y}_{t-1}, x') - \mathbb{E}_0[\phi^I(\vartheta, Y_1, x, \tilde{Y}_{t-1}, x')]|$ for the conditioning sets $\mathcal{F}$ that appear in the lemma. Define $\phi_{\vartheta_T(j)}^I : = \phi^I(\vartheta, \tilde{Z}_{t-1}^j) - \mathbb{E}_0[\phi^I(\vartheta, \tilde{Z}_{t-1}^j)|\tilde{Y}_{-m}^n]$, so that $\mathbb{E}_0[\phi_{\vartheta_T(j)}^I|\tilde{Y}_{-m}^n] = \mathbb{E}_0[\tilde{Y}_{\vartheta_T(j)}^n|\tilde{Y}_{-m}^n]$. Henceforth, we suppress the subscript $\vartheta$ from $\phi_{\vartheta_T(j)}^I$ and $\tilde{\phi}_{\vartheta_T(j)}$.

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

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

$$E(X_{t_1} - EX_{t_1}) \cdots (X_{t_j} - EX_{t_j}) = \text{cov}[X_{t_1}, (X_{t_2} - EX_{t_2}) \cdots (X_{t_j} - EX_{t_j})]$$

$$= \text{cov}[(X_{t_1} - EX_{t_1}) \cdots (X_{t_{j-1}} - EX_{t_{j-1}}), X_{t_j}].$$

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

\begin{align}
\{ (a,b,c) \} & \in D_1 \\
\max_{(a,b,c) \in D_1} \| E^c_0 [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}} ] E_0^- [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} ] \| & \leq C \rho^{(t_2-t_1)+}\| \phi^{T(5)}_{\Theta(4)} \|, \\
\max_{(a,b,c) \in D_2} \| E^c_0 [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}} ] E_0^- [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} ] \| & \leq C \rho^{(t_3-t_2)+}\| \phi^{T(6)}_{\Theta(6)} \|, \\
\max_{(a,b,c) \in D_3} \| E^c_0 [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}} ] E_0^- [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} ] \| & \leq C \rho^{(t_4-t_3)+}\| \phi^{T(7)}_{\Theta(7)} \|.
\end{align}

For $j = 4$, parts (a) and (b) follow from combining (91)–(92) and (97) with $k = 2$ because, for example, $\Phi^{1,1,1,1}_{\Theta(4)}[F] \leq C \rho^{(t_2-t_1)+}\| \phi^{T(5)}_{\Theta(4)} \|$ from (91)–(92) and $\Phi^{1,1,1,1}_{\Theta(4)}[F] = \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}] = \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}] = \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}]$ from (92) and (97). For $j = 5$, parts (a) and (b) follow from combining (91), (93) and (97) with $k = 2,3,4$. For $j = 6$, first, $\Phi^{T(6)}_{\Theta(6)}[F]$ is bounded by $C \rho^{(t_2-t_1)+}\| \phi^{T(6)}_{\Theta(6)} \|$ from (91). Second, let $A = E^c_0 [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}} ] = E_0^c [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} ] E_0^c [ \phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} ]$, then all the terms on the right hand side of $\Phi^{T(6)}_{\Theta(6)}[F]$ in (61) except for $A$ are bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$ from (91) and (95), and $A$ is bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$ from (97). Therefore, $\Phi^{T(6)}_{\Theta(6)}[F]$ is bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$. Third, let $B_1 = B_2 = - \sum_{(a,b,c,d), \{e,f\}} \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}]$ and $B_2 = - \sum_{(a,b,c,d), \{e,f\}} \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}]$ the all terms on the right hand side of $\Phi^{T(6)}_{\Theta(6)}[F]$ except for $B_1 + B_2$ are bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$ from (95)–(96). Further, we can write $B_2$ as

\begin{align}
B_2 & = - \sum_{(a,b,c,d), \{e,f\}} \text{cov}_\gamma[\phi_{t_1}^{\ell_{t_1}} \phi_{t_2}^{\ell_{t_2}} \phi_{t_3}^{\ell_{t_3}} \phi_{t_4}^{\ell_{t_4}}] \leq C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|.
\end{align}

Therefore, $\Phi^{T(6)}_{\Theta(6)}[F]$ is bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$. From a similar argument, $\Phi^{T(6)}_{\Theta(6)}[F]$ is also bounded by $C \rho^{(t_4-t_3)+}\| \phi^{T(6)}_{\Theta(6)} \|$, and parts (a) and (b) follow. □
We next present the result that extends Lemmas 13 and 17 of DMR. Let \( r_{\mathcal{I}(1)} = q_i \); \( r_{\mathcal{I}(2)} = q_i / 2 \) if \( i_1 = i_2 \) and \((q_i \land q_{i_2}) / 2 \) if \( i_1 \neq i_2 \); \( r_{\mathcal{I}(3)} = q_i / 3 \) if \( i_1 = i_2 = i_3 \), \((q_i / 2 \land q_{i_2} / 2) \) if \( i_1 \neq i_2 = i_3 \), \((q_i / 3 \land q_{i_2} / 3) \) if \( i_1 \neq i_2 \neq i_3 \) are distinct; \( r_{\mathcal{I}(4)} = q_i / 4 \) if \( i_1 = i_2 = i_3 = i_4 \), \((q_i / 2 \land q_{i_2} / 2 \land q_{i_4} / 2) \) if \( i_1 \neq i_2 \neq i_3 \neq i_4 \); \( r_{\mathcal{I}(5)} = q_i / 5 \) if \( i_1 = i_2 = i_3 = i_4 = i_5 \), \((q_i / 4 \land q_{i_2} / 4) \) if \( i_1 \neq i_2 \neq i_3 \neq i_4 \neq i_5 \).

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

\[
(a) \quad \sup_{x \in X} \sup_{\varnothing \in \mathcal{N}^*} |\sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing)| \leq K_{\mathcal{I}(j)}(k + m)^{7/4}, \quad \mathbb{P}_{\mathcal{I}}\text{-a.s.},
\]

\[
(b) \quad \sup_{x \in X} \sup_{\varnothing \in \mathcal{N}^*} |\sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing)| \leq K_{\mathcal{I}(j)}(k + m)^{7/16}, \quad \mathbb{P}_{\mathcal{I}}\text{-a.s.},
\]

\[
(c) \sup_{x \in X} \sup_{\varnothing \in \mathcal{N}^*} \sup_{m \geq 0} |\sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing)| \leq M_{\mathcal{I}(j)}, \quad \mathbb{P}_{\mathcal{I}}\text{-a.s.}
\]

(d) Uniformly in \( \varnothing \in \mathcal{N}^* \), \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \) and \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \) converge \( \mathbb{P}_{\mathcal{I}}\text{-a.s.} \) and in \( L^{\mathcal{I}(j)}(\mathbb{P}_{\mathcal{I}}) \) to \( \sum_{j,k} \Theta_{j,k}(\varnothing) \) in \( L^{\mathcal{I}(j)}(\mathbb{P}_{\mathcal{I}}) \) as \( m \to \infty \).

**Proof of Lemma 3.** Let \( ||\phi|\|_\infty := \sup_{\varnothing \in \mathcal{N}^*} |\phi(\varnothing,y,x,\varnothing_{t-1},x')| \) for each \( j \). Part (c) follows from parts (a) and (b), \( \sum_{x \in X} \sup_{\varnothing \in \mathcal{N}^*} |\sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing)| \leq M_{\mathcal{I}(j)}, \quad \mathbb{P}_{\mathcal{I}}\text{-a.s.} \) and the distributional equivalence between \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \) and \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \). Part (d) follows from parts (a)-(c) because parts (a)-(c) imply that \{ \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \) \}_{m \geq 0} and \{ \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) \) \}_{m \geq 0} are uniform \( L^{\mathcal{I}(j)}(\mathbb{P}_{\mathcal{I}})\)-Cauchy sequences with respect to \( \varnothing \in \mathcal{N}^* \) that converge to the same limit and \( L^{\mathcal{I}(j)}(\mathbb{P}_{\mathcal{I}}) \) is complete.

We prove parts (a) and (b) by extending the argument in the proof of Lemmas 13 and 17 in DMR. Recall \( \mathcal{T} := (t_1, \ldots, t_2) \). For part (a), define, suppressing the dependence of \( \Theta_{\mathcal{T}}(j) \) on \( \varnothing \) and \( \mathcal{I}(j) \),

\[
A_{\mathcal{T}(j)} := \begin{cases} 
\Phi_{\mathcal{T}(j)} \left[ Y_{-m,k} X_{-m} = x \right] - \Phi_{\mathcal{T}(j)} \left[ Y_{-m,k} X_{-m} = x \right] & \text{if } \max\{t_1, \ldots, t_2\} < k, \\
- \Phi_{\mathcal{T}(j)} \left[ Y_{-m,k} X_{-m} = x \right] + \Phi_{\mathcal{T}(j)} \left[ Y_{-m,k} X_{-m} = x \right], & \text{otherwise},
\end{cases}
\]

\[
A_{\mathcal{T}(j,\ell,k)} := A_{t_1, t_2, \ldots, t_{\ell-1}} \cdot k^{\ell} \cdot k, \quad \text{where } \mathcal{T}(j,\ell,k) := (T(j-\ell,k), \ldots, k).
\]

Then, we can write \( \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) - \sum_{j,k,m,x} \Theta_{j,k,m}(\varnothing) = \sum_{\mathcal{T}(j) \in \{-m+1, \ldots, k\}} A_{\mathcal{T}(j)} = \Delta_a + \Delta_b \), where

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

and \( \sum_{\ell=1}^{j-1} \) is defined as 0 when \( j = 1 \). From Lemma 2 and the symmetry of \( A_{\mathcal{T}(j)} \), \( \Delta_a \) is bounded.
by $CB_{j,k,m}M_{j,k,m}^{(j)}$, where

$$B_{j,k,m} := \sum_{-m+1 \leq t_1 \leq \ldots \leq t_j \leq k-1} \left( \rho^{(m+t_1-1)_+} \wedge \rho^{(t_2-t_1-1)_+} \wedge \ldots \wedge \rho^{(t_j-t_{j-1}-1)_+} \wedge \rho^{(k-1-t_j-1)_+} \right)$$

$$= \sum_{1 \leq t_1 \leq \ldots \leq t_j \leq k+m-1} \left( \rho^{(t_1-1)_+} \wedge \rho^{(t_2-t_1-1)_+} \wedge \ldots \wedge \rho^{(t_j-t_{j-1}-1)_+} \wedge \rho^{(k+m-1-t_j-1)_+} \right),$$

$$M_{j,k,m}^{(j)} := \max_{-m+1 \leq t_1, \ldots, t_j \leq k-1} ||\phi_{t_1}^j||_{\infty}||\phi_{t_2}^j||_{\infty} \cdots ||\phi_{t_j}^j||_{\infty}.$$ 

From $(t-1)_+ \geq [t/2]$ and Lemma 15, $B_{j,k,m}$ is bounded by $C_{j2}(\rho)^{(k+m-1)/4j}$. 

We proceed to derive a bound on $M_{j,k,m}^{(j)}$. Define $||\phi_t^j||_{\infty} := \max_{t=-\infty}^{\infty} (|t| \lor 1)^{-2}||\phi_t^j||_{\infty}$. When $i_1 = i_2 = \cdots = i_j$, from Lemma 16 we have $M_{j,k,m}^{(j)} \leq (k + m)^{j+1}||\phi_t^j||_{\infty}$. In the other cases, from Lemma 16 and the relation $2|xy| \leq x^2 + y^2$, $3|xyz| \leq |x|^3 + |y|^3 + |z|^3$, $|xy_1 y_2 y_3| \leq (|xy_1|^2 + |y_2|^2 + |y_3|^2)/2 \leq (x^4 + y_2^2 y_3^2)/4$, $|xy_1 y_2 y_3 y_4| \leq (|xy_1|^2 + y_2^2 y_3^2 y_4^2)/2 \leq (x^4 + y_2^2 y_3^2 y_4^2)/4$, we can bound $M_{j,k,m}^{(j)}$ by

$$j = 2$$ and $i_1 \neq i_2$:

$$(k + m)^2(||\phi_{i_1}^j||_{\infty}^2 + ||\phi_{i_2}^j||_{\infty}^2),$$

$$j = 3$$ and $i_1 \neq i_2 = i_3$:

$$(k + m)^3(||\phi_{i_1}^j||_{\infty}^3 + ||\phi_{i_2}^j||_{\infty}^3 + ||\phi_{i_3}^j||_{\infty}^3),$$

$$j = 3$$ and $i_1, i_2, i_3$ are distinct:

$$(k + m)^2(||\phi_{i_1}^j||_{\infty}^3 + ||\phi_{i_2}^j||_{\infty}^3 + ||\phi_{i_3}^j||_{\infty}^3),$$

$$j = 4$$ and $i_1 \neq i_2 = i_3 = i_4$:

$$(k + m)^3(||\phi_{i_1}^j||_{\infty}^4 + ||\phi_{i_2}^j||_{\infty}^4 + ||\phi_{i_3}^j||_{\infty}^4),$$

$$j = 4$$ and $i_1 = i_2 \neq i_3 = i_4$:

$$(k + m)^3(||\phi_{i_1}^j||_{\infty}^4 + ||\phi_{i_2}^j||_{\infty}^4 + ||\phi_{i_3}^j||_{\infty}^4),$$

$$j = 5$$ and $i_1 \neq i_2 = i_3 = i_4 = i_5$:

$$(k + m)^3(||\phi_{i_1}^j||_{\infty}^5 + ||\phi_{i_2}^j||_{\infty}^5 + ||\phi_{i_3}^j||_{\infty}^5).$$

Therefore, $\Delta_d$ is bounded by the right hand side of part (a). From Lemmas 2 and 15, $\Delta_b$ is bounded by $C^{(k+m-1)/4(j-1)}M_{j,k+1,m}^{(j)}$, and part (a) of the lemma follows.

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

$$D_{T(j),m'}(x) := \begin{cases}
\Phi_{\theta_{T(j)}}^{(j)}[Y_{-m'}^{k}, X_{-m'} = x] - \Phi_{\theta_{T(j)}}^{(j)}[Y_{-m'}^{k-1}, X_{-m'} = x], & \text{if } \max\{t_1, \ldots, t_j\} < k, \\
\Phi_{\theta_{T(j)}}^{(j)}[Y_{-m'}^{k}, X_{-m'} = x], & \text{otherwise}.
\end{cases}$$

and define $D_{T(j),m}(x)$ similarly. Then, we can write $\Delta_{j,k,m}(\theta) = \sum_{T(j) \in \{-m+1, \ldots, k\}} D_{T(j),m}(x)$ and $\Delta_{j,k,m'}(\theta) = \sum_{T(j) \in \{-m'+1, \ldots, k\}} D_{T(j),m'}(x)$.

$$\Delta_d := \sum_{t_1=-m'+1}^{k} \sum_{t_2=-m'+1}^{k} \cdots \sum_{t_j=-m'+1}^{k} D_{T(j),m',x}.$$ 

By the same argument as part (a), $\Delta_{j,k,m}(\theta) - \Delta_c$ is bounded by the right hand of part (a). For
Furthermore, these bounds hold uniformly in $\Delta_d$, observe that, with $M_j := \max_{1 \leq j \leq j(\ell)}$,

$$|\Delta_d| \leq M_j \sum_{\ell = 1}^j \sum_{t_1 = -m' + 1}^{-m} \sum_{t_2 = -m' + 1}^{-m} \cdots \sum_{t_{\ell-1} = -m' + 1}^{-m} \sum_{t_{\ell} = -m' + 1}^{-m} \sum_{t_{\ell+1} = -m + 1}^{k} \cdots \sum_{t_j = -m + 1}^{k} |D_{T(j),m',x}|$$

$$\leq j M_j \sum_{t_1 = -m' + 1}^{-m} \sum_{t_2 = -m' + 1}^{-m} \cdots \sum_{t_j = -m' + 1}^{-m} |D_{T(j),m',x}|$$

$$\leq j M_j j! \sum_{t_1 = -m + 1}^{-m + 1} \sum_{t_2 = t_1}^{-m + 1} \sum_{t_j = t_{j-1}}^{-m + 1} |D_{T(j),m',x}|.$$
Lemma 5. Let $f(\mu, \sigma^2)$ denote the density of $N(\mu, \sigma^2)$. Then

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

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

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

and hence the stated result follows.

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

$$\begin{align*}
(a) & \quad \frac{\nabla \lambda_\vartheta \overline{p}_{\psi^*}(Y_k|Y_0^{k-1})}{\overline{p}_{\psi^*}(Y_k|Y_0^{k-1})} = \varrho K(\vartheta), \\
(b) & \quad \frac{\nabla \lambda_\vartheta \overline{p}_{\psi^*}(Y_k|Y_0^{k-1})}{\overline{p}_{\psi^*}(Y_k|Y_0^{k-1})} = b(\alpha) \frac{\nabla \lambda_\vartheta \overline{p}_{\psi^*}(Y_k|Y_0^{k-1})}{\overline{p}_{\psi^*}(Y_k|Y_0^{k-1})} + \varrho K(\vartheta). \end{align*}$$

Proof of Lemma 6. Part (a) holds if

$$\nabla \lambda_\vartheta \overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1})/\overline{p}_{\psi^*}(Y_k|Y_0^{k-1}) = 0,$$

(98)

because (i) $\nabla \lambda_\vartheta \overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1}) - \nabla \lambda_\vartheta \overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1}) = \nabla \varrho \nabla \lambda_\vartheta \overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1}) \varrho$ for $\varrho \in [0, \vartheta]$ from the mean value theorem, (ii) $\overline{p}_{\psi^*}(Y_k|Y_0^{k-1})$ does not depend on the value of $\varrho$, and (iii) $\nabla \varrho \nabla \lambda_\vartheta \overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1})/\overline{p}_{\psi^* \alpha}(Y_k|Y_0^{k-1}) \in L^2(\mathbb{P}_{\psi^*})$ uniformly in $\vartheta \in \mathcal{A}^*$ from Proposition 6(d).
We proceed to show (98). Let \( \nabla^4 \ell_t^* := \nabla_{\lambda^3} \log g_t^* \) with \( \nabla \ell_t^* = \nabla^1 \ell_t^* \). Observe that

\[
\nabla_{\lambda^3} \log \bar{p}_{\psi^* \alpha^0}(Y_0^k | \bar{Y}_0) = \sum_{t=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^3 \ell_t^* \bigg| \bar{Y}_0^k \right] + 3 \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^2 \ell_{t_1}^* \nabla \ell_{t_2}^* \bigg| \bar{Y}_0^k \right] \\
+ \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \sum_{t_3=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \bigg| \bar{Y}_0^k \right] = \sum_{t=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla_{\lambda^3} g_t^*/g_t^* \bigg| \bar{Y}_0^k \right] = 0,
\]

where the first equality follows from Lemma 1, the second equality holds because (i) \( X_t \) is serially independent when \( \varrho = 0 \), (ii) \( \nabla \ell_t^* = d_{1t} \nabla \mu f_t^*/f_t^* \) and \( \nabla^2 \ell_t^* = d_{2t} \nabla \mu f_t^*/f_t^* - \left( d_{1t} \nabla \mu f_t^*/f_t^* \right)^2 \), and (iii) \( \mathbb{E}_{\psi^* \alpha^0}[d_{1t} | \bar{Y}_0] = \mathbb{E}_{\psi^* \alpha^0}[d_{2t} | \bar{Y}_0] = 0 \) from (29), the third equality follows from (89), and the last equality follows from (29). (98) follows from (99) because (i) \( \nabla_{\lambda^3} \bar{p}_{\psi^* \pi}(Y_k | \bar{Y}_0^{k-1}) / \bar{p}_{\psi^* \pi}(Y_k | \bar{Y}_0^{k-1}) = \nabla_{\lambda^3} \log \bar{p}_{\psi^* \pi}(Y_0^k | \bar{Y}_0) - \nabla_{\lambda^3} \log \bar{p}_{\psi^* \pi}(Y_0^{k-1} | \bar{Y}_0) \) from (89) and \( \nabla_{\lambda} \bar{p}_{\psi^* \pi}(Y_k | \bar{Y}_0^{k-1}) = 0 \). Therefore, part (a) is proven.

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

\[
\nabla_{\lambda^3} \bar{p}_{\psi^* \alpha^0}(Y_k | \bar{Y}_0^{k-1}) = b(\alpha) \nabla_{\lambda^3} \bar{p}_{\psi^* \alpha^0}(Y_k | \bar{Y}_0^{k-1}) / \bar{p}_{\psi^* \pi}(Y_k | \bar{Y}_0^{k-1}).
\]

Using Lemma 1, noting that \( X_t \) is serially independent when \( \varrho = 0 \), collecting the terms, and using (89), we obtain

\[
\nabla_{\lambda^3} \log \bar{p}_{\psi^* \alpha^0}(Y_0^k | \bar{Y}_0) = \sum_{t=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^4 \ell_t^* \bigg| \bar{Y}_0^k \right] + 4 \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^3 \ell_{t_1}^* \nabla \ell_{t_2}^* \bigg| \bar{Y}_0^k \right] + 3 \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^2 \ell_{t_1}^* \nabla^2 \ell_{t_2}^* \bigg| \bar{Y}_0^k \right] \\
+ 6 \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \sum_{t_3=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ (\nabla^2 \ell_{t_1}^* - \mathbb{E}[\nabla^2 \ell_{t_1}^* | \bar{Y}_0]) \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \bigg| \bar{Y}_0^k \right] \\
+ \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \sum_{t_3=1}^{k} \sum_{t_4=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \nabla \ell_{t_4}^* \bigg| \bar{Y}_0^k \right] - 3 \left( \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \bigg| \bar{Y}_0^k \right] \right)^2 \\
- 6 \sum_{t_1=1}^{k} \sum_{t_2=1}^{k} \sum_{t_3=1}^{k} \sum_{t_4=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla \ell_{t_1}^* \nabla \ell_{t_2}^* \nabla \ell_{t_3}^* \nabla \ell_{t_4}^* \bigg| \bar{Y}_0^k \right] \\
\sum_{t=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla^4 \ell_t^* + 4 \nabla^3 \ell_t^* \nabla \ell_t^* + \left( 3 \nabla \ell_t^* \right)^2 + 6 \nabla^2 \ell_t^* \nabla \ell_t^* \nabla \ell_t^* + \left( \nabla \ell_t^* \right)^4 \bigg| \bar{Y}_0^k \right] = \sum_{t=1}^{k} \mathbb{E}_{\psi^* \alpha^0} \left[ \nabla_{\lambda^3} g_t^*/g_t^* \bigg| \bar{Y}_0^k \right].
\]
Repeating the argument in the proof of Proposition 19, we can establish that

\[ \nabla \log p_{\psi^*}(y) = \nabla \log p_{\psi^*}(y) - \nabla \log p_{\psi^*}(\tilde{y}) \]

from (89), \( \nabla \log p_{\psi^*}(\tilde{y}) = 0 \), and \( \nabla \log p_{\psi^*}(y) = 0 \). This is because \( \psi^* \) is a common continuous limit law given by \( E_{\psi^*}[\nabla \mu^* g^*_t/|\tilde{y}_t^k|] = b(\alpha)E_{\psi^*}[\nabla \mu^* g^*_t/|\tilde{y}_t^k|] \) from (29). Therefore, part (b) is proven.

\[ \square \]

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

\[ (a) \quad \frac{\nabla \log p_{\psi^*}(\tilde{y})}{\nabla \log p_{\psi^*}(\tilde{y})} = (1 - \alpha)(1 - 2\alpha) \frac{\nabla \mu^*_k}{f^*_k} + \rho K(\theta), \]

\[ (b) \quad \frac{\nabla \log p_{\psi^*}(\tilde{y})}{\nabla \log p_{\psi^*}(\tilde{y})} = (1 - \alpha)(1 - 6\alpha + 6\alpha^2) \frac{\nabla \mu^*_k}{f^*_k} + \rho K(\theta). \]

**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

\[ (A) \quad \nabla \log p_{\psi^*}(\tilde{y})/\nabla \log p_{\psi^*}(\tilde{y}) = (1 - \alpha)(1 - 2\alpha) \frac{\nabla \mu^*_k}{f^*_k}, \]

\[ (B) \quad \nabla \log p_{\psi^*}(\tilde{y})/\nabla \log p_{\psi^*}(\tilde{y}) = (1 - \alpha)(1 - 6\alpha + 6\alpha^2) \frac{\nabla \mu^*_k}{f^*_k}. \]

Observe that \( \nabla \log p_{\psi^*}(\tilde{y}) = \sum_{t=1}^k E_{\psi^*}[\nabla \mu^*_k g^*_t/|\tilde{y}_t^k|] \) in (99) and the equality (101) in the proof of Lemma 6 still hold under the assumptions of Proposition 12 if we use (43) in place of (29). Consequently, (A) and (B) follow from (42), (43), and the argument of the proof of Lemma 6, giving the stated result.

\[ \square \]

**Lemma 8.** Suppose that the assumptions of Propositions 15 hold. Let \( C_\eta \) be a set of sequences \( \{\mathbb{P}_{n_\eta}\} \) with \( \eta_n = \eta_n + h_n/\sqrt{n} \) for all bounded \( h_n \), where \( \mathbb{P}_{n_\eta} := \mathbb{P}_{\eta_\theta} = \prod_{k=1}^n f_k(\eta, 0) \) is the probability measure under \( \eta_n \) with \( \lambda_n = 0 \). Then, for every sequence \( \{\mathbb{P}_{n_\eta}\} \) in \( C_\eta \), the LRTS converges in distribution to a common continuous limit law given by \( \sup_{\theta \in \Theta_\eta} (\tilde{t}_{L_\eta} L_{\eta_n} \tilde{t}_{L_\eta}) \) in Propositions 15.

**Proof of Lemma 8.** Repeating the argument in the proof of Proposition 19, we can establish that the stochastic process \( \nu_n(s_{ik}) \) converges weakly as \( n \to \infty \) under \( \mathbb{P}_{n_\eta}^{\eta_n + h_n/\sqrt{n}} \). By continuous mapping theorem, this implies that \( \tilde{t}_{L_\eta} L_{\eta_n} \tilde{t}_{L_\eta} := \max_{\lambda_\theta \in \Theta_\eta} \{2t_{\lambda_\theta} \tilde{t}_{L_\eta} t_{\lambda_\theta} - t_{\lambda_\theta} \tilde{t}_{L_\eta} t_{\lambda_\theta} \} \) converges weakly to a tight random element as \( n \to \infty \) under \( \mathbb{P}_{n_\eta}^{\eta_n + h_n/\sqrt{n}} \). Then, it follows from Theorem 18.14 of van der Vaart (1998) that, for every \( \epsilon > 0 \), there exists finite \( N(\epsilon) \) and \( \{\varrho_1, \ldots, \varrho_N(\epsilon)\} \) such that

\[ \limsup_{n \to \infty} \mathbb{P}_{n_\eta}^{\eta_n + h_n/\sqrt{n}} \left( \sup_{\theta \in \Theta_\eta} \left| \tilde{t}_{L_\eta} L_{\eta_n} \tilde{t}_{L_\eta} - \max_{1 \leq t \leq N(\epsilon)} \left( \tilde{t}_{L_\eta} L_{\eta_n} \tilde{t}_{L_\eta} \right) \right| > \epsilon \right) < \epsilon. \]

Therefore, it suffices to show that, for any \( \{\varrho_1, \ldots, \varrho_N(\epsilon)\} \) finite \( N \), the joint distributions of the finite dimensional vectors \( (\tilde{t}_{L_\eta} L_{\eta_n} \tilde{t}_{L_\eta} : 1 \leq i \leq N) \) are asymptotically the same distribution for
all \( \mathbb{P}_{\eta^*+h_\eta/\sqrt{n}} \) with bounded \( h_\eta \). From (88) with \( \vartheta_n = (\eta_n, \lambda_n, \pi_n) = (\eta^* + h_\eta/\sqrt{n}, 0, \pi) \), we have

\[
\log \frac{d\mathbb{P}_{\eta^*+h_\eta/\sqrt{n}}}{d\mathbb{P}_{\eta^*}} = h_\eta' \nu(s_m) - \frac{1}{2} h_\eta' \mathcal{I}_n h_\eta + o_{\mathbb{P}_{\eta^*}}(1) \quad \text{under } \mathbb{P}_{\eta^*}.
\]  

(103)

Define \( s_{\lambda \eta \rho k} := s_{\lambda k} - \mathcal{I}_{\lambda} \eta \mathcal{I}_{\eta}^{-1} \mathcal{I}_{\rho} \) so that \( G_{\lambda \eta \rho n} = \nu(s_{\lambda \eta \rho k}) \). Then, because \( \lim_{k \to \infty} \mathbb{E}_{\mathbb{P}_{\eta^*}}(s_{\lambda \eta \rho k} s_{\eta k}) = 0 \) for any \( \varrho \in \Theta_{\eta^*} \), we have from (103) that \( (G_{\lambda \eta \rho n_1}, ..., G_{\lambda \eta \rho n_N}) \) under \( \mathbb{P}_{\eta^*} \) converges in distribution under \( \mathbb{P}_{\eta^*} \) to a Gaussian distribution with mean 0 and covariance \( \left( \begin{array}{c} \Sigma \\ 0 \\ 0 \end{array} \right) \), where \( \Sigma \) is the asymptotic covariance of \( (G_{\lambda \eta \rho n_1}, ..., G_{\lambda \eta \rho n_N}) \). It follows from Le Cam’s third lemma that the joint distribution of \( (G_{\lambda \eta \rho n_1}, ..., G_{\lambda \eta \rho n_N}) \) under \( \mathbb{P}_{\eta^*+h_\eta/\sqrt{n}} \) is the same as that of \( (G_{\lambda \eta \rho n_1}, ..., G_{\lambda \eta \rho n_N}) \) under \( \mathbb{P}_{\eta^*} \), and, therefore, by continuous mapping theorem, the joint distributions of \( (t_{\lambda i} \mathcal{I}_{\lambda}, \eta \mathcal{I}_{\rho}, \tilde{t}_{\lambda i}; 1 \leq i \leq N) \) are asymptotically the same distribution for all \( \mathbb{P}_{\eta^*+h_\eta/\sqrt{n}} \) with bounded \( h_\eta \). \( \square \)

10.3.3 Bounds on difference in state probabilities and conditional moments

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

**Proof of Lemma 9.** For any \( \epsilon > 0 \), we have

\[
\mathbb{P}(\max_{1 \leq i \leq n} |X_i| > \epsilon n^{1/q}) \leq \sum_{1 \leq i \leq n} \mathbb{P}(|X_i| > \epsilon n^{1/q}) \leq \epsilon^{-q} n^{-1} \sum_{1 \leq i \leq n} \mathbb{E}(|X_i|^q \mathbb{I}\{|X_i| > \epsilon n^{1/q}\})
\]

by a version of Markov inequality. As \( n \to \infty \), the right hand side tends to 0 by the Dominated Convergence Theorem. \( \square \)

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

In the following, \( x_m \) denotes “\( X_m = x_m \).”

**Lemma 10.** Suppose Assumptions 1-2 hold and \( \vartheta_x \in \Theta_x \). For all \( -m \leq t_1 \leq t_2 \) with \( -m < n \) and all probability measures \( \mu_1 \) and \( \mu_2 \) on \( \mathcal{B}(\mathcal{X}) \),

\[
\left\| \sum_{x_m \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in [\overline{Y}_{-m}^n, x_m, W_{m-1}^n]) \mu_1(x_m) - \sum_{x_m \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1}^{t_2} \in [\overline{Y}_{-m}^n, x_m, W_{m-1}^n]) \mu_2(x_m) \right\|_{TV} \\
\leq \rho^{t_1+1+m}.
\]

**Proof of Lemma 10.** We assume \( t_1 > -m \) because the stated result holds trivially when \( t_1 = -m \). We prove the lemma by showing that the stated bound holds for \( X_{t_1}^{t_2} \in A \) for any \( A \subseteq \mathcal{X}^{t_2-t_1} \). Observe that Lemma 1 of DMR still holds when \( W_{m-1}^n \) is added to the conditioning variable because Assumption 1 implies that \( \{(X_k, Y_k)\}_{k=0}^\infty \) is a Markov chain given \( \{W_k\}_{k=0}^\infty \). Therefore,
\{X_t\}_{t \geq -m} is a Markov chain when conditioned on \{\overline{Y}^n_m, W^n_{-m+1}\}. Choose \(B \in \mathcal{B}(\mathcal{X})\) so that 
\[\mathbb{P}_{\vartheta_x}(X_{t_1} \in B|\overline{Y}^n_m, W^n_{-m+1}) \neq \{0, 1\}\], then it follows from the Markov property of \(\{X_t\}_{t \geq -m}\) that 
\[
\sum_{x_{-m} \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|\overline{Y}^n_m, x_{-m}, W^n_{-m+1})\mu_1(x_{-m}) - \sum_{x_{-m} \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|\overline{Y}^n_m, x_{-m}, W^n_{-m+1})\mu_2(x_{-m}) 
= \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|X_{t_1} \in B, \overline{Y}^n_m, W^n_{-m+1}) \times \left[ \sum_{x_{-m} \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1} \in B|\overline{Y}^n_m, x_{-m}, W^n_{-m+1})\mu_1(x_{-m}) 
- \sum_{x_{-m} \in \mathcal{X}} \mathbb{P}_{\vartheta_x}(X_{t_1} \in B|\overline{Y}^n_m, x_{-m}, W^n_{-m+1})\mu_2(x_{-m}) \right].
\]
The stated result follows because \(\sup_{m} \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|\cdot) \leq 1\) and the supremum of the term inside the brackets over \(B \in \mathcal{B}(\mathcal{X})\) is bounded by \(\rho^{n+1+m}\) from Corollary 1 of DMR.

**Lemma 11.** Suppose Assumptions 1-2 hold and \(\vartheta \in \Theta_x\). For all \(-m \leq t_1 \leq t_2 \leq n - 1\),
\[
\left\| \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|\overline{Y}^n_m, x_{-m}, W^n_{-m+1}) - \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in \cdot|\overline{Y}^{n-1}_m, x_{-m}, W^n_{-m+1}) \right\|_{TV} \leq \rho^{n-1-t_2}.
\]The same bound holds when \(x_{-m}\) is dropped from the conditioning variables.

**Proof of Lemma 11.** We show that the stated bound holds for any \(A \subseteq \mathcal{X}^{t_2-t_1-1}\). Observe that the time-reversed process \(\{Z_{n-k}\}_{0 \leq k \leq n-m}\) is Markov when conditioned on \(W^n_{-m+1}\). Consequently, for \(k = n, n - 1\), \(\mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|X_{t_2}, \overline{Y}^n_m, x_{-m}, W^n_{-m+1}) = \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|X_{t_2}, \overline{Y}^{n-1}_m, x_{-m}, W^n_{-m+1})\) holds. Choose \(B \in \mathcal{B}(\mathcal{X})\) so that \(\mathbb{P}_{\vartheta_x}(X_{t_2} \in B|\overline{Y}^n_m, W^n_{-m+1}) \neq \{0, 1\}\), then it follows that 
\[
\mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|\overline{Y}^n_m, x_{-m}, W^n_{-m+1}) - \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in \cdot|\overline{Y}^{n-1}_m, x_{-m}, W^n_{-m+1}) 
= \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in A|X_{t_2} \in B, \overline{Y}^n_m, x_{-m}, W^n_{-m+1}) \times \left[ \mathbb{P}_{\vartheta_x}(X_{t_1} \in B|\overline{Y}^n_m, x_{-m}, W^n_{-m+1}) - \mathbb{P}_{\vartheta_x}(X_{t_1} \in B|\overline{Y}^{n-1}_m, x_{-m}, W^n_{-m+1}) \right].
\]

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

**Lemma 12.** Suppose Assumptions 1-2 hold and \(\vartheta \in \Theta_x\). For all \(-m \leq t_1 \leq t_2 < t_3 \leq t_4\) with 
\(-m < n\),
\[
\left\| \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in \cdot, X^{t_4}_{t_3} \in \cdot|\overline{Y}^n_m, x_{-m}, W^n_{-m+1}) - \mathbb{P}_{\vartheta_x}(X^{t_2}_{t_1} \in \cdot|\overline{Y}^{n-1}_m, x_{-m}, W^n_{-m+1}) \mathbb{P}_{\vartheta_x}(X^{t_4}_{t_3} \in \cdot|\overline{Y}^n_m, x_{-m}, W^n_{-m+1}) \right\|_{TV} \leq \rho^{t_3-t_2}.
\]

59
The same bound holds when $x_m$ is dropped from the conditioning variables.

**Proof of Lemma 12.** We suppress the conditioning variable $W_{m+1}^n$. We show that the stated bound holds for any $(A, B) \subseteq \mathcal{X}^{t_4-t_1+1}$. Observe that

\[
\mathbb{P}_{\varrho_x}(X_{t_1}^{t_4} \in A, X_{t_3}^{t_4} \in B|Y_{-m}^n, x_m) - \mathbb{P}_{\varrho_x}(X_{t_1}^{t_4} \in A|Y_{-m}^n, x_m)\mathbb{P}_{\varrho_x}(X_{t_3}^{t_4} \in B|Y_{-m}^n, x_m)
= \mathbb{P}_{\varrho_x}(X_{t_1}^{t_4} \in A|Y_{-m}^n, x_m)\mathbb{P}_{\varrho_x}(X_{t_3}^{t_4} \in B|Y_{-m}^n, x_m) - \mathbb{P}_{\varrho_x}(X_{t_1}^{t_4} \in B|Y_{-m}^n, x_m)
\]

The term inside the brackets is bounded by $\rho^{t_3-t_2}$ from Lemma 10. The stated bound then follows immediately.

The following corollaries are used for proving Lemma 2. Their proofs are omitted because they follow straightforwardly from the Lemma 10-12 and the fact that, for any two probability measures $\mu_1$ and $\mu_2$, $\mathbb{P}_{\varrho_x}(x; \varrho_x(x)) = \int f(x) d\mu_1(x) - \int f(x) d\mu_2(x) = 2||\mu_1 - \mu_2||_{TV}$ (see, e.g., Levin et al. (2009, Proposition 4.5)).

**Corollary 2.** Suppose that Assumptions 1-2 hold, $f(Y_{-m}^n, X_{t_1}^{t_4}, W_{m+1}^n; \varrho) \in L^2(\mathbb{P}_{\varrho_x})$, and $g(Y_{-m}^n, X_{t_1}^{t_4}, W_{m+1}^n; \varrho) \in L^2(\mathbb{P}_{\varrho_x})$. Define $||f||_\infty := \sup_{\varrho \in \mathbb{N}^+} \max_{t_1 \leq t_2} f_\varrho(Y_{-m}^n, X_{t_1}^{t_4}, W_{m+1}^n)$, and define $||g||_\infty$ similarly. Let $f_m^n(x_{t_1}^{t_4}; \varrho) := f(Y_{-m}^n, X_{t_1}^{t_4}, W_{m+1}^n; \varrho)$ and $g_m^n(x_{t_1}^{t_4}; \varrho) := g(Y_{-m}^n, X_{t_1}^{t_4}, W_{m+1}^n; \varrho)$. Then, suppressing the conditioning variables $W_{m+1}^n$ and $W_{m+1}^n$ from the conditioning sets,

(a) For all $-m \leq t_1 \leq t_2 < t_3 \leq t_4$,

\[
|\text{cov}_{\varrho_x}(f_m^n(X_{t_1}^{t_4}; \varrho), g_m^n(X_{t_1}^{t_4}; \varrho)|Y_{-m}^n, x_m)| \leq 2\rho^{t_3-t_2}||f_m^n||_\infty ||g_m^n||_\infty,
\]

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

\[
|\mathbb{E}_{\varrho_x}[f_m^n(X_{t_1}^{t_4}; \varrho)|Y_{-m}^n, x_m] - \mathbb{E}_{\varrho_x}[f_m^n(X_{t_1}^{t_4}; \varrho)|Y_{-m}^n] | \leq 2\rho^{m+t_1}||f_m^n||_\infty,
\]

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

\[
|\mathbb{E}_{\varrho_x}[f_m^n(X_{t_1}^{t_4}; \varrho)|Y_{-m}^n, x_m] - \mathbb{E}_{\varrho_x}[f_m^n(X_{t_1}^{t_4}; \varrho)|Y_{-m}^{n-1}, x_m] | \leq 2\rho^{n-1-t_2}||f_m^n||_\infty.
\]

**Lemma 13.** Suppose $X_k$ is a stationary Markov process with the state space $\mathcal{X} = \{1, 2\}$ and the transition matrix

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

Let $q_k := \mathbb{I}\{X_k = 1\}$, $\alpha := \mathbb{E}[q_k]$, and $\varrho := p_{11} + p_{22} - 1$. Then, (a) $\mathbb{E}(q_k - \alpha)^2 = \alpha(1 - \alpha)$, $\mathbb{E}(q_k - \alpha)^3 = \alpha(1 - \alpha)(1 - 2\alpha)$, $\mathbb{E}(q_k - \alpha)^4 = \alpha(1 - \alpha)(3\alpha^2 - 3\alpha + 1)$, and $\text{corr}(X_k, X_{k+\ell}) = \varrho^\ell$. 

60
(b) For all \( k > m \) and any probability measures \( \mu_1 \) and \( \mu_2 \) on \( \mathcal{B}(\mathcal{X}) \),
\[
\left\| \sum_{j \in \mathcal{X}} \mathbb{P}(X_k \in \cdot | X_m = j) \mu_1(j) - \sum_{j \in \mathcal{X}} \mathbb{P}(X_k \in \cdot | X_m = j) \mu_2(j) \right\|_{TV} \leq g^{k-m}.
\]

(c) For all \( t_1 \leq t_2 \leq \cdots \leq t_k < t_{k+1} \leq \cdots \leq t_{\ell} \), \( \| \mathbb{P}(X_{t_k}^t \in \cdot, X_{t_{k+1}}^{t_0} \in \cdot) - \mathbb{P}(X_{t_k}^t \in \cdot) \mathbb{P}(X_{t_{k+1}}^{t_0} \in \cdot) \|_{TV} \leq C g^{k+1-t_k} \).

Proof of Lemma 13. For part (a), the first three results follow from the property of a Bernoulli random variable, and the last result holds because \( X_k = 2 - q_k \) and Hamilton (1994, p. 684) shows that \( q_k \) follows an AR(1) process with the autoregressive coefficient \( p_{11} - p_{22} - 1 \). For part (b), decompose \( P \) as
\[
P = U\Omega U^{-1}, \quad U = (t \ R) = \begin{pmatrix} 1 & \frac{1-p_{11}}{1-p_{11}-p_{22}} \\ 1 & \frac{1-p_{22}}{1-p_{11}-p_{22}} \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \pi \\ u' \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 & 0 \\ 0 & \varrho \end{pmatrix},
\]
where \( \pi := (\mathbb{P}(X_k = 1), \mathbb{P}(X_k = 2)) = (1-p_{22}, 1-p_{11})/(2-p_{11}-p_{22}) \) is a vector of the stationary distribution of \( X_k \), and \( u := (1,-1)' \) is an eigenvector of \( P \) associated with the eigenvalue \( \varrho \). Let \( v_1 = (\mathbb{P}(X_m = 1|\mu_1), \mathbb{P}(X_m = 2|\mu_1))' \) denote the probability mass function of \( X_m \) under \( \mu_1 \), and define \( v_2 \) similarly. Let \( e = (1,0)' \), then, in view of \( v_1't = v_2't = 1 \),
\[
\sum_{j \in \mathcal{X}} \mathbb{P}(X_k = 1|X_m = j) \mu_1(j) - \sum_{j \in \mathcal{X}} \mathbb{P}(X_k = 1|X_m = j) \mu_2(j) = v_1' P^{k-m} e - v_2' P^{k-m} e = (v_1 - v_2)'(t\pi + R \varrho^{k-m} u')e = (v_1 - v_2)'R \varrho^{k-m}.
\]
Similarly, we obtain \( \sum_{j \in \mathcal{X}} \mathbb{P}(X_k = 2|X_m = j) \mu_1(j) - \sum_{j \in \mathcal{X}} \mathbb{P}(X_k = 2|X_m = j) \mu_2(j) = -(v_1 - v_2)'R \varrho^{k-m} \), and the stated result follows because \( \max_{v_1,v_2} |(v_1 - v_2)'R| = 1 \). Part (c) follows from repeating the proof of Lemmas 10 and 12 using part (b). \( \square \)

10.3.4 The sums of powers of \( \rho \)

Lemma 14. For all \( \rho \in (0,1) \), \( c \geq 1 \), \( q \geq 1 \), and \( b > a \),
\[
\sum_{t=-\infty}^{\infty} \rho^{\lfloor t-a \rfloor/q} \rho^{\lfloor b-t \rfloor/q} \leq \frac{q(c+1)\rho^{\lfloor (b-a)/(c+1) \rfloor}}{1-\rho},
\]
\[
\sum_{t=-\infty}^{\infty} \rho^{\lfloor t-a \rfloor} \rho^{\lfloor b-t \rfloor} \leq \frac{q(c+1)\rho^{\lfloor (b-a)/(c+1) \rfloor}}{1-\rho}.
\]
Proof of Lemma 14. The first result holds because the left hand side is bounded by

\[
\sum_{t=-\infty}^{\lfloor (a+bc)/(c+1) \rfloor} \rho^{\lfloor (b-t)/q \rfloor} + \sum_{t=\lfloor (a+bc)/(c+1) \rfloor + 1}^{\infty} \rho^{\lfloor (t-a)/cq \rfloor} \\
\leq q\rho^{\lfloor (b-\lfloor (a+bc)/(c+1) \rfloor)/q \rfloor}/(1 - \rho) + cq\rho^{\lfloor \lfloor (a+bc)/(c+1) \rfloor+1 - a)/cq \rfloor}/(1 - \rho) \\
\leq q(1 + c)\rho^{\lfloor (b-a)/(c+1)q \rfloor}/(1 - \rho).
\]

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 15. For all \(\rho \in (0, 1)\), \(k \geq 1\), \(q \geq 1\), and \(n \geq 0\),

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

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

Proof of Lemma 15. When \(k = 1\), the stated result follows from Lemma 14 with \(c = 1\). We first show that the following holds for \(k \geq 2\):

\[
\sum_{t_1 \leq t_2 \leq \ldots \leq t_k \leq n} \left( \rho^{\lfloor (t_1-t)/q \rfloor} \wedge \rho^{\lfloor (t_2-t)/q \rfloor} \wedge \ldots \wedge \rho^{\lfloor (t_{k-1}-t)/q \rfloor} \wedge \rho^{\lfloor (n-t)/q \rfloor} \right) \leq q^{k-1}(k+1)\rho^{\lfloor (n-t)/q \rfloor}/(1 - \rho)^{k-1}. \tag{104}
\]

We prove (104) by induction. When \(k = 2\), it follows from Lemma 14 with \(c = 1\) that

\[
\sum_{t_2=t_1}^{n} \left( \rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \rho^{\lfloor (t_3-t_2)/q \rfloor} \wedge \ldots \wedge \rho^{\lfloor (t_{n+1}-t_2)/q \rfloor} \wedge \rho^{\lfloor (n-t_1)/q \rfloor} \right) \leq 2q\rho^{\lfloor (n-t_1)/2q \rfloor}/(1 - \rho),
\]

giving (104). Suppose (104) holds when \(k = \ell\). Then (104) holds when \(k = \ell + 1\) because, from Lemma 14,

\[
\sum_{t_1 \leq t_2 \leq \ldots \leq t_{\ell+1} \leq n} \left( \rho^{\lfloor (t_1-t)/q \rfloor} \wedge \rho^{\lfloor (t_2-t)/q \rfloor} \wedge \ldots \wedge \rho^{\lfloor (t_{\ell+1}-t)/q \rfloor} \wedge \rho^{\lfloor (n-t_{\ell+1})/q \rfloor} \right) \\
\leq \sum_{t_2=t_1}^{n} \left( \rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \sum_{t_3=\ldots=t_{\ell+1} \leq n} \left( \rho^{\lfloor (t_3-t)/q \rfloor} \wedge \ldots \wedge \rho^{\lfloor (t_{\ell+1}-t)/q \rfloor} \wedge \rho^{\lfloor (n-t_{\ell+1})/q \rfloor} \right) \right) \\
\leq q^{\ell-1}(1 - \rho)^{\ell-1} \sum_{t_2=t_1}^{n} \left( \rho^{\lfloor (t_2-t_1)/q \rfloor} \wedge \rho^{\lfloor (n-t)/q \rfloor} \right) \\
\leq q^{\ell}(\ell+1)!\rho^{\lfloor (n-t_1)/(\ell+1)q \rfloor},
\]
and hence (104) holds for all \( k \geq 2 \). We proceed to show the stated result. Observe that

\[
\sum_{0 \leq t_1 \leq t_2 \leq \cdots \leq t_k \leq n} \left( \rho_{t_1/q} \wedge \rho_{(t_2-t_1)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q} \right)
\leq \sum_{t_1=0}^{n/2} \sum_{t_2 \geq t_1} \cdots \sum_{t_k \geq t_{k-1}} \sum_{t_k=t_1}^{n-t_k} \left( \rho_{t_1/q} \wedge \rho_{(t_2-t_1)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q} \right)
\leq \sum_{t_1=0}^{n/2} \sum_{t_2 \geq t_1} \cdots \sum_{t_k \geq t_{k-1}} \sum_{t_k=t_1}^{n-t_k} \left( \rho_{(t_2-t_1)/q} \wedge \rho_{(t_3-t_2)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q} \right)
\leq 2 \sum_{t_1=0}^{n/2} \sum_{t_2 \geq t_1} \cdots \sum_{t_k \geq t_{k-1}} \sum_{t_k=t_1}^{n-t_k} \rho_{(t_2-t_1)/q} \wedge \rho_{(t_3-t_2)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q}
\leq 2 \sum_{t_1=0}^{n/2} \sum_{t_2 \geq t_1} \cdots \sum_{t_k \geq t_{k-1}} \sum_{t_k=t_1}^{n-t_k} \left( \rho_{(t_2-t_1)/q} \wedge \rho_{(t_3-t_2)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q} \right)
\leq 2 \sum_{t_1=0}^{n/2} \sum_{t_2 \geq t_1} \cdots \sum_{t_k \geq t_{k-1}} \sum_{t_k=t_1}^{n-t_k} \left( \rho_{(t_2-t_1)/q} \wedge \rho_{(t_3-t_2)/q} \wedge \cdots \wedge \rho_{(t_k-t_{k-1})/q} \wedge \rho_{(n-t_k)/q} \right),
\]

where the first equality holds by symmetry, and the second equality follows from \( n - t_k \geq t_1 \). From (104), the right hand side is no larger than \( q^{k-1}(k+1)(1-\rho)^{1-k} \sum_{t_1=0}^{n/2} \rho_{(n-t_1)/q} \leq q^{k}(k+1)! (1-\rho)^{-k} \rho^{[n/2q]} \), 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 16.** 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

\[
\max_{-m+1 \leq t_1 \leq \cdots \leq t_k \leq -t_k} a_{t_1} \cdots a_{t_k} \leq (k+m)^{\ell+1} A_\ell,
\]

where \( A_\ell := \sum_{t=\infty}^{-\infty} |t| \wedge 1 \left| \sum_{t=\infty}^{-t} a_t \right|^2 \).

**Proof of Lemma 16.** When \( \ell = 1 \), the stated result follows from \( \max_{-m+1 \leq t_1 \leq \cdots \leq t_k} a_{t_1} \cdots a_{t_k} \leq \sum_{t=-m+1}^{k} a_t = \sum_{t=-m+1}^{k} |t| \wedge 1|^2 |t| \wedge 1 \right|^2 a_t \leq (k+m)^2 \sum_{t=-\infty}^{-1} |t| \wedge 1 \right| a_t \). When \( \ell \geq 2 \), from the Hölder’s inequality, we have \( \max_{-m+1 \leq t_1 \leq \cdots \leq t_k} a_{t_1} a_{t_2} \cdots a_{t_k} \leq \left( \sum_{t=-m+1}^{k} a_t \right)^\ell = \left( \sum_{t=-m+1}^{k} |t| \wedge 1 \right|^{2/\ell} |t| \wedge 1 \right|^{\ell-1} \sum_{t=m}^{k} \right| a_t \right|^{\ell-1} A_\ell = (k+m)^{1+2/(\ell-1)} A_\ell \leq (k+m)^{\ell+1} A_\ell \).

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

**Lemma 17.** For \( \alpha > 0 \), \( q > 0 \), and \( c_{j,t} \geq 0 \), define \( c_j^{\infty} := \sum_{t=-\infty}^{-1} \rho_{|t|/q} c_j \). For all \( \rho \in (0,1) \), \( k \geq 1 \), and \( 0 \leq m \leq m' \),

\[
\sum_{t_1=-m'+1}^{t_1} \sum_{t_2 \leq t_3 \leq \cdots \leq t_k \leq t_6} \left( \rho_{(k-1-t_6)/q} \wedge \rho_{(t_6-t_5)/q} \wedge \rho_{(t_5-t_4)/q} \wedge \rho_{(k-t_4)/q} \wedge \rho_{(t_3-t_2)/q} \wedge \rho_{(t_2-t_1)/q} \right) \prod_{j=1}^{6} c_{j,t_j} \leq \rho_{(k+1+m)/2q\alpha}^{c_j^{\infty}} \left( \rho^{3/2\alpha} \right)^{\prod_{j=2}^{6} c_{j,q}^{\infty}} \left( \rho^{1/4\alpha} \right),
\]

(105)
where \((a_j, b_j)\) are defined recursively with \((a_2, b_2) = (1, 1)\) and, for \(j \geq 3\),

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

\(a_j, b_j \geq 3/2\) for all \(j\). Direct calculations using Matlab produce \(a_7 \approx 334.5406\).

**Proof of Lemma 17.** First, observe that the following result holds for \(a, b > 1/4, t_1 \leq 0\), and \(t_j, t_{j+1} \geq t_1\):

\[
(a) \text{ if } t_j \leq \frac{at_j + t_1}{a + b}, \quad \text{then } \frac{|t_j|}{4a} \leq \frac{a(4a + 1)t_{j+1} + (2a - 1)t_1}{4a(a + b)} - t_j.
\]

\[
(b) \text{ if } t_j \geq \frac{at_j + t_1}{a + b}, \quad \text{then } \frac{|t_j|}{4a} \leq \frac{b}{a} - \frac{a(4b - 1)t_{j+1} + (2a + 4b + 1)t_1}{4a(a + b)}.
\]

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

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

We proceed to derive the stated bound. It follows from (a) and (b) and \(|x + y| \geq |x| + |y|\) that, with \(\tau_j = (a_j t_{j+1} + t_1)/(a_j + b_j)\),

\[
\sum_{t_j = -m'}^{k} \rho^{[\frac{(t_j + 1 - t_j)}{4}]} \rho^{[\frac{(b_j t_j - t_j)}{4}]} c_j t_j \leq \rho^{\frac{a_j (4b_j - 1)t_{j+1} + (2a_j - 1)t_1}{4a_j (a_j + b_j)}} \left( \sum_{t_j \leq \tau_j} \rho^{\frac{a_j (4a_j + 1)t_{j+1} + (2a_j - 1)t_1}{4a_j (a_j + b_j)} - \frac{t_j}{4}} + \sum_{t_j \geq \tau_j} \rho^{\frac{b_j t_j - a_j (4b_j - 1)t_{j+1} + (2a_j + 4b_j + 1)t_1}{4a_j (a_j + b_j)}} \right) c_j t_j
\]

\[
\leq \rho^{\frac{a_j (4b_j - 1)t_{j+1} + (2a_j - 1)t_1}{4a_j (a_j + b_j)}} \left( \sum_{j} \rho^{\frac{b_j t_j - a_j (4b_j - 1)t_{j+1} + (2a_j + 4b_j + 1)t_1}{4a_j (a_j + b_j)}} \right) c_j t_j
\]

\[
= \rho^{\frac{b_j + t_{j+1} - t_j}{a_j + 4}} \left( \sum_{j} \rho^{\frac{b_j t_j - a_j (4b_j - 1)t_{j+1} + (2a_j + 4b_j + 1)t_1}{4a_j (a_j + b_j)}} \right).
\]

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
\[
\sum_{t_1=-m'}^{r} \rho_1^{\left|\frac{b_\gamma(k-1)-t_1}{a_\gamma}\right|} \cdot c_{1t_1} \cdot \prod_{j=2}^{6} c_{j|q}^{\infty} \left(\rho^{1/4a_j}\right).
\]
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 \(b_\gamma(k-1) \geq k-1\) and \(|t_1| \leq k-1-2t_1-m\), the sum is bounded by
\[
\sum_{t_1=-m'}^{r} \rho_1^{\left|\frac{k-1-t_1}{a_\gamma}\right|} \cdot c_{1t_1} = \rho_1^{\left|\frac{k-1+m}{2a_\gamma}\right|} \sum_{t_1=-m'}^{r} \rho_1^{\left|\frac{k-1-2t_1-m}{2a_\gamma}\right|} \cdot c_{1t_1} \leq \rho_1^{\left|\frac{k-1+m}{2a_\gamma}\right|} c_{1q}^{\infty} \left(\rho^{1/2a_j}\right),
\]
and the stated result follows.

Lemma 18. When the regime \(h\) and \(h+1\) are combined into one regime in an \(M_0+1\)-regime model with \(\vartheta_{M_0+1,x}\), the transition probability of \(X_k\) equals the transition probability of \(X_k\) under \(\vartheta_{M_0,x}\) if and only if \(\vartheta_{M_0+1,x} \in \Theta_{ph} \times \Pi^{*}_h\).

Proof of Lemma 18. The “only if” part is trivial. To prove the “if” part, let \(h = M_0\) without loss of generality. Let \(X_k \sim P_{\vartheta_{M_0+1}}\) and \(\tilde{X}_k \sim P_{\vartheta_{M_0}}\). The stated result follows if we show that, for any \(\vartheta_{M_0+1} \in \Theta_{M_0}\), (a) \(P_{\vartheta_{M_0+1}}(X_k+1 = j|X_k \in J_{M_0}) = \vartheta_{M_0}^{*}(\tilde{X}_k+1 = j|X_k = M_0)\) for \(j \leq M_0 - 1\) and (b) \(P_{\vartheta_{M_0+1}}(X_k+1 \in J_{M_0}|X_k \in J_{M_0}) = \vartheta_{M_0}^{*}(\tilde{X}_k+1 = M_0|X_k = M_0)\). For (a), observe that
\[
P_{\vartheta_{M_0+1}}(X_k+1 = j|X_k \in J_{M_0}) = \frac{P_{\vartheta_{M_0+1}}(X_k+1 = j) - P_{\vartheta_{M_0+1}}(\{X_k+1 = j\} \cap \{X_k \in J_{M_0}\})}{1 - P_{\vartheta_{M_0+1}}(X_k \in J_{M_0})}.
\]
If \(p_{ij} = p^{*}_{ij}\) and \(j = p^{*}_{j}\) for all \(i,j \leq M_0 - 1\), then we have, for any \(j \leq M_0 - 1\), \(P_{\vartheta_{M_0+1}}(X_k+1 = j) = \vartheta_{M_0}^{*}(\tilde{X}_k+1 = j), P_{\vartheta_{M_0+1}}(X_k \in J_{M_0}) = \vartheta_{M_0}^{*}(\tilde{X}_k \leq M_0 - 1),\) and
\[
P_{\vartheta_{M_0+1}}(\{X_k+1 = j\} \cap \{X_k \in J_{M_0}\}) = \sum_{i \leq M_0 - 1} P_{\vartheta_{M_0+1}}(X_k+1 = j|X_k = i)P_{\vartheta_{M_0+1}}(X_k = i)
= \sum_{i \leq M_0 - 1} \vartheta_{M_0}^{*}(\tilde{X}_k+1 = j|X_k = i)\vartheta_{M_0}^{*}(\tilde{X}_k = i)
= \vartheta_{M_0}^{*}(\{\tilde{X}_k+1 = j\} \cap \{\tilde{X}_k \leq M_0 - 1\}).
\]
Substituting them to (108) gives \(P_{\vartheta_{M_0+1}}(X_k+1 = j|X_k \in J_{M_0}) = \vartheta_{M_0}^{*}(\tilde{X}_k+1 = j|X_k = M_0)\), and (a) follows. Because \(P_{\vartheta_{M_0+1}}(X_k+1 \in J_{M_0}|X_k \in J_{M_0}) = 1 - \sum_{j \leq M_0 - 1} P_{\vartheta_{M_0+1}}(X_k+1 = j|X_k \in J_{M_0})\), (b) follows from (a), and the stated result follows.

References


