Testing for the Mixture Hypothesis with Misspecified Distributions in Exponential Family

JIN SEO CHO
School of Economics and Finance
Victoria University of Wellington
P.O. Box 600, Wellington, 6001, New Zealand
Email: JinSeo.Cho@vuw.ac.nz

HALBERT WHITE
Department of Economics
University of California, San Diego
9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.
Email: hwhite@weber.ucsd.edu

First version: Feb. 17, 2001. This version: Feb. 29, 2008

Abstract

We examine the likelihood ratio (LR) statistic testing for the mixture hypothesis when the mixtures of distributions in exponential family are incorrectly specified.

The analysis assuming the correct model specification condition in Cho and White (2007, *Econometrica*) cannot be exactly applied to the misspecified models in a way similar to identified models. Further, the LR statistic often becomes degenerate under the null that the single component based model attains the same likelihood as the mixture. We thus provide a set of regularity conditions for the LR statistic to be non-degenerate asymptotically under the null. Given our regularity conditions, the LR statistic converges weakly to the square of half-normal random variable, non-central chi-square random variable or the maximum of these under the null, depending on the model scopes and data generating processes considered in the text. This is a different consequence from the correctly specified model case, in which the asymptotic null distribution is given as a functional of a Gaussian process.

Key Words: Mixture Hypothesis; Model Misspecification; Likelihood Ratio Statistic; Exponential Family Distributions.
Subject Class: C12, C22, C32, C52.

Acknowledgements: We have benefited from discussions with Zongwu Cai, Robert Davies, Ying Fang, Cheng Hsiao, and other participants at Hidden Markov Models and Complex Systems (Wellington, Dec., 2005), ASC/NZSA (Auckland, Jul., 2006), and Sino-Korean Econometrics Workshop (Xiamen, Dec., 2007).

1 Introduction

Mixture models are very popular in empirical economics. In labor economics, Heckman and Singer (1984) specify the mixture of Weibull distributions to estimate unemployment duration data consistently. In macroeconomics, Hamilton (1989) develops the mixture of autoregressive processes into a regime switching model and examines the business cycle of the postwar US real GNP. This popularity is not restricted only to economics but also reaches most disciplines.

Careful mixture specifications are emphasized in the literature. Most of all, its flexibility to generate another model by appending additional component (or regime) has the consequence to generate another model with components more than necessary, not estimating the parameters of interests consistently. Therefore the prior literature recommends testing first the hypotheses that there is a single component say, versus multiple components, say two. This recommendation can be found out in Hartigan (1985), Ghosh and Sen (1985), Chernoff and Lander (1995), Dacunha-Castelle and Gassiat (1999), Chen and Chen (2001), Cho and White (2007), and the references therein. Mainly, they examine the asymptotic behaviors of standard log-likelihood ratio (LR) statistics under the single component hypothesis.

These emphases are deduced from correct model specification condition, and the literature hasn't shed light on the same issue under misspecified model context. In particular, this issue is important given that misspecified models play a central role in economic data analyses, and that there is a strong tendency to interpret correctly specified models as a variation of misspecified models. As pointed out by White (1982), the failure to specify models correctly is followed by information matrix inequality.

The main goal of this paper, therefore, is to examine how the LR statistic behaves asymptotically under the null when the null model is misspecified and the same hypotheses are tested. For this, we follow the model analysis posited by Cho and White (2007) and exploit their regularity conditions. Also, we focus our interests to exponential family distributions. That is, we examine the mixtures of distributions in exponential family. This is not only because they are popular, but also because we can contrast our results with correctly specified models. As we show below, correctly specified models cannot be restated in the framework of misspecified models due to the exponential family distribution condition, mainly.

Our main results can be summarized as follows. First, when models are misspecified, the large sample distributions of the LR statistic cannot be derived in the same way as for correctly specified models. As we show in more detail below, this is mainly because the LR statistic is degenerate on the boundary null parameter space. Second, we therefore provide a set of regularity conditions to prevent the LR statistic from being degenerate. Our regularity conditions are pertained to the null parameter space not involving boundary parameters. Certainly, correct model specification assumption is sufficient to have non-degenerate LR statistic on the non-boundary null parameter space, but our regularity conditions nest the conditions of correct model specification as a special case. Thus, we can interpret the results for correctly specified models as special cases of ours under the null not involving boundary parameters. Third, under our maintained misspecification assumption, we show that the LR statistic weakly converges to the square of half-normal random variable, non-central chi-square random variable or the maximum of these. It depends on the model scopes suggested below specifically. Thus, the asymptotic null distribution of the LR statistic for correctly specified models are different from misspecified models, because their asymptotic null distributions are also determined by the null parameter space involving boundary parameters.

The plan of this paper is as follows. In Section 2, we describe the data and mixture models of interests. We expound main results in Section 3 by providing the asymptotic null distributions of the LR statistic. By following Cho and White (2007), we consider two different cases. The first is the case in which null distributions of the LR statistic can be approximated by a fourth-order Taylor expansion, and the second is the case requiring much higher-order approximations. Section 4 contains the conclusion, and we collect all the mathematical proofs in the Appendix.

2 Data Generating Process and Misspecified Mixtures

To proceed our discussion in a manageable way, we assume the following data generating process (DGP) and model conditions.

A1 (DGP): *An observed data set* $\{X_t \in \mathbb{R}^d\}_{t=1}^n (d \in \mathbb{N})$ *is generated as a sequence of strictly stationary and ergodic sequence, and for each* $t = 1, \cdots, n$, *the conditional distribution function of* X_t *given* X^{t-1} *is a measurable function* $g(\,\cdot\,|X^{t-1})$, *where* X^{t-1} *is the smallest* σ*-algebra adapted to* $(X_{t-1}, \cdots, X_1)$.

A2 (MODEL): (*i*) *A model for* $g(\,\cdot\,|X^{t-1})$ *is given as a misspecified mixture of* $\{f(\,\cdot\,|X^{t-1};\theta^j) : \theta^j := (\theta_0, \theta_j) \in \tilde{\Theta}\}$, *where* $\tilde{\Theta} := \Theta_0 \times \Theta_* \in \mathbb{R}^{r_0+1}$; and Θ_0 and Θ_* *are convex and compact sets in* $\mathbb{R}^{r_0}$ *and* $\mathbb{R}$ *respectively* ($j = 1, 2$).
(*ii*) *For each* t, $f(\,\cdot\,|X^{t-1};\theta^j)$ *is of the exponential form*

$$f(x|X^{t-1};\theta^j) = d(x)\exp\left(\sum_{i=1}^{k} a_i\left(X^{t-1};\theta^j\right) b_i(x) + c\left(X^{t-1};\theta^j\right)\right),$$

where $d : \mathbb{R}^d \mapsto \mathbb{R}$ is a measurable function; $a_i(x^{t-1}; \cdot) : \tilde{\Theta} \mapsto \mathbb{R}$; and $b_i^2(X_t)$ is integrable (with respect to the DGP) for all $i = 1, \cdots, k$. Further, there is a measurable function $T : \mathbb{R}^k \mapsto \mathbb{R}$ such that $T(b_1(x), \cdots, b_k(x)) = x$.

A3 (DIFFERENTIATION): (*i*) *$f(X_t|X^{t-1}; \cdot) \in \mathcal{C}^{(1)}(\tilde{\Theta})$ almost surely (a.s.), where $\mathcal{C}^{(k)}(\tilde{\Theta})$ denotes a set of functions k-times continuously differentiable.*

The assumption A2 implies that for $(\pi, \theta) \in [0, 1] \times \Theta$,

$$\prod_{t=1}^{n} \{\pi f(X_t|X^{t-1}; \theta^1) + (1-\pi) f(X_t|X^{t-1}; \theta^2)\}$$

is specified for the probability density function (PDF) of $(X_n, X_{n-1}, \cdots, X_1)$, where $\theta := (\theta_0, \theta_1, \theta_2)$ and $\Theta := \Theta_0 \times \Theta_* \times \Theta_*$, and this accommodates a considerable range of the models popular for empirical applications. In particular, the mixture of linear regression normal residuals belongs to this model specification. For notational simplicity, we abbreviate $f(X_t|X^{t-1}; \theta^j)$ as $f_t(\theta^j)$ throughout this paper. Also, we omit the argument of functions given above and below, so that as an example, $f_t(\cdot)$ can be also denoted as f_t. Additionally, for each $(\pi, \theta) \in [0, 1] \times \Theta$, we let

$$\ell_t(\pi, \theta) := \log(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2)),$$

and for each θ^1, $\tilde{\ell}_t(\theta^1) = \log(f_t(\theta^1))$; and we assume the following to apply the large sample theory to the above set-up.

A4 (EXISTENCES): (*i*) *For each (π, θ), $E[\ell_t(\pi, \theta)]$ exists and is finite uniformly on $[0, 1] \times \Theta$.*
(*ii*) *$(\pi^\dagger, \theta_0^\dagger, \theta_1^\dagger, \theta_2^\dagger)$ maximizes $E[\ell_t]$ on $[0, 1] \times \Theta$.*
(*iii*) *(θ_0^*, θ_*) uniquely maximizes $E[\tilde{\ell}_t]$ in the interior of $\tilde{\Theta}$.*

These are the minimal regularity conditions on which we can proceed with our presentation without involving too much complication. As we shortly encounter below, the model $\tilde{\ell}_t$ is associated with the null hypothesis that the single component based model satisfies the standard model identification condition in the literature, A4(iii). Meanwhile, we cannot maintain the same hypothesis under the alternative model, ℓ_t. Under the null, it turns out that the model ℓ_t is not identified. That's why the identification condition is missing in A4(ii).

We provide additional regularity conditions as our discussions move on. These are provided to apply mainly central limit theorems (CLTs) to the data of interest. For simplicity, we will say that a strictly stationary and ergodic sequence, say $\{Y_t \in \mathbb{R}^d\}$, satisfies *an adapted mixingale* (*AMXG*) condition, if (*i*) $\{Y_t, \mathcal{F}_t\}$ is an adapted mixingale of size -1; and (*ii*) the asymptotic variance of $n^{-1/2}\sum_{t=1}^{n} Y_t$ exists, where $\mathcal{F}_t$ is a σ-algebra adapted to $(Y_t, \cdots, Y_1)$. It's well-known that $E[Y_t] = 0$ and $n^{-1/2}\sum Y_t \overset{\text{A}}{\sim} N(0, D)$ under the AMXG condition, where D is the asymptotic variance of $n^{-1/2}\sum Y_t$ (see theorem 3 in Scott (1973)). Given the AMXG condition, a test statistic of interest is not necessarily non-degenerate, because D can be singular. If the same sequence additionally has var$(n^{-1/2}\sum Y_t)$ which is uniformly positive definite with respect to n, then we will say that $\{Y_t \in \mathbb{R}^d\}$ satisfies a *non-degenerate adapted mixingale sequence* (*NAMXG*) condition.

3 Asymptotic Null Distribution of Likelihood Ratio Statistic

Given these, suppose further that a researcher tests whether a single component based Kullback-Leibler information level is equivalent to that obtained by the mixture. Formally, the researcher wishes to choose one of the following hypotheses: for the unknown θ_*,

$$H_0 : \pi^\dagger = 1 \text{ and } \theta_1^\dagger = \theta_*;\ \pi^\dagger = 0 \text{ and } \theta_2^\dagger = \theta_*;\ \text{or } \theta_1^\dagger = \theta_2^\dagger = \theta_*;\ \text{versus } H_1 : \pi^\dagger \in (0,1) \text{ and } \theta_1^\dagger \neq \theta_2^\dagger,$$

and makes a decision by the LR statistic computed as

$$LR_n := 2(L_n(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) - L_n(1, \hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n, \theta_2)),$$

where for each $(\pi, \theta) \in [0,1] \times \Theta$,

$$L_n(\pi, \theta_0, \theta_1, \theta_2) := \sum_{t=1}^{n} \ell_t(\pi, \theta_0, \theta_1, \theta_2);$$

$(\hat{\pi}_n^a, \hat{\theta}_0^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a)$ and $(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)$ are the quasi-maximum likelihood estimators (QMLEs) obtained under the alternative and the null respectively; and θ_2 in the null is a placeholder whose value is irrelevant. Note that the misspecification assumption on the mixture in A2 implies that the null model is misspecified; and that $\theta_0^\dagger = \theta_0^*$ under the null.

The given hypotheses are exactly the same as those for correctly specified models. The null hypothesis contains composite irregular conditions, which makes it hard to apply the conventional model analysis. To take account this aspect, we may partition the null parameter space as follows:

(i) $H_{01} : \pi^\dagger = 1$ and $\theta_1^\dagger = \theta_*$;

(ii) $H_{02} : \theta_1^\dagger = \theta_2^\dagger = \theta_*$;

(iii) $H_{03}: \pi^\dagger = 0$ and $\theta_2^\dagger = \theta_*$.

Under H_{01} (resp. H_{03}), $\pi^\dagger$ is on the boundary and θ_2 (resp. θ_1) is not identified, so that Davies' (1977, 1987) identification problem arises. That is, there is a nuisance parameter identified only under the alternative. Also, $\pi^\dagger$ is on the boundary of the parameter space, implying that interiority condition violates as well. Similarly, $\pi^\dagger$ is not identified under H_{02}, but more seriously, the first-order derivative turns out to be zero identically under H_{02} as explored below. Due to these, the asymptotic null behavior of the LR statistic cannot be obtained by the conventional second-order Taylor expansion.

Nevertheless, we cannot carry over the same model analysis for correctly specified models to misspecified models. This is mainly because there is no longer a "true parameter" vector corresponding to the DGP. For the misspecified case, the key element for testing the given hypotheses is an information gain obtained by forming a mixture. Thus, we need to define multiple components (or regimes) in a different way from correctly specified models. We will say that *there is an additional component* (*or regime*) in data if there is any information gain by forming a mixture. This interpretation is more flexible than those maintained by correctly specified models, because the global maximum of the Kullback-Leibler information level is attained at the true parameter vector corresponding to the DGP.

We examine the null distribution of the LR statistic under each hypothesis. Before proceeding our discussion, we note that exactly the same analysis can be carried out under H_{01} and H_{03} due to the symmetry of mixture. Thus, we don't examine them separately.

3.1 Null Distribution under H_{01} or H_{03}

The exponential family model assumption has a significant implication for the LR statistic under H_{01} (or H_{03}). Under H_{01}, the LR statistic turns out to be negligible in probability. For this examination, we assume the following regularity condition.

A5 (MOMENTS): *There exists a sequence of positive*, *strictly stationary*, *and ergodic random variables* $\{M_t\}$ *with* $E[M_t] < \Delta < \infty$, (*i*) $\sup_{(\pi,\theta)} |\nabla_j \ell_t(\pi, \theta)| \leq M_t$ *for all* $j \in \{\pi, \theta_{01}, \cdots, \theta_{0r}, \theta_1\}$.

We impose this condition for the existence of a finite moment of the first-order derivatives, and this condition is weaker than the second-order derivative based regularity condition for standard econometric models. Given this condition, the following lemma promotes our further discussion on the misspecified mixtures.

LEMMA 1: (*i*) *Given A1, A2, A3*(*i*) *and A5*(*i*), *if for some* $(\theta_0^\star, \theta_\star) \in \tilde{\Theta}$, *the PDF of* $X_t|X^{t-1}$ *is identical to* $f(\,\cdot\,|X^{t-1}; \theta_0^\star, \theta_\star)$ *then for all* $\theta_2 \in \Theta_*$, $E[\nabla_\pi \ell_t(1, \theta_0^\star, \theta_\star, \theta_2)|X^{t-1}] = 0$ *a.s.*
(*ii*) *Given A1, A2, A3*(*i*) *and A5*(*i*), *if for some* $(\theta_0^\star, \theta_\star) \in \tilde{\Theta}$, $E[\nabla_\pi \ell_t(1, \theta_0^\star, \theta_\star, \theta_2)|X^{t-1}] = 0$ *a.s. for all* $\theta_2 \in \Theta_*$ *then the PDF of* $X_t|X^{t-1}$ *is the same as* $f(\,\cdot\,|X^{t-1}; \theta_0^\star, \theta_\star)$ *a.s.*

We prove Lemma 1 using the fact that the moment generating function (MGF) of $X_t|X^{t-1}$ can be derived from the first-order condition (FOC), and it can be further used for distributional identification. Lemma 1(ii) implies that the model misspecification assumption does not allow us to impose the zero first–order derivative condition uniformly on the set

$$\{(\pi, \theta_0, \theta_1, \theta_2) \in [0,1] \times \Theta : \pi = 0, \theta_0 = \theta_0^*, \theta_2 = \theta_*\} \\ \cup \{(\pi, \theta_0, \theta_1, \theta_2) \in [0,1] \times \Theta : \pi = 1, \theta_0 = \theta_0^*, \theta_1 = \theta_*\}.$$

As a straightforward example, suppose that an identically and independently distributed data set, $\{X_t \in \mathbb{R}\}$, whose variance is known as one, is modeled as a mixture of normals with unknown means. That is, the log-likelihood of X_t is specified as

$$\ell_t(\pi, \mu_1, \mu_2) = \log(\pi\phi(\mu_1, X_t) + (1-\pi)\phi(\mu_2, X_t)),$$

where $\phi(\mu, X_t)$ is the PDF of normal random variable with mean μ and variance one. Then, the FOC in Lemma 1(ii) can be rephrased as $E[\exp(\tau X_t)] = \exp(\tau\mu^\star + \frac{1}{2}\tau^2)$, where $\tau := (\mu_2 - \mu^\star)$, which implies that the MGF of X_t is that of $N(\mu^\star, 1)$. There is one-to-one relationship between MGF and distribution function, so that we cannot impose the zero-FOC together with the model misspecification assumption. This aspect also entails that the regularity conditions for misspecified models in the literature are not easy to maintain under exponential mixture context. For example, Andrews (2001) considers identification problem on boundary in a general model framework and assumes that the FOC holds on the boundary, which cannot hold under our model condition. We understand his condition as a device to obtain a practically conservative null test distribution. Instead, we impose a little bit stronger assumption.

A6 (BOUNDARY DERIVATIVE): $E[\nabla_\pi \ell_t(1, \theta_0^*, \theta_*, \cdot)|\mathcal{F}_{t-1}] > 0$ (*or* $E[\nabla_\pi \ell_t(0, \theta_0^*, \cdot, \theta_*)|\mathcal{F}_{t-1}] < 0$) *a.s. on* $\Theta_* \setminus \{\theta_*\}$.

This assumption is not implied by Lemma 1, and it may restrict the DGPs under consideration, because there can be DGPs that violate the FOC non–uniformly on $\Theta_* \setminus \{\theta_*\}$. Nevertheless, this assumption is not so strong as it appears. Bierens (1990) considers MGF as an essential diagnostic tool for correct specification. Rephrasing his lemma 1 shows that $E[\nabla_\pi \ell_t(1, \theta_0^*, \theta_*, \cdot)|X^{t-1}] = 0$ a.s. on a set whose Lebesgue measure is at most zero, if the model for $X_t|X^{t-1}$ is misspecified. Thus, the generality of our analysis is not significantly compromised by our assumption.

The LR statistic is negligible in probability under H_{01} (or H_{03}) by this condition. For each $\theta_2 \neq \theta_*$, the information level increases as π approaches one, and further higher information level is achieved by letting π bigger than one, which is infeasible by our assumptions on the mixtures. Accordingly, $\hat{\pi}_n^a(\theta_2)$ converges to one and further equals one a.s., where for each θ_2, we denote

$$\{\hat{\pi}_n^a(\theta_2), \hat{\theta}_{0,n}^a(\theta_2), \hat{\theta}_{1,n}^a(\theta_2)\} := \arg\max_{(\pi,\theta_0,\theta_1)} L_n(\pi, \theta_0, \theta_1, \theta_2).$$

This aspect is strengthened by applying the strong uniform law of large numbers (SULLN). That is, for each $\epsilon > 0$, $\hat{\pi}_n^a(\cdot) = 1$ a.s. over $\Theta_*(\epsilon) := \Theta_* \setminus \{\theta \in \Theta_* : |\theta - \theta_*| < \epsilon\}$, implying that the alternative model reduces to the null model, and the maximum level of information is identical to that attained under the null. By this, the asymptotic null distribution of the LR statistic under H_{01} (or H_{03}) can be formally given as follows.

THEOREM 1: *Given A1, A2, A3(i), A4, A6, A5(i) and* H_{01} *(or* H_{03}*), for each* $\epsilon > 0$,

$$LR_n^{(1)}(\epsilon) = o_p(1),$$

where $(\hat{\pi}_n^a(\epsilon), \hat{\theta}_{n0}^a(\epsilon), \hat{\theta}_{1,n}^a(\epsilon), \hat{\theta}_{2,n}(\epsilon))$ *maximizes* $L_n(\pi, \theta_0, \theta_1, \theta_2)$ *over* $[0,1] \times \Theta_0 \times \Theta_*(\epsilon) \times \Theta_*(\epsilon)$, *and*

$$LR_n^{(1)}(\epsilon) := 2\left\{ L_n(\hat{\pi}_n^a(\epsilon), \hat{\theta}_{n0}^a(\epsilon), \hat{\theta}_{1,n}^a(\epsilon), \hat{\theta}_{2,n}(\epsilon)) - L_n(1, \hat{\theta}_{0,n}^n, \theta_{1,n}^n, \theta_2) \right\}.$$

As the main idea of Theorem 1 is already expounded, we don't prove it in the Appendix. Theorem 1 also implies that the probability law of the LR statistic can be determined by H_{02}, when it is non-degenerate under H_{02}.

The negligible LR statistic under H_{01} (or H_{03}) is a different consequence from that for correctly specified models. For correctly specified models, the LR statistic weakly converges to a functional of Gaussian processes under H_{01} (or H_{03}).

3.2 Null Distribution under H_{02}

Even if models are misspecified, the LR statistic can be non-degenerate under H_0, unless it is degenerate under H_{02}. To examine this, we first consider the convergence limit of the QMLE and its asymptotic distribution under H_0. For this, we assume the following regularity conditions.

A3 (DIFFERENTIATION): (*ii*) $f(X_t|X^{t-1};\,\cdot\,) \in \mathcal{C}^{(2)}(\tilde{\Theta})$ *a.s.*

A5 (MOMENTS): (*ii*) $\sup_{\theta^1} |\nabla_{i_1} \tilde{\ell}_t(\theta^1)| \leq M_t$.
(*iii*) $\sup_{\theta^1} |\nabla_{i_1} \nabla_{i_2} \tilde{\ell}_t(\theta^1)| \leq M_t$ *for* $i_1, i_2 \in \{\theta_{01}, \cdots, \theta_{0r}, \theta_1\}$.

A7 (CLT): (*i*) $\{\nabla_{\theta^1} \tilde{\ell}_t(\theta_0^*, \theta_*)\}$ *satisfies the NAMXG condition and* $\lambda_{\min}(-A_*) \in (0, \infty)$, *where* A_* *denotes* $E[\nabla^2_{\theta^1} \tilde{\ell}_t(\theta_0^*, \theta_*)]$, *and* $\lambda_{\min}(\,\cdot\,)$ *is the minimum eigenvalue of a given matrix respectively.*

Assumptions A5(ii, iii) and A7(i) provide the sufficient conditions for obtaining the asymptotic distribution of the QMLE under the null. As these are quite well established in the literature, we take them for granted. (See White (1994)). We can formally claim the following.

THEOREM 2: (*i*) *Given A1, A2, A3*(*i*), *A4*(*i, ii*), *A5*(*i*), *A6, and* H_0, $(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) \to \{(\pi, \theta_0^*, \theta_1, \theta_2) \in [0,1] \times \Theta : \theta_1 = \theta_2 = \theta_*\}$ *a.s.*
(*ii*) *Given A1, A2, A3*(*i*), *A4*(*i, iii*), *A5*(*ii*), $(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) \to (\theta_0^*, \theta_*)$ *a.s.*
(*iii*) *Given A1, A2, A3*(*ii*), *A4*(*i, iii*), *A5*(*ii, iii*) *and A7*(*i*),

$$n^{1/2}\{(\hat{\theta}_{0,n}^{n\prime}, \hat{\theta}_{1,n}^{n\prime})' - (\theta_0^{*\prime}, \theta_*')'\} \overset{\text{A}}{\sim} N(0, (-A_*)^{-1} B_* (-A_*)^{-1}),$$

where B_* *is the asymptotic covariance matrix of* $n^{-1/2} \sum \nabla_{\theta^1} \tilde{\ell}_t(\theta_0^*, \theta_*)$.

As Theorems 2(ii and iii) are well known, we do not prove them in the Appendix (see theorem 6.4 in White (1994)). Note that the convergence limit of the QMLE in Theorem 2(i) is the same as the null parameter space given by H_{02}. It is mainly because the expected log-likelihood function on the boundary ($\pi = 0$ or 1) violates the FOC, but not under H_{02}. Thus, the QMLE converges to the null parameter space associated with H_{02}, and the asymptotic null distribution of the LR statistic is mainly determined by H_{02}.

The asymptotic null distribution of the LR statistic cannot be approximated by the conventional second-order Taylor expansion. As we show below, the first-order derivative is identically zero under H_{02}. Thus, its associated first-order derivative cannot determine the asymptotic null distribution.

Numerous authors examine this aspect under correctly specified mixture context. Neyman (1959) develops the C(α) statistic theory as a device for locally optimal test statistics and recommends exploiting the C(α) statistic using higher order derivatives when lower order derivatives are identically zero under the null. Neyman and Scott (1965) apply this recommendation to the mixtures of Poisson distributions, and Lindsay (1995) sustains its general applicability. Goffinet, Loisel, and Laurent (1992) also consider the same problem in the mixture of normal context and approximate the log-likelihood functions by fourth-order Taylor expansions. In a general framework, Dacunha-Castelle and Gassiat (1999) approximate general mixtures by fourth-order expansions and locally conic parametrization. Cho and White (2007) note that many popular mixtures need to be approximated by higher-order expansions than the fourth-order and extend mixture model scopes to those requiring eighth-order expansions. This virtually can arises if the second-order derivative is also identically zero under H_{02}. The mixture of normals with unknown different means and common variance belongs to this case.

We separately consider both cases with non-zero and zero second-order derivatives. Our discussion proceeds in a way to provide regularity conditions to have non-degenerate LR statistic and also intend them to accommodate correctly specified models as a special case under H_{02}. Even if this attempt is infeasible under H_{01} (or H_{03}) by Lemma 1, we can do this under H_{02} mainly by the absence of the boundary parameter problem.

3.2.1 Non-Zero Second-Order Derivative Case

We follow the notation in Cho and White (2007). For each $(\pi, \theta_2) \in (0,1) \times \Theta_*$, we let

$$\{\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)\} := \arg\max_{\theta^1 \in \tilde{\Theta}} L_n(\pi, \theta_0, \theta_1, \theta_2),$$

so that it satisfies the FOCs, implying that for each (π, θ_2),

$$\nabla_{\theta_0} L_n(\pi, \tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2), \theta_2) = \sum \frac{\pi f_t^{(1,0)}(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)) + (1-\pi) f_t^{(1,0)}(\tilde{\theta}_{0,n}(\theta_2), \theta_2)}{\pi f_t(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}_{0,n}(\theta_2), \theta_2)} \equiv 0,$$

$$\nabla_{\theta_1} L_n(\pi, \tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2), \theta_2) = \sum \frac{(1-\pi) f_t^{(0,1)}(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2))}{\pi f_t(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}_{0,n}(\theta_2), \theta_2)} \equiv 0,$$

where $f_t^{(i,j)} := \nabla_{\theta_0}^i \nabla_{\theta_1}^j f_t$. We omit π from the arguments of $(\tilde{\theta}_{0,n}, \tilde{\theta}_{1,n})$, as its role is paltry in terms of the asymptotic null distribution. As it turns out, π exists as a coefficient of the second-order derivative, so that proving its tightness becomes trivial. This also implies that Davies' (1977, 1987) identification problem is not a key issue under H_{02}. For each (π, θ_2), we further let

$$h_t(\theta_2) := f_t^{(0,1)}(\tilde{\theta}_{0,n}(\theta_2), \theta_2), \quad k_t(\theta_2) := f_t^{(0,1)}(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)),$$

$$m_t(\pi, \theta_2) := \pi f_t^{(1,0)}(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)) + (1-\pi) f_t^{(1,0)}(\tilde{\theta}_{0,n}(\theta_2), \theta_2),$$

$$g_t(\pi, \theta_2) := [\pi f_t(\tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}_{0,n}(\theta_2), \theta_2)]^{-1},$$

and $\tilde{L}_n(\pi, \theta_2) := L_n(\pi, \tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2), \theta_2)$. These are devised to rephrase higher-order derivatives in a shorthand notation. For example, using these, the identities given above can be written in a much simpler form.

$$\tilde{M}_n^{(1)}(\pi, \theta_2) := \nabla_{\theta_0} L_n(\pi, \tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2), \theta_2) = \sum m_t(\pi, \theta_2) g_t(\pi, \theta_2) \equiv 0, \tag{1}$$

$$\tilde{K}_n^{(1)}(\pi, \theta_2) := \nabla_{\theta_1} L_n(\pi, \tilde{\theta}_{0,n}(\theta_2), \tilde{\theta}_{1,n}(\theta_2), \theta_2) = (1-\pi) \sum k_t(\theta_2) g_t(\pi, \theta_2) \equiv 0, \tag{2}$$

where the superscript in the left-hand side (LHS), (1), denotes the first-order derivative of the given functions with respect to θ_2. Later, we will use similar superscripts when differentiating these functions more than once. Also, a simplified form of $\tilde{L}_n^{(1)}(\pi, \theta_2)$ can be given as

$$\tilde{L}_n^{(1)}(\pi, \theta_2) = (1-\pi) \sum h_t(\theta_2) g_t(\pi, \theta_2), \tag{3}$$

because the derivative terms involving θ_0 and θ_1 are zero. We further let

$$\hat{r}_t^{(i,j)} := f_t^{(i,j)}(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) / f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n),$$

$$\hat{r}_t^{(1)} := \nabla_{\theta^1} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) / f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n);$$

and unless confusion arises otherwise, for a given function of (π, θ_2), say q_t, $q_t(\pi, \hat{\theta}_{1,n}^n)$ is abbreviated as $\hat{q}_t$. Thus, for example, $g_t(\pi, \hat{\theta}_{1,n}^n)$ and $h_t(\hat{\theta}_{1,n}^n)$ will be denoted as $\hat{g}_t$ and $\hat{h}_t$ respectively.

We can use these to show that $\tilde{L}_n^{(1)}(\pi, \theta_2)$ is identically zero under H_{02}. Specifically, by the definitions of h_t and k_t, $\hat{h}_t := h_t(\hat{\theta}_{1,n}^n) = \hat{k}_t := k_t(\hat{\theta}_{1,n}^n)$, because $\hat{\theta}_{0,n}^a(\pi) = \tilde{\theta}_{0,n}(\hat{\theta}_{2,n}^a(\pi))$, $\hat{\theta}_{1,n}^a(\pi) = \tilde{\theta}_{1,n}(\hat{\theta}_{2,n}^a(\pi))$, $\hat{\theta}_{0,n}^n = \tilde{\theta}_{0,n}(\hat{\theta}_{1,n}^n)$ and $\hat{\theta}_{1,n}^n = \tilde{\theta}_{1,n}(\hat{\theta}_{1,n}^n)$ under the alternative and the null model assumptions respectively, where for each π,

$$\{\hat{\theta}_{0,n}^a(\pi), \hat{\theta}_{1,n}^a(\pi), \hat{\theta}_{2,n}^a(\pi)\} := \arg\max_{(\theta_0, \theta_1, \theta_2)} L_n(\pi, \theta_0, \theta_1, \theta_2).$$

Thus, it follows that

$$\tilde{L}_n^{(1)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi) \sum \hat{h}_t \hat{g}_t = (1-\pi) \sum \hat{k}_t \hat{g}_t = \tilde{K}_n^{(1)}(\pi, \hat{\theta}_{1,n}^n) = 0. \tag{4}$$

This aspect is important because the LR statistic can be constructed using $\tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)$, as well. Note that the LR statistic can be also computed as

$$\max_{(\pi, \theta_2)} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\},$$

and (4) implies that the first-order derivative evaluated under the null parameter estimate, $\hat{\theta}_{1,n}^n$, is identically zero.

Due to this, we have to approximate the log-likelihood function by much higher-order derivatives. For non-zero second-order derivative cases, appropriate expansion order turns out to be four. The prior differentiability order condition is therefore strengthened as follows.

A3 (DIFFERENTIATION): (*iii*) $f(X_t|X^{t-1};\cdot) \in \mathcal{C}^{(4)}(\tilde{\Theta})$ *a.s.*

For each π, the fourth-order approximation can be written as

$$\begin{aligned}\tilde{L}_n(\pi,\theta_2) = \tilde{L}_n(\pi,\hat{\theta}_{1,n}^n) &+ \frac{1}{2}n^{-1/2}\tilde{L}_n^{(2)}(\pi,\hat{\theta}_{1,n}^n)\{n^{1/4}(\theta_2-\hat{\theta}_{1,n}^n)\}^2 \\ &+ \frac{1}{3!}n^{-3/4}\tilde{L}_n^{(3)}(\pi,\hat{\theta}_{1,n}^n)\{n^{1/4}(\theta_2-\hat{\theta}_{1,n}^n)\}^3 \\ &+ \frac{1}{4!}n^{-1}\tilde{L}_n^{(4)}(\pi,\hat{\theta}_{1,n}^n)\{n^{1/4}(\theta_2-\hat{\theta}_{1,n}^n)\}^4 + o_p((\theta_2-\hat{\theta}_{1,n}^n)^4)\end{aligned} \tag{5}$$

under the regularity conditions given below, where the second to the fourth-order derivatives are

$$\tilde{L}_n^{(2)}(\pi,\hat{\theta}_{1,n}^n) = (1-\pi)\sum(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t = (1-\pi)(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}, \tag{6}$$

$$\tilde{L}_n^{(3)}(\pi,\hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t + 2(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(1)}\}, \tag{7}$$

$$\tilde{L}_n^{(4)}(\pi,\hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(3)}-\hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t^{(1)} + 3(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(2)}\}. \tag{8}$$

Note that these involve computing $\hat{\tilde{\theta}}_{i,n}^{(j)} := \tilde{\theta}_i^{(j)}(\hat{\theta}_{1,n}^n)$ ($i=0,1$, and $j=1,2,3$), as well, and the asymptotic behavior of each derivative determines the asymptotic distribution of the LR statistic.

Further regularity conditions are needed for the components constituting the right-hand side (RHS) of (5). Given the DGP and the fourth-order expansion in (5), the second and fourth-order derivatives are the components we can apply the conventional CLT and the SULLN respectively. This also implies that $\tilde{L}_n^{(3)}(\pi,\hat{\theta}_{1,n}^n)$ needs to be $o_p(n^{3/4})$ not to affect the asymptotic behaviors of the second and fourth-order derivatives. We provide regularity conditions for these below. Cho and White (2007) show that the asymptotic variance of the second-order derivative can be consistently estimated by applying the SULLN to the fourth-order derivative under the correct specification assumption, leading to the information matrix equality. Under our model assumptions, the information matrix equality doesn't hold, but we can still proceed with analysis parallel to correctly specified models.

We assume for the second-order derivative that the sequence

$$\left\{ z_t^{\{1\}} := \left(r_t^{(1)}(\theta_*)', r_t^{(0,2)}(\theta_*) \right)' \right\}$$

satisfies the NAMXG condition, where $r_t^{(i,j)}(\theta_*) := f_t^{(i,j)}(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)$, so that $\sum \hat{r}_t^{(0,2)} = O_p(n^{1/2})$ but not $\sum \hat{r}_t^{(0,2)} = o_p(n^{1/2})$. Also, note that $r_t^{(1)}(\theta_*) = \nabla_{\theta^1} \tilde{\ell}(\theta_0^*, \theta_*)$. If $\sum \hat{r}_t^{(0,2)} = O_p(n)$ but not $o_p(n)$, the LR statistic has a trivial asymptotic behavior. It is degenerate when $n^{-1} \sum \hat{r}_t^{(0,2)}$ converges to a negative number; otherwise, the LR statistic is not bounded in probability, implying that this occasion belongs to the alternative events. Further, if $\sum \hat{r}_t^{(0,2)} = o_p(n^{1/2})$, the second-order derivative cannot be the main component forming the asymptotic null distribution, and the next order derivative needs to be examined. The NAMXG condition avoids these aspects.

We assume for the third-order derivative that the sequence

$$\left\{ z_t^{\{2\}} := \left(z_t^{\{1\}}, r_t^{(0,3)}(\theta_*), r_t^{(1,1)}(\theta_*)' \right)' \right\}$$

satisfies the AMXG condition. This is required to have the third-order derivative not affecting the asymptotic behaviors of the second and fourth-order derivatives. Note that each element in the RHS of (7) is obtained as follows.

$$\sum (\hat{h}_t^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(1)} = (1 - \hat{\tilde{\theta}}_{1,n}^{(1)}) \sum \hat{r}_t^{(0,2)} \{ (-\pi \hat{\tilde{\theta}}_{1,n}^{(1)} + \pi - 1) \hat{r}_t^{(0,1)} - \hat{\tilde{\theta}}_{0,n}^{(1)\prime} \hat{r}_t^{(1,0)} \}, \tag{9}$$

$$\sum (\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t = \sum \{ -\hat{\tilde{\theta}}_{1,n}^{(2)} \hat{r}_t^{(0,2)} + (1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^2) \hat{r}_t^{(0,3)} + 2(1 - \hat{\tilde{\theta}}_{1,n}^{(1)}) \hat{\tilde{\theta}}_{0,n}^{(1)\prime} \hat{r}_t^{(1,2)} \}. \tag{10}$$

Given this, we show in the Appendix that $\sum \hat{r}_t^{(0,2)} = O_p(n^{1/2})$ implies $(\hat{\tilde{\theta}}_{0,n}^{(1)\prime}, \pi \hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi)' = O_p(n^{-1/2})$, so that (9) and (10) are $O_p(n^{1/2})$ by the AMXG condition and the SULLN, leading to that $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = o_p(n^{2/3})$.

We assume for the fourth-order derivative that for each π, $-\Xi^{(4)}(\pi) \in (0, \infty)$, where

$$\Xi^{(4)}(\pi) := \left[\frac{1-\pi}{\pi} \right] \left[\frac{1 - 3\pi + 3\pi^2}{\pi^2} \right] R_*^{(0,4)} - 3 \left[\frac{1-\pi}{\pi} \right]^2 \times$$
$$\left[R_*^{(0,2)(0,2)} - \begin{bmatrix} R_*^{(1,0)(0,2)} + R_*^{(1,2)} \\ R_*^{(0,1)(0,2)} \end{bmatrix}' \begin{bmatrix} R_*^{(1,0)(1,0)} - R_*^{(2,0)} & R_*^{(1,0)(0,1)\prime} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix}^{-1} \begin{bmatrix} R_*^{(1,0)(0,2)} + R_*^{(1,2)} \\ R_*^{(0,1)(0,2)} \end{bmatrix} \right],$$

$R_*^{(i,j)} := E[r_t^{(i,j)}(\theta_*)]$, and $R_*^{(i,j)(k,\ell)} := E[r_t^{(i,j)}(\theta_*) r_t^{(k,\ell)}(\theta_*)']$. This is what can be obtained by applying the SULLN to $n^{-1} \tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$, and it is assumed to have a locally concave expected log-likelihood function with respect to θ_2. This assumption also implies that $R_*^{(0,4)} \leq 0$, and the inverse matrix of $\Xi^{(4)}(\pi)$ exists. We recapitulate these assumptions into the following.

A7 (CLT): (*ii*) $\{z_t^{\{1\}}\}$ *and* $\{z_t^{\{2\}}\}$ *satisfy the NAMXG and the AMXG conditions respectively, and for each* $\pi \in (0,1)$, $-\Xi^{(4)}(\pi) \in (0,\infty)$.

Further regularity conditions are needed for the desired results. Given the moment conditions, we cannot apply the SULLN to the products of involved derivatives. For example, one of the components in $\Xi^{(4)}(\pi)$ is $R_*^{(0,2)(0,2)}$, which has to be estimated by $n^{-1}\sum\{\hat{r}_t^{(0,2)}\}^2$, but the prior moment conditions do not ensure the existence of its limit. We therefore strengthen the prior moment condition into the following.

A5 (MOMENTS) (*iv*) *For* $i_1, \cdots, i_4 \in \{\theta_{01}, \theta_{02}, \cdots, \theta_{0r_0}, \theta_1\}$,
(*a*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$;
(*b*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$;
(*c*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$;
(*d*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4} f_t(\theta^1)/f_t(\theta^1)| \leq M_t$.

Note that the highest moment order is four, and the fourth-order derivative is assumed to be uniformly bounded by M_t to ensure the SULLN.

Using these key conditions, each derivative can be shown to exhibit the following asymptotic behavior.

LEMMA 2: *Given A1, A2, A3*(*iii*), *A4, A5*(*iv*), *A7*(*ii*), *and* H_{02}, *for each* π,
(*i*) $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = (\frac{1-\pi}{\pi})\sum \hat{r}_t^{(0,2)} + o_p(n^{1/2})$;
(*ii*) $n^{-1/2}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) \Rightarrow (\frac{1-\pi}{\pi})G_1$, *where* $G_1 \sim N(0, V_1)$ *and* $V_1 := \text{avar}[n^{-1/2}\sum \hat{r}_t^{(0,2)}]$;
(*iii*) $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(*iv*) $n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = \Xi^{(4)}(\pi) + o_p(1)$.

There are numerous remarks relevant to Lemma 2. First, Lemma 2(iii) can be obtained by imposing other conditions. For example, we could impose $\sum \hat{r}_t^{(0,2)}\hat{r}_t^{(1,1)} = O_p(n^{1/2})$ and $(-\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + \pi - 1) = O_p(1)$ to have the first component in the RHS of (9) be $O_p(n^{1/2})$. But, these are not satisfied by correctly specified models, so that we cannot maintain the model analysis parallel to that for correctly specified models. Second, the unidentified parameter, π, is present as a coefficient function of $\sum \hat{r}_t^{(0,2)}$ as given in Lemma 2(i), implying that its tightness trivially holds, and leading to a trivial identification problem when compared to correctly specified models. Third, dynamic misspecification needs to be accommodated to compute the variance of G_1. Fourth, if all of $\sum \hat{r}_t^{(0,4)}$, $\sum \hat{r}_t^{(1,2)}$ and $\sum \hat{r}_t^{(2,0)}$ are $o_p(n)$ and dynamic misspecification is not involved, then $n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ estimates $-3\left[\frac{1-\pi}{\pi}\right]^2 \Omega^{(2)}$, where for $j = 1, 2, \cdots$,

$$\Omega^{(j)} := R_*^{(0,j)(0,j)} - \begin{bmatrix} R_*^{(1,0)(0,j)} \\ R_*^{(0,1)(0,j)} \end{bmatrix}' \begin{bmatrix} R_*^{(1,0)(1,0)} & R_*^{(1,0)(0,1)\prime} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix}^{-1} \begin{bmatrix} R_*^{(1,0)(0,j)} \\ R_*^{(0,1)(0,j)} \end{bmatrix}$$

whenever it exists. If models are correctly specified, $n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ also estimates this, and also it is the asymptotic variance of $n^{-1/2}\sum \hat{r}_t^{(0,2)}$. Thus, information matrix equality follows, though model correct specification is not implied by this.

Lemma 2 simplifies the asymptotic behavior of (5). For each π, $n^{-3/4}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = o_p(1)$ by Lemma 2(iii), so that (5) virtually reduces to

$$\begin{aligned} \tilde{L}_n(\pi, \theta_2) = \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n) &+ \frac{1}{2!} n^{-1/2}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/4}(\theta_2 - \hat{\theta}_{1,n}^n)\}^2 \\ &+ \frac{1}{4!} n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/4}(\theta_2 - \hat{\theta}_{1,n}^n)\}^4 + o_p(1), \end{aligned} \tag{11}$$

and for each π,

$$\sup_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow \sup_{\xi} \left[\frac{1-\pi}{\pi}\right] G_1 \xi^2 + \frac{1}{12}\Xi^{(4)}(\pi)\xi^4, \tag{12}$$

where ξ captures the asymptotic behavior of $n^{1/4}(\theta_2 - \hat{\theta}_{1,n}^n)$. Note that ξ^2 and ξ^4 cannot be less than zero, implying that the RHS of (12) is zero when G_1 is negative. Equation (12) can be further maximized with respect to π, which delivers the asymptotic null distribution of the LR statistic. Under our maintained assumptions, the following Theorem 3 can be obtained.

THEOREM 3: *Given A1, A2, A3(iii), A4, A5(iv), A6, A7(ii), and* H_{02}, *for each* $\epsilon \in (0, 1/2)$,
(*i*) *for each* π, $\max_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow 3[\frac{1-\pi}{\pi}]^2[-\Xi^{(4)}(\pi)]^{-1}\max[0, G_1]^2$;
(*ii*) $LR_n^{(2)}(\epsilon) := \max_{\pi\in(\epsilon,1-\epsilon)} \max_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow 3[-\Xi^{(4)}(1/2)]^{-1}\max[0, G_1]^2$.

There are several remarks. First, if the model is correctly specified, the square of half standard normal random variable is obtained as the weak limit of the LR statistic. This holds because of the information matrix equality. Second, the convergence rate of the LR statistic is $n^{1/4}$, which is different from the standard LR statistic. This is mainly because ξ in the RHS of (12) captures the asymptotic behavior of $n^{1/4}(\theta_2 - \hat{\theta}_{1,n}^n)$. Third, the probability limit of $\hat{\pi}_n^a$ may or may not exist, depending on the sign of G_1. If G_1 is negative, the RHS of (12) is attained when $\xi = 0$, and π does not matter for the maximum of (12). Otherwise, the maximum of $[\frac{1-\pi}{\pi}]^2[-\Xi^{(4)}(\pi)]$ is attained when $\pi = 1/2$, which is the main argument of Theorem 3(ii).

3.2.2 Zero Second-Order Derivative Case

For many other popular mixtures, we note that $f_t^{(0,2)}(\theta_0, \theta_1) = \alpha' f_t^{(1,0)}(\theta_0, \theta_1) + \beta f_t^{(0,1)}(\theta_0, \theta_1)$ for each (θ_0, θ_1) and for some non-zero $(\alpha', \beta)' \in \mathbb{R}^{r_0+1}$, so that

$$\sum \hat{r}_t^{(0,2)} = \alpha' \sum \hat{r}_t^{(1,0)} + \beta \sum \hat{r}_t^{(0,1)} = 0, \tag{13}$$

implying that $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = 0$ by Lemma 2(i). In this case, $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)$ cannot be exploited to approximate the relevant likelihood functions, either. Much higher-order approximation is necessary. We examine the higher-order derivatives combined with this additional condition (13).

By (13), stronger results can be obtained. For each π, we show in the Appendix that $(\pi\hat{\hat{\theta}}_{1,n}^{(1)} + 1 - \pi) = 0$ and $\hat{\hat{\theta}}_{0,n}^{(1)} = 0$, so that

$$\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = -\left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right] \sum \hat{r}_t^{(0,3)} \tag{14}$$

after substituting these to (9) and (10). Thus, the third-order derivative can constitute the asymptotic null distribution, so that a sixth-order Taylor expansion can be exploited. Also, the fourth and fifth-order derivatives need to be negligible in a way for the third and sixth-order derivatives to determine the asymptotic null distribution of the LR test statistic. More specifically, we note that

$$\begin{aligned} \tilde{L}_n(\pi, \theta_2) = \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n) &+ \frac{1}{3!} n^{-1/2} \tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/6}(\theta_2 - \hat{\theta}_{1,n}^n)\}^3 \\ &+ \frac{1}{4!} n^{-2/3} \tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/6}(\theta_2 - \hat{\theta}_{1,n}^n)\}^4 \\ &+ \frac{1}{5!} n^{-5/6} \tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/6}(\theta_2 - \hat{\theta}_{1,n}^n)\}^5 \\ &+ \frac{1}{6!} n^{-1} \tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n)\{n^{1/6}(\theta_2 - \hat{\theta}_{1,n}^n)\}^6 + o_p((\theta_2 - \hat{\theta}_{1,n}^n)^6) \end{aligned} \tag{15}$$

under the regularity conditions provided below. As before, the conventional CLT and the SULLN are applied to the third and sixth-order derivatives respectively, and if the fourth and fifth-order derivatives are negligible under the hypothesis, then the desired property follows. The regularity conditions for these are provided below.

With this approximation, nevertheless, there are two further important remarks we need to pay attention to. First, the sixth-order expansion cannot approximate the log-likelihood function for every π. If $\pi = 1/2$, $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = 0$ by (14), so that further higher-order approximation is needed. The eighth-order expansion turns out to be relevant for this case, and accordingly ordered further regularity conditions need to be provided. Second, the desired non-degeneracy of the LR statistic at $\pi = 1/2$ is not necessary to have non-degenerate LR statistic. The LR statistic can be still non-degenerate by its non-degeneracy at $\pi \neq 1/2$. We thus consider two cases separately: (*i*) degenerate LR statistic at $\pi = 1/2$; and (*ii*) non-degenerate LR statistic at $\pi = 1/2$. If models are correctly specified, we cannot assume the first case.

(*i*) WHEN LR STATISTIC IS DEGENERATE AT $\pi = 1/2$. By (15), much stronger order of differentiability is imposed.

A3 (DIFFERENTIATION): (*iv*) $f(X_t|X^{t-1};\cdot) \in \mathcal{C}^{(6)}(\tilde{\Theta})$ *a.s., and for some* $(\alpha', \beta)' \in \mathbb{R}^{r_0+1}$ ($\neq 0$), $f_t^{(0,2)} = \alpha' f_t^{(1,0)} + \beta f_t^{(0,1)}$.

Further, the LR statistic is non-degenerate (resp. degenerate) at $\pi \neq 1/2$ (resp. $\pi = 1/2$) if the fourth and fifth-order derivatives are $O_p(n^{1/2})$, and the sixth-order derivative converges to a negative value for every π for the concavity of the log-likelihood function. Our regularity conditions considered below attain these.

For the third-order derivative, we assume that

$$\left\{ w_t^{\{1\}} := \left(r_t^{(1)}(\theta_*)', r_t^{(0,3)}(\theta_*) \right)' \right\}$$

satisfies the NAMXG condition. It is clear from (14) that $\sum \hat{r}_t^{(0,3)}$ is the main component constituting the asymptotic null distribution of the LR statistic. If $n^{-1} \sum \hat{r}_t^{(0,3)}$ converges to a positive real number, then the LR statistic is not bounded in probability, thus the null behavior cannot be captured. On the other hand, if it converges to a negative real number, the LR statistic is degenerate. We thus need that $\sum \hat{r}_t^{(0,3)} = O_p(n^{1/2})$ but not $o_p(n^{1/2})$. Our assumption on $\{w_t^{\{1\}}\}$ attains this.

For $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ and $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n)$, we assume that

$$\left\{ w_t^{\{2\}} := \left(w_t^{\{1\}\prime}, r_t^{(0,4)}(\theta_*), r_t^{(0,5)}(\theta_*), r_t^{(1,1)}(\theta_*)', r_t^{(1,2)}(\theta_*)', r_t^{(1,3)}(\theta_*)', \text{vech}(r_t^{(2,0)}(\theta_*))' \right)' \right\}$$

satisfies the AMXG condition. This assumption virtually implies that all of $\sum \hat{r}_t^{(1,1)}$, $\sum \hat{r}_t^{(2,0)}$, $\sum \hat{r}_t^{(1,2)}$, $\sum \hat{r}_t^{(0,4)}$, $\sum \hat{r}_t^{(1,3)}$ and $\sum \hat{r}_t^{(0,5)}$ are $O_p(n^{1/2})$ and also that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ and $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n)$ are $O_p(n^{1/2})$, because they are the main components constituting these derivatives. They are also implied by correctly specified models, but the reverse does not hold. Note that $r_t^{(2,0)}(\theta_*)$ is an $r_0 \times r_0$ symmetric random matrix, so that the 'vech' operator is used for $\{w_t^{\{2\}}\}$ to avoid counting duplicated elements.

We assume for the sixth-order derivative that for each π,

$$\begin{aligned}\Xi^{(6)}(\pi) :=& -10\left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right]^2\Omega^{(3)} \\ &+\left[\frac{1-\pi}{\pi}\right]\left[\frac{1-5\pi+10\pi^2-10\pi^3+5\pi^4}{\pi^4}\right]R_*^{(0,6)} \\ &-3\left[\frac{1-\pi}{\pi}\right]^2\left[\frac{1-3\pi+3\pi^2}{\pi^2}\right]\alpha' R_*^{(1,4)}+45\left[\frac{1-\pi}{\pi}\right]^3\alpha' R_*^{(2,2)}\alpha \\ &+5\left[\frac{1-\pi}{\pi}\right]^2\left[\pi-3\beta\left(\frac{1-\pi}{\pi}\right)\right]\alpha' R_*^{(2,1)}\alpha-15\left[\frac{1-\pi}{\pi}\right]^3\alpha' M_*^{(1,0)}<0,\end{aligned} \tag{16}$$

where $M_*^{(k,\ell)} := E[\{\nabla^k_{\theta_0}\nabla^\ell_{\theta_1}[\alpha' f_t^{(2,0)}(\theta_0^*,\theta_*)\alpha]\}/f_t(\theta_0^*,\theta_*)]$. Note that $M_*^{(1,\ell)}$ and $M_*^{(2,\ell)}$ are an $r_0\times 1$ vector and an $r_0\times r_0$ matrix respectively. In particular, we devise these to avoid a tensor notation, and $\Xi^{(6)}(\pi)$ can be estimated consistently by applying the SULLN to the sixth-order derivative. If models are correctly specified, it is identical to $-10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2\Omega^{(3)}$, because $R_*^{(i,j)}$ and $M_*^{(1,0)}$ vanish to zero. We also assume that $\Xi^{(6)}(1/2)$ is negative, which leads to a degenerate LR statistic at $\pi=1/2$. These conditions can be recapitulated as follows.

A7 (CLT): (*iii*) $\{w_t^{\{1\}}\}$ *and* $\{w_t^{\{2\}}\}$ *satisfy the NAMXG and the AMXG conditions respectively, and for each* $\pi\in(0,1)$, $-\Xi^{(6)}(\pi)\in(0,\infty)$.

In addition to this, the previous moment conditions need to be strengthened to apply the SULLN.

A5 (MOMENTS) (*v*) *For* $i_1,\cdots,i_6\in\{\theta_{01},\theta_{02},\cdots,\theta_{0r_0},\theta_1\}$,
(*a*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}f_t(\theta^1)/f_t(\theta^1)|^4\le M_t$;
(*b*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}\nabla_{i_2}f_t(\theta^1)/f_t(\theta^1)|^4\le M_t$;
(*c*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}f_t(\theta^1)/f_t(\theta^1)|^4\le M_t$;
(*d*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}\cdots\nabla_{i_4}f_t(\theta^1)/f_t(\theta^1)|^2\le M_t$;
(*e*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}\cdots\nabla_{i_5}f_t(\theta^1)/f_t(\theta^1)|^2\le M_t$;
(*f*) $\sup_{\theta^1\in\tilde\Theta}|\nabla_{i_1}\cdots\nabla_{i_6}f_t(\theta^1)/f_t(\theta^1)|\le M_t$.

By these additional regularity conditions, the following asymptotic behaviors are obtained.

LEMMA 3: *Given A1, A2, A3(iv), A4, A5(v), A7(iii) and* H_{02},
(*i*) $n^{-1/2}\tilde L_n^{(3)}(\pi,\hat\theta_{1,n}^n)\Rightarrow[-\frac{(1-\pi)(1-2\pi)}{\pi^2}]G_2$, *where* $G_2\sim N(0,V_2)$ *and* $V_2:=\text{avar}[n^{-1/2}\sum\hat r_t^{(0,3)}]$;
(*ii*) $\tilde L_n^{(4)}(\pi,\hat\theta_{1,n}^n)=O_p(n^{1/2})$;
(*iii*) $\tilde L_n^{(5)}(\pi,\hat\theta_{1,n}^n)=O_p(n^{1/2})$;

(*iv*) $n^{-1}\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = \Xi^{(6)}(\pi) + o_p(1)$.

As Lemma 3 is parallel to Lemma 2 in terms of its contents, we do not iterate our explanation. Instead, we directly obtain the asymptotic null distribution of the LR statistic. For each π, Lemmas 3(ii and iii) reduce (15) to

$$\tilde{L}_n(\pi, \theta_2) = \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n) + \frac{1}{3!}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^3 + \frac{1}{6!}\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^6 + o_p(1),$$

and for each π, by Lemma 3(i and iv)

$$\sup_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow \sup_{\xi} \frac{2}{3!}\left[-\frac{(1-\pi)(1-2\pi)}{\pi^2}\right] G_2\xi^3 + \frac{2}{6!}\Xi^{(6)}(\pi)\xi^6. \tag{17}$$

There is no sign condition in (17) as for (12), because ξ^3 can be negative. Given this, the asymptotic null distribution can be obtained by maximizing (17) with respect to π. Formally, we can state our main claims as follows.

THEOREM 4: *Given A1, A2, A3(iv), A4, A5(v), A7(iii) and* H_{02},
(*i*) *for each* π, $\max_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow 10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[-\Xi^{(6)}(\pi)]^{-1}G_2^2$;
(*ii*) $LR_n^{(2)}(\epsilon) \Rightarrow 10[\frac{(1-\pi^\ddagger)(1-2\pi^\ddagger)}{(\pi^\ddagger)^2}]^2[-\Xi^{(6)}(\pi^\ddagger)]^{-1}G_2^2$, *where* $\pi^\ddagger := \arg\min[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[\Xi^{(6)}(\pi)]^{-1}$.

As a remark relevant to Theorem 4, the LR statistic is degenerate when $\pi = 1/2$. It's mainly because $\Xi^{(6)}(1/2) < 0$, and the third-order derivative is identical to zero, so that the RHS of (17) is $\frac{2}{6!}\Xi^{(6)}(\pi)\xi^6$, whose maximum is zero when $\xi = 0$. Thus, $\pi^\ddagger$ cannot be $1/2$.

(*ii*) WHEN LR STATISTIC IS NOT DEGENERATE AT $\pi = 1/2$. If the sixth and seventh-order derivatives are $O_p(n^{1/2})$ at $\pi = 1/2$, the LR statistic can be non-degenerate at $\pi = 1/2$ by requiring that log-likelihood functions can be approximated by an eighth-order expansion. Also, this entails that the derivatives from the fifth to the seventh-orders at $\pi = 1/2$ need to be negligible not to affect the fourth and eighth-order derivatives. We provide regularity conditions for these.

First of all, the order of differentiability has to be strengthened, and the moment condition also has to reach the eighth-order derivatives. These can be provided as follows.

A3 (DIFFERENTIATION): (*v*) $f(X_t|X^{t-1};\cdot) \in \mathcal{C}^{(8)}(\tilde{\Theta})$ *a.s., and for some* $(\alpha', \beta)' \in \mathbb{R}^{r_0+1}$ ($\neq 0$), $f_t^{(0,2)} = \alpha' f_t^{(1,0)} + \beta f_t^{(0,1)}$.

A5 (MOMENTS) (*vi*) *For* $i_1, \cdots, i_8 \in \{\theta_{01}, \theta_{02}, \cdots, \theta_{0r_0}, \theta_1\}$, (*a*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$;

(*b*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \nabla_{i_2} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$;

(*c*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \nabla_{i_2} \nabla_{i_3} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$;

(*d*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \cdots \nabla_{i_4} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$;

(*e*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \cdots \nabla_{i_5} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$;

(*f*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \cdots \nabla_{i_6} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$;

(*g*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \cdots \nabla_{i_7} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$;

(*h*) $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} \cdots \nabla_{i_8} f_t(\theta^1)/f_t(\theta^1)| \leq M_t$.

Note that the highest moment order is four. The additional condition (13) gives rise to zero coefficients for the moments greater than four. We don't have to pay attention to the moment orders greater than four. Nevertheless, the fourth-order derivative is assumed to have a finite fourth-order moment, which is stronger than A5(v), and as before, the final-order derivative is assumed to have a finite first-order moment.

We assume for the fourth-order derivative that

$$\left\{ w_t^{\{3\}} := \left(w_t^{\{1\}\prime}, s_t \right)' \right\}$$

satisfies the NAMXG condition, where

$$s_t := r_t^{(0,4)}(\theta_*) - 6\beta r_t^{(0,3)}(\theta_*) + 6\alpha' r_t^{(1,1)}(\theta_*)\beta - 6\alpha' r_t^{(1,2)}(\theta_*) + 3\alpha' r_t^{(2,0)}(\theta_*)\alpha.$$

This assumption is motivated from the notice that

$$\begin{aligned} \tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n) = & \sum \left\{ 3 \left(\hat{\tilde{\theta}}_{0,n}^{(2)} - \alpha \right)' \hat{r}_t^{(1,2)} - \left(3\hat{\tilde{\theta}}_{0,n}^{(2)} \right)' \hat{r}_t^{(2,0)} \alpha \right\} \\ & + \sum \left\{ \hat{r}_t^{(0,4)} + 3 \left(0.5\hat{\tilde{\theta}}_{1,n}^{(2)} - \beta \right) \hat{r}_t^{(0,3)} - 3 \left(0.5\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} + \beta\hat{\tilde{\theta}}_{0,n}^{(2)} \right)' \hat{r}_t^{(1,1)} \right\} \end{aligned} \tag{18}$$

using (8) and the facts that $(\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi) = 0$ and $\hat{\tilde{\theta}}_{0,n}^{(1)} = 0$. In the Appendix, we show that $\hat{\tilde{\theta}}_{0,n}^{(2)} + \alpha = o_p(1)$ and $\hat{\tilde{\theta}}_{1,n}^{(2)} + 2\beta = o_p(1)$. Thus, it's not hard to derive from these and the AMXG condition for $\{w_t^{\{2\}}\}$ that $\tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n) = \sum \hat{s}_t + o_p(n^{1/2}) = O_p(n^{1/2})$, where

$$\hat{s}_t := \hat{r}_t^{(0,4)} - 6\beta\hat{r}_t^{(0,3)} + 6\alpha'\hat{r}_t^{(1,1)}\beta - 6\alpha'\hat{r}_t^{(1,2)} + 3\alpha'\hat{r}_t^{(2,0)}\alpha.$$

Nevertheless, $n^{-1/2} \sum \hat{s}_t$ can be degenerate, and to prevent this, we assume that $\{s_t\}$ satisfies the NAMXG condition. Also, $\{w_t^{\{1\}}\}$ is included in $\{w_t^{\{3\}}\}$ to approximate the log-likelihood functions at $\pi \neq 1/2$ by the sixth-order expansion.

We assume for the fifth to the seventh-order derivatives that

$$\left\{w_t^{\{4\}} := \left(s_t, w_t^{\{2\}}, r_t^{(0,6)}(\theta_*), r_t^{(1,4)}(\theta_*)', \text{vech}(r_t^{(2,1)}(\theta_*))', \text{vech}(r_t^{(2,2)}(\theta_*))', \text{wech}(r_t^{(3,0)}(\theta_*))'\right)'\right\}$$

satisfies the AMXG condition, where

$$\text{wech}(r_t^{(3,0)}(\theta_*)) = \frac{1}{f_t(\theta_0^*, \theta_*)}\left[\frac{\partial^3}{\partial^3\theta_{01}^3} f_t(\theta_0^*, \theta_*) \cdots \frac{\partial^3}{\partial\theta_{0i}\partial\theta_{0j}\partial\theta_{0k}} f_t(\theta_0^*, \theta_*) \cdots \frac{\partial^3}{\partial^3\theta_{0r_0}^3} f_t(\theta_0^*, \theta_*)\right]$$

such that $1 \leq i \leq j \leq k \leq r_0$. The 'wech' operator is devised not to count duplicated elements in a symmetric three-dimensional matrix. By this, it follows that all of $\sum \hat{r}_t^{(0,6)}$, $\sum \hat{r}_t^{(1,4)}$, $\sum \hat{r}_t^{(2,1)}$, $\sum \hat{r}_t^{(2,2)}$ and $\sum \hat{r}_t^{(3,0)}$ are $O_p(n^{1/2})$. Further, they are the components forming the fifth to the seventh-order derivatives, so that $\tilde{L}_n^{(5)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$, $\tilde{L}_n^{(6)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$ and $\tilde{L}_n^{(7)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$ follow.

Finally, we assume for the eighth-order derivative that

$$\begin{aligned}\Xi^{(8)}(1/2) :=& R_*^{(0,8)} - 28R_*^{(0,7)}\beta - 35\Omega^{(s)}\\ &+ 367.5\alpha' R_*^{(1,5)}\beta - 28\alpha' R_*^{(1,6)} + 420\alpha'(M_*^{(1,1)}\beta - M_*^{(1,2)})\\ &- 1260\alpha' R_*^{(2,3)}\alpha\beta + 210\alpha' R_*^{(2,4)}\alpha + 105\alpha' M_*^{(2,0)}\alpha < 0,\end{aligned}$$

where

$$\Omega^{(s)} := E\left[s_t^2\right] - \begin{bmatrix} E[s_t \cdot r_t^{(1,0)}(\theta_*)] \\ E[s_t \cdot r_t^{(0,1)}(\theta_*)] \end{bmatrix}' \begin{bmatrix} R_*^{(1,0)(1,0)} & R_*^{(1,0)(0,1)'} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix}^{-1} \begin{bmatrix} E[s_t \cdot r_t^{(1,0)}(\theta_*)] \\ E[s_t \cdot r_t^{(0,1)}(\theta_*)] \end{bmatrix}.$$

Applying the SULLN to the eighth-order derivative estimates $\Xi^{(8)}(1/2)$, and $\Omega^{(s)}$ is the variance of the regression errors obtained by regressing s_t against $r_t^{(1)}(\theta_*)$. Also, if models are correctly specified, $\Omega^{(s)}$ is the asymptotic variance of the fourth-order derivative, and information matrix equality follows from this. We repeat these conditions in a compact way as follows.

A7 (CLT): (*iv*) $\{w_t^{\{3\}}\}$ *and* $\{w_t^{\{4\}}\}$ *satisfy the NAMXG and the AMXG conditions respectively, and for each* $\pi \in (0, 1)$, $-\Xi^{(8)}(\pi) \in (0, \infty)$.

By A7(iv), the sixth-order derivative has a different asymptotic behavior from the prior case. Note that the AMXG condition for $\{w_t^{\{4\}}\}$ implies that all the components not involving $\Omega^{(3)}$ in (16) are zeros, so that $\Xi^{(6)}(\pi) = -10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2\Omega^{(3)}$.

Using these regularity conditions, we can provide the asymptotic behavior of each derivative as follows.

LEMMA 4: *Given A1*, *A2*, *A3*(*v*), *A4*, *A5*(*vi*), *A7*(*iv*), *and* H_{02},

(*i*) $\tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n) = \sum \hat{s}_t + o_p(n^{1/2})$;
(*ii*) $n^{-1/2}\tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n) \Rightarrow G_3$, *where* $G_3 \sim N(0, V_3)$ *and* $V_3 := \text{avar}[n^{-1/2}\sum \hat{s}_t]$;
(*iii*) $\tilde{L}_n^{(5)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(*iv*) $\tilde{L}_n^{(6)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(*v*) $\tilde{L}_n^{(7)}(1/2, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(*vi*) $n^{-1}\tilde{L}_n^{(8)}(1/2, \hat{\theta}_{1,n}^n) = \Xi^{(8)}(1/2) + o_p(1)$.

With regard to Lemma 4, if $\pi \neq 1/2$, the same conclusion is obtained as for Lemma 3, because the maintained assumptions for Lemma 4 are stronger than those for Lemma 3. As another remark, applying the SULLN to the sixth-order derivative for $\pi \neq 1/2$ may estimate the asymptotic variance of the third-order derivative and result in information matrix equality unless dynamic misspecification is involved.

Given Lemma 4, it is now straightforward to derive the asymptotic null distribution of the LR statistic. Proving procedures are similar to the prior cases. That is, if $\pi = 1/2$, we can write the log-likelihood function as

$$\tilde{L}_n(1/2, \theta_2) = \tilde{L}_n(1/2, \hat{\theta}_{1,n}^n) + \frac{1}{4!}\tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^4 + \frac{1}{8!}\tilde{L}_n^{(8)}(1/2, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^8 + o_p(1)$$

by Lemmas 4(iii, iv, and v), so that

$$\sup_{\theta_2} 2\{\tilde{L}_n(1/2, \theta_2) - \tilde{L}_n(1/2, \hat{\theta}_{1,n}^n)\} \Rightarrow \sup_{\xi} \frac{2}{4!}G_3\xi^4 + \frac{2}{8!}\Xi^{(8)}(1/2)\xi^8 = 35[-\Xi^{(8)}(1/2)]^{-1}\max[0, G_3]^2.$$

If $\pi \neq 1/2$, then

$$\sup_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(1/2, \hat{\theta}_{1,n}^n)\} \Rightarrow [\Omega^{(3)}]^{-1}G_2^2$$

by Theorem 4(ii) and the fact that $\Xi^{(6)}(\pi) = -10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2\Omega^{(3)}$. Thus, the nuisance parameter π is free in this case, and the overall LR statistic has the asymptotic distribution equal to that of

$$\max\left\{[\Omega^{(3)}]^{-1}G_2^2,\ 35[-\Xi^{(8)}(1/2)]^{-1}\max[0, G_3]^2\right\}$$

under H_{02}. Further, G_2 and G_3 can be dependent. Their asymptotic covariance needs to accommodate dynamic misspecifications as well. Formally, the null asymptotic distribution can be given as Theorem 5.

THEOREM 5: *Given A1, A2, A3*(*v*), *A4*, *A5*(*vi*), *A7*(*iv*), *and* H_{02},
(*i*) *for each* $\pi \neq 1/2$, $\max_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow [\Omega^{(3)}]^{-1}G_2^2$;
(*ii*) *for* $\pi = 1/2$, $\max_{\theta_2} 2\{\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)\} \Rightarrow 35[-\Xi^{(8)}(1/2)]^{-1}\max[0, G_3]^2$;
(*iii*) $LR_n^{(2)}(\epsilon) \Rightarrow \max\{[\Omega^{(3)}]^{-1}G_2^2, 35[-\Xi^{(8)}(1/2)]^{-1}\max[0, G_3]^2\}$ *such that* $(G_2, G_3)' \sim N(0, \Sigma)$, *where* Σ *is the asymptotic covariance matrix between* $n^{-1/2}\sum \hat{r}_t^{(0,3)}$ *and* $n^{-1/2}\sum \hat{s}_t$.

3.3 Null Distribution of the LR Statistic under H_0

Given the asymptotic null distributions under H_{01} (or H_{03}) and H_{02}, we can combine them to obtain the asymptotic distribution under H_0. Mainly, as the null distribution is degenerate under H_{01} (or H_{03}), the distribution under H_0 is identical to that under H_{02}. This holds by the converging-together lemma. The following Corollary contains the key results of this paper.

COROLLARY: (*i*) *Given A1*, *A2*, *A3*(*iii*), *A4*, *A5*(*iv*), *A6*, *A7*(*ii*), *and* H_0,

$$LR_n \Rightarrow 3[-\Xi^{(4)}(1/2)]^{-1}\max[0, G_1]^2.$$

(*ii*) *Given A1*, *A2*, *A3*(*iv*), *A4*, *A5*(*v*), *A6*, *A7*(*iii*), *and* H_0,

$$LR_n \Rightarrow -10\left[\frac{(1-\pi^{\ddagger})(1-2\pi^{\ddagger})}{(\pi^{\ddagger})^2}\right]^2[\Xi^{(6)}(\pi^{\ddagger})]^{-1}G_2^2.$$

(*iii*) *Given A1*, *A2*, *A3*(*v*), *A4*, *A5*(*vi*), *A6*, *A7*(*iv*), *and* H_0,

$$LR_n \Rightarrow \max\left\{[\Omega^{(3)}]^{-1}G_2^2,\ 35[-\Xi^{(8)}(1/2)]^{-1}\max[0, G_3]^2\right\}.$$

The above Corollary is practically useful because the non-degenerate distributions are obtained by imposing mild conditions. Unless these are satisfied, the LR statistic is degenerate under the null, and so it is reasonable to apply the critical values based upon these as conservative critical values. Also, it can be problematic not to be aware of whether the LR statistic is degenerate or not at $\pi = 1/2$ to be able to apply the given Corollary. We recommend testing first whether $\Xi^{(6)}(1/2) = 0$ or not to choose the relevant critical values given in Corollaries (ii) and (iii). Finally, when models are presumably misspecified, it is also a practically useful aspect not to have to obtain critical values involving Gaussian processes. In general, it is indeed another challenge to obtain the exact distributions of Gaussian processes unless their covariance structures are trivial.

4 Conclusion

Mixture models are very popular in empirical applications, and also misspecified model analysis is often carried out in empirical applications in a way to accommodate information matrix inequality. In this paper, we examine how the misspecified model assumption can affect the asymptotic null distribution of the LR statistic designed to test for a mixture hypothesis.

The main results of this paper can be restated as follows. First, the asymptotic null distribution of the LR statistic for misspecified models cannot be interpreted in a way to conciliate that of correctly specified models. It's mainly because the asymptotic null distribution of the LR statistic cannot be given as a functional of a Gaussian process as obtained for correctly specified models. This is mainly because of the exponential family assumption we adopt to contrast our results with that for correctly specified models. The boundary parameter problem combined with the exponential family assumption prevents model analysis from being similar to that for correctly specified models. Second, we provide regularity conditions to have non-degenerate LR statistic. The consequence of these regularity conditions is that the LR statistic weakly converges to the square of half-normal random variable, non-central chi-square random variable or the maximum of these. It depends on the model conditions we encounter and the DGPs under which our regularity conditions are sustained.

5 Appendix: Proofs

First, we provide supplementary claims used to prove the main claims.

LEMMA A1: (*i*) *Given A1*, *A2*, *A3*(*i*) *and A5*(*i*), $\sup_{(\pi,\theta)} |n^{-1}\sum \ell_t(\pi,\theta) - E[\ell_t(\pi,\theta)]| \to 0$ *a.s.*
(*ii*) *Given A1*, *A2*, *A3*(*i*) *and A5*(*i*), $\sup_{(\pi,\theta)} \|n^{-1}\sum \nabla_{(\pi,\theta)}\ell_t(\pi,\theta) - E[\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)]\|_\infty \to 0$ *a.s.*

Proof of Lemma A1: (*i and ii*) See Cho and White (2007, Lemma A1). ■

Proof of Lemma 1: (*i*) By the conditions, $E[\nabla_\pi \ell_t(1,\theta_0^\star,\theta_\star,\theta_2)|X^{t-1}] = 1 - \int f_t(\theta_0^\star,\theta_2)dx_t = 0$.
(*ii*) Using the exponential family and the first–order condition, it follows that

$$
\begin{aligned}
& E[\exp\{\sum_{i=1}^k [a_i(X^{t-1};\theta_0^\star,\theta_2) - a_i(X^{t-1};\theta_0^\star,\theta_\star)]b_i(X_t)\}|X^{t-1}] \\
&= \exp[c(X^{t-1};\theta_0^\star,\theta_\star) - c(X^{t-1};\theta_0^\star,\theta_2)] \\
&= \int d(x)\exp[\sum_{i=1}^k a_i(X^{t-1};\theta_0^\star,\theta_2)b_i(x) + c(X^{t-1};\theta_0^\star,\theta_\star)]dx \\
&= \int d(x)\exp\{\sum_{i=1}^k [a_i(X^{t-1};\theta_0^\star,\theta_2) - a_i(X^{t-1};\theta_0^\star,\theta_\star) + a_i(X^{t-1};\theta_0^\star,\theta_\star)]b_i(x) + c(X^{t-1};\theta_0^\star,\theta_\star)\}dx \\
&= \int \exp\{\sum_{i=1}^k [a_i(X^{t-1};\theta_0^\star,\theta_2) - a_i(X^{t-1};\theta_0^\star,\theta_\star)]b_i(x)\}f_t(\theta_0^\star,\theta_\star)dx.
\end{aligned}
$$

Thus, it follows that $E[\exp[\sum_{i=1}^{k} \tau_i(X^{t-1})\, b_i(X_t)]|X^{t-1}] = \int \exp[\sum_{i=1}^{k} \tau_i(X^{t-1})\, b_i(x)] f_t(\theta_0^\star, \theta_\star) dx$, where $\tau_i(X^{t-1}) := a_i(X^{t-1}; \theta_0^\star, \theta_2) - a_i(X^{t-1}; \theta_0^\star, \theta_\star)$. That is, the conditional moment generating function of $(b_1(X_t) \cdots b_k(X_t)|X^{t-1})$ is the same as that generated by $f_t(\theta_0^\star, \theta_\star)$. Further, by A2, there is a function, $T : \mathbb{R}^k \mapsto \mathbb{R}$ such that $T(b_1(x), \cdots, b_k(x)) = x$, thus the conditional distribution of $X_t|X^{t-1}$ must be $F(\,\cdot\,|X^{t-1}, \theta_0^\star, \theta_\star)$. ■

Proof of Theorem 2: (*i*) Given the SULLN in Lemma A1 and H_0, it follows that for all $\omega \in F$, $P(F) = 1$, and $\varepsilon > 0$, there exists $n_1^*(\omega, \varepsilon)$ such that if $n > n_1^*(\omega, \varepsilon)$, then $(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) \in I_1(\varepsilon) \cup I_2(\varepsilon) \cup I_3(\varepsilon)$, where $I_1(\varepsilon) := \{(\pi, \theta_0, \theta_1, \theta_2) \in [0,1] \times \Theta : \|(\pi, \theta_0, \theta_1) - (1, \theta_0^*, \theta_*)\| < \varepsilon\} \setminus I_2(\varepsilon)$, $I_2(\varepsilon) := \{(\pi, \theta_0, \theta_1, \theta_2) \in [0,1] \times \Theta : \|(\theta_0, \theta_1, \theta_2) - (\theta_0^*, \theta_*, \theta_*)\| < \varepsilon\}$ and $I_3(\varepsilon) := \{(\pi, \theta_0, \theta_1, \theta_2) \in [0,1] \times \Theta : \|(\pi, \theta_0, \theta_2) - (0, \theta_0^*, \theta_*)\| < \varepsilon\} \setminus I_2(\varepsilon)$.

From the differentiability in A3(*i*), $E[\nabla_\pi \ell_t]$ is continuous on $[0,1] \times \Theta$ by Ranga Rao (1962), and $E[\nabla_\pi \ell_t(\pi, \theta)|_{(\pi=0)}] < 0$ and $E[\nabla_\pi \ell_t(\pi, \theta)|_{(\pi=1)}] > 0$ on $\Theta_* \setminus \{\theta_*\}$ by A6, so that we can choose δ to make $E[\nabla_\pi \ell_t]$ be uniformly positive and negative on $I_1(\delta)$ and $I_3(\delta)$ respectively.

For given δ, we let $K(\delta) := \inf_{I_1(\delta)} E[\nabla_\pi \ell_t(\pi, \theta)]$ and apply the SULLN to $\{n^{-1} \sum_{t=1}^{n} \nabla_\pi \ell_t\}$. Then for any $\omega \in F'$, $P(F') = 1$, there is an $n_2^*(\omega, K(\delta))$ such that if $n > n_2^*(\omega, K(\delta))$ then

$$\sup_{(\pi,\theta) \in I_1(\delta)} |n^{-1} \sum \nabla_\pi \ell_t(\pi, \theta) - E[\nabla_\pi \ell_t(\pi, \theta)]| < K(\delta)/2,$$

implying that $n^{-1} \sum \nabla_\pi \ell_t(\pi, \theta) > K(\delta)/2$ uniformly on $I_1(\delta)$. Hence,

$$\sup_{(\pi,\theta) \in I_1(\delta)} n^{-1} \sum \ell_t(\pi, \theta) = \sup_{(\pi,\theta) \in I_1(\delta)} n^{-1} \sum \ell_t(1, \theta) = \sup_{(\pi,\theta) \in I_1(\delta)} n^{-1} \sum \log(f_t(\theta^1)).$$

Likewise, using a similar argument, $\sup_{(\pi,\theta) \in I_3(\delta)} n^{-1} \sum \ell_t(\pi, \theta) = \sup_{(\pi,\theta) \in I_3(\delta)} n^{-1} \sum \log(f_t(\theta^2))$, which is identical to $\sup_{(\pi,\theta) \in I_1(\delta)} n^{-1} \sum \log(f_t(\theta^1))$.

Finally, if we let ε be equal to δ and denote it as η then it follows that for all $\omega \in F \bigcap F'$, $P(F \bigcap F') = 1$, $(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) \in I_1(\eta) \cup I_2(\eta) \cup I_3(\eta)$, and $(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) \notin I_1(\eta) \cup I_3(\eta)$, because

$$\begin{aligned} \sup_{(\pi,\theta) \in I_1(\eta)} n^{-1} \sum_{t=1}^{n} \log(f_t(\theta^1)) &= \sup_{(\pi,\theta) \in I_3(\eta)} n^{-1} \sum_{t=1}^{n} \log(f_t(\theta^2)) \\ &= \sup_{(\pi,\theta) \in I_2(\delta) : \theta_1 = \theta_2} n^{-1} \sum_{t=1}^{n} \ell_t(\pi, \theta) \leq \sup_{(\pi,\theta) \in I_2(\delta)} n^{-1} \sum_{t=1}^{n} \ell_t(\pi, \theta). \end{aligned}$$

Thus, $(\hat{\pi}_n^a, \hat{\theta}_{0,n}^a, \hat{\theta}_{1,n}^a, \hat{\theta}_{2,n}^a) \in (I_1(\eta) \cup I_2(\eta) \cup I_3(\eta)) \setminus (I_1(\eta) \cup I_3(\eta)) = I_2(\eta)$ with probability one. This completes the proof. ■

We make use of the following supplementary claims to prove the main argument.

LEMMA B1: *Given A1, A2, A3(iii), A4(i, iii), and A5(iv), for* $i_1, \cdots, i_4 \in \{\theta_{01}, \theta_{02}, \cdots, \theta_{0r_0}, \theta_1\}$,
(*i*) $n^{-1}\sum\{\nabla_{i_1} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^4]$ *a.s.*;
(*ii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\nabla_{i_2} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*iii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*iv*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\} \to E[\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)]$ *a.s.*

Proof of Lemma B1: (*i to iv*) Given the assumptions in A5(iv), these are straightforward by applying the ULLN and Theorem 2(ii). ■

LEMMA C1: *Given A1, A2, A3(iii), A4, A5(iv), A7(ii) and* H_0, *for each* π,
(*i*) $(\hat{\tilde{\theta}}_{0,n}^{(1)\prime}, \pi\hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi) = O_p(n^{-1/2})$;
(*ii*) $(\hat{\tilde{\theta}}_{0,n}^{(2)\prime}, \pi\hat{\tilde{\theta}}_{1,n}^{(2)})' = O_p(1)$;
(*iii*) $(\hat{\tilde{\theta}}_{0,n}^{(3)\prime}, \pi\hat{\tilde{\theta}}_{1,n}^{(3)})' = O_p(1)$.

Proof of Lemma C1: (*i*) We can differentiate (1) and (2) with respect to θ_2, and we obtain from these that $\tilde{M}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = 0$, implying that

$$\begin{bmatrix} \sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)}) & \sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)\prime}) \\ \sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)}) & \sum(-\pi^{-1}\hat{r}_t^{(0,2)} + \hat{r}_t^{(0,1)(0,1)}) \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(1)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(1)} \end{bmatrix} = \begin{bmatrix} (1-\pi)\sum(\hat{r}_t^{(1,1)} - \hat{r}_t^{(1,0)(0,1)}) \\ -(1-\pi)\sum\hat{r}_t^{(0,1)(0,1)} \end{bmatrix},$$

where $\hat{r}_t^{(i,j)(k,\ell)} := \hat{r}_t^{(i,j)}\hat{r}_t^{(k,\ell)\prime}$. Given this, we can apply Lemma B1(i to ii). Then

$$\begin{bmatrix} -R_*^{(2,0)} + R_*^{(1,0)(1,0)} & R_*^{(1,0)(0,1)\prime} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(1)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(1)} \end{bmatrix} = -(1-\pi) \begin{bmatrix} R_*^{(1,0)(0,1)} \\ R_*^{(0,1)(0,1)} \end{bmatrix} + o_p(1)$$

because $\sum\hat{r}_t^{(1,1)} = O_p(n^{1/2})$ and $\sum\hat{r}_t^{(0,2)} = O_p(n^{1/2})$ by A7(ii). Thus, it follows that $\hat{\tilde{\theta}}_{0,n}^{(1)} = o_p(1)$ and $(\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi) = o_p(1)$. The desired result follows from this and that the inverse of the coefficient matrix of $[\hat{\tilde{\theta}}_{0,n}^{(1)\prime}, \pi\hat{\tilde{\theta}}_{1,n}^{(1)}]'$ exists by A7(ii). Specifically, after some algebra,

$$\begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(1)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi \end{bmatrix} = \frac{1-\pi}{\pi} \begin{bmatrix} \hat{A}_n^{-1}\hat{B}_n\hat{E}_n^{-1}\sum\hat{r}_t^{(0,2)} \\ -\hat{E}_n^{-1}\sum\hat{r}_t^{(0,2)} \end{bmatrix},$$

where $\hat{A}_n := \sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)})$, $\hat{B}_n := \sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)})$, and $\hat{E}_n := -\pi^{-1}\sum\hat{r}_t^{(0,2)} + \sum\hat{r}_t^{(0,1)(0,1)} - \hat{B}_n\hat{A}_n^{-1}\hat{B}_n$. Note that $\hat{A}_n = O_p(n)$, $\hat{B}_n = O_p(n)$, $\hat{E}_n = O_p(n)$, and $\sum\hat{r}_t^{(2,0)} = O_p(n^{1/2})$, so that the desired result follows.

(*ii*) Likewise, if $(\tilde{\theta}_0^{(1)}(\theta_2)', \tilde{\theta}_1^{(1)}(\theta_2))$ is plugged back to $\tilde{M}_n^{(2)}(\pi,\theta_2)$ and $\tilde{K}_n^{(2)}(\pi,\theta_2)$, then another set of identities is obtained, and we differentiate them as before. Then

$$\begin{bmatrix} n^{-1}\sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)}) & n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)'}) \\ n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)}) & n^{-1}\sum(-\pi^{-1}\hat{r}_t^{(0,2)} + \hat{r}_t^{(0,1)(0,1)}) \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(2)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(2)} \end{bmatrix}$$
$$= \begin{bmatrix} (\pi(\hat{\tilde{\theta}}_{1,n}^{(1)})^2 + 1 - \pi)n^{-1}\sum(\hat{r}_t^{(1,2)} - \hat{r}_t^{(1,0)(0,2)}) \\ -n^{-1}\sum\{(3\pi(\hat{\tilde{\theta}}_{1,n}^{(1)})^2 + 2(1-\pi)\hat{\tilde{\theta}}_{1,n}^{(1)} + 1 - \pi)\hat{r}_t^{(0,1)(0,2)} + (\hat{\tilde{\theta}}_{1,n}^{(1)})^2\hat{r}_t^{(0,3)}\} \end{bmatrix}$$
$$+ O_p(\hat{\tilde{\theta}}_{0,n}^{(1)}) + O_p(\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + (1-\pi)),$$

where the last two remainders are $O_p(1)$ terms with coefficient $\hat{\tilde{\theta}}_{0,n}^{(1)}$ or $\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + (1-\pi)$. If we apply Lemmas B1($i$ to iv) and Lemma C1(i), then

$$\begin{bmatrix} E[-r_t^{(2,0)}(\theta_*) + r_t^{(1,0)(1,0)}(\theta_*)] & E[r_t^{(1,0)(0,1)'}(\theta_*)] \\ E[r_t^{(1,0)(0,1)}(\theta_*)] & E[r_t^{(0,1)(0,1)}(\theta_*)] \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(2)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(2)} \end{bmatrix}$$
$$= \left[\frac{1-\pi}{\pi}\right] \begin{bmatrix} E[r_t^{(1,2)}(\theta_*) - r_t^{(1,0)(0,2)}(\theta_*)] \\ -E[r_t^{(0,1)(0,2)}(\theta_*)] \end{bmatrix} + o_p(1).$$

Thus, the desired result trivially follows from the fact that the inverse of the coefficient matrix of $[\hat{\tilde{\theta}}_{0,n}^{(2)'}, \pi\hat{\tilde{\theta}}_{1,n}^{(2)}]'$ exists by A7(ii).

(*iii*) We can iterate the same procedure. That is, if $(\tilde{\theta}_0^{(2)}(\theta_2)', \tilde{\theta}_1^{(2)}(\theta_2))$ is plugged back to $\tilde{M}_n^{(3)}(\pi,\theta_2)$ and $\tilde{K}_n^{(3)}(\pi,\theta_2)$, then another set of identities is obtained, and we can differentiate them. Then,

$$\begin{bmatrix} n^{-1}\sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)}) & n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)'}) \\ n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)}) & n^{-1}\sum(-\pi^{-1}\hat{r}_t^{(0,2)} + \hat{r}_t^{(0,1)(0,1)}) \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(3)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(3)} \end{bmatrix}$$
$$= \begin{bmatrix} (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})n^{-1}\sum(-\hat{r}_t^{(1,3)} + \hat{r}_t^{(1,0)(0,3)}) \\ n^{-1}\sum\{-(\frac{1-\pi}{\pi})^3\hat{r}_t^{(0,4)} + 3(\frac{1-\pi}{\pi})^2\hat{r}_t^{(0,2)(0,2)} + (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})\hat{r}_t^{(0,1)(0,3)}\} \end{bmatrix}$$
$$+ O_p(\hat{\tilde{\theta}}_{0,n}^{(1)}) + O_p(\pi\hat{\tilde{\theta}}_{1,n}^{(1)} + (1-\pi)) + O_p(1).$$

Applying Lemma B1(i to iv) and A7(ii) completes the proof. ■

Proof of Lemma 2: (*i and ii*) Note that $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}$. Given this, we can apply the mean value theorem, so that for some $(\hat{\theta}_{n,0}^m, \hat{\theta}_{n,1}^m)$,

$$n^{-1/2}\sum\hat{r}_t^{(0,2)} = n^{-1/2}\sum r_t^{(0,2)}(\theta_0^*, \theta_*) + n^{-1}\sum\nabla_{\theta^1} r_t^{(0,2)}(\hat{\theta}_{n,0}^m, \hat{\theta}_{n,1}^m)\sqrt{n}[(\hat{\theta}_{n,0}^{n'}, \hat{\theta}_{n,1}^n) - (\theta_0^{*'}, \theta_*)]'.$$

By applying Lemma B1(ii), it follows that $n^{-1}\sum\nabla_{\theta^1} r_t^{(0,2)}(\hat{\theta}_{n,0}^m,\hat{\theta}_{n,1}^m)$ converges to $E[\nabla_{\theta^1} r_t^{(0,2)}(\theta_0^*,\theta_*)]$ a.s., and also $n^{-1/2}\sum r_t^{(0,2)}(\theta_0^*,\theta_*)$ obeys the CLT by the NAMXG condition, so that we can apply the asymptotic normality to $n^{-1/2}\sum\hat{r}_t^{(0,2)}$. That is, $n^{-1/2}\sum\hat{r}_t^{(0,2)} \Rightarrow G_1 \sim N(0,V_1)$. Further, $\hat{\tilde{\theta}}_{1,n}^{(1)} = -(1-\pi)/\pi + o_p(1)$ by Lemma C1(i). Combining these delivers the desired result.

(iii) Note that $\tilde{L}_n^{(3)}(\pi,\hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t + 2(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(1)}\}$. Given this,

$$\sum(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(1)} = (1-\hat{\tilde{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}\{(-\pi\hat{\tilde{\theta}}_{1,n}^{(1)}+\pi-1)\hat{r}_t^{(0,1)} - \hat{\tilde{\theta}}_{0,n}^{(1)\prime}\hat{r}_t^{(1,0)}\},$$

$$\sum(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t = \sum\{-\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,2)} + (1-(\hat{\tilde{\theta}}_{1,n}^{(1)})^2)\hat{r}_t^{(0,3)} + 2(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\hat{\tilde{\theta}}_{0,n}^{(1)\prime}\hat{r}_t^{(1,2)}\}$$

by (9) and (10). First, we can apply Lemma B1(ii) to $\sum\hat{r}_t^{(0,2)}\hat{r}_t^{(0,1)}$ and $\sum\hat{r}_t^{(0,2)}\hat{r}_t^{(1,0)}$, so that they are $O_p(n)$. Also, $(-\pi\hat{\tilde{\theta}}_{1,n}^{(1)}+\pi-1) = O_p(n^{1/2})$ and $\hat{\tilde{\theta}}_{0,n}^{(1)} = O_p(n^{1/2})$ by Lemma C1(i), implying that $\sum(\hat{h}_t^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(1)} = O_p(n^{1/2})$. Next, $\sum\hat{r}_t^{(0,2)}$ and $\sum\hat{r}_t^{(0,3)}$ are $O_p(n^{1/2})$ by A7(ii); $\hat{\tilde{\theta}}_{1,n}^{(1)}$ and $\hat{\tilde{\theta}}_{1,n}^{(2)}$ are $O_p(1)$ by Lemmas C1(i and ii); and $\hat{r}_t^{(1,2)} = O_p(n)$ by applying Lemma B1(iii). This implies that $\sum(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t = O_p(n^{1/2})$. Hence, for each π, $\tilde{L}_n^{(3)}(\pi,\hat{\theta}_{1,n}^n) = O_p(n^{1/2})$.

(iv) By some algebra, it's not hard to derive that

$$\begin{aligned}\tilde{L}_n^{(4)}(\pi,\hat{\theta}_{1,n}^n) &= \left[\frac{1-\pi}{\pi}\right]\sum\left\{(1+\hat{\tilde{\theta}}_{1,n}^{(1)}+(\hat{\tilde{\theta}}_{1,n}^{(1)})^2)\hat{r}_t^{(0,4)} - 3(\pi(\hat{\tilde{\theta}}_{1,n}^{(1)})^2+1-\pi)\hat{r}_t^{(0,2)(0,2)}\right\}\\ &+\left[\frac{1-\pi}{\pi}\right]\sum\left\{3\hat{\tilde{\theta}}_{0,n}^{(2)}(\hat{r}_t^{(1,2)}-\hat{r}_t^{(1,0)(0,2)})+3\pi\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,1)(0,2)}\right\}+o_p(n).\end{aligned}$$

Also, Lemmas C1(i and ii) show that for each π, $\hat{\tilde{\theta}}_{1,n}^{(1)} = -(1-\pi)/\pi + o_p(1)$ and

$$\begin{aligned}\begin{bmatrix}\hat{\tilde{\theta}}_{0,n}^{(2)}\\ \pi\hat{\tilde{\theta}}_{1,n}^{(2)}\end{bmatrix} = \left[\frac{1-\pi}{\pi}\right]&\begin{bmatrix}-\sum\hat{r}_t^{(2,0)}+\sum\hat{r}_t^{(1,0)(1,0)} & -\sum\hat{r}_t^{(1,1)}+\sum\hat{r}_t^{(1,0)(0,1)\prime}\\ -\sum\hat{r}_t^{(1,1)}+\sum\hat{r}_t^{(1,0)(0,1)} & -\pi^{-1}\sum\hat{r}_t^{(0,2)}+\sum\hat{r}_t^{(0,1)(0,1)}\end{bmatrix}^{-1}\times\\ &\begin{bmatrix}\sum\hat{r}_t^{(1,2)}-\sum\hat{r}_t^{(1,0)(0,2)}\\ -\sum\hat{r}_t^{(0,1)(0,2)}\end{bmatrix}+o_p(1).\end{aligned}$$

We can combine this with Lemma B1(i to iii), obtaining that for each π, $n^{-1}\tilde{L}_n^{(4)}(\pi,\hat{\theta}_{1,n}^n) = [\frac{1-\pi}{\pi}]\Xi^{(4)}(\pi) + o_p(1)$. This completes the proof. ■

Proof of Theorem 3: (i) By (12), for each π,

$$2(\tilde{L}_n(\pi,\theta_2)-\tilde{L}_n(\pi,\hat{\theta}_{1,n}^n)) \Rightarrow \left[\frac{1-\pi}{\pi}\right]G_1\xi^2 + \frac{1}{12}\Xi^{(4)}(\pi)\xi^4.$$

The maximum of the RHS depends on the sign of G_1. If $G_1 < 0$, then the maximum is attained when $\xi = 0$, so that $\max_{\theta_2} 2(\tilde{L}_n(\pi,\theta_2)-\tilde{L}_n(\pi,\hat{\theta}_{1,n}^n)) \to 0$ in probability. Otherwise, the FOC implies that the maximum of the RHS is attained when $\xi^2 = 6(\frac{1-\pi}{\pi})G_1[\Xi^{(4)}(\pi)]^{-1}$. Thus, by combining these, $\max_{\theta_2} 2(\tilde{L}_n(\pi,\theta_2) - \tilde{L}_n(\pi,\hat{\theta}_{1,n}^n)) \Rightarrow -3[\frac{1-\pi}{\pi}]^2[\Xi^{(4)}(\pi)]^{-1}\max[0,G_1]^2$.

(*ii*) By applying the continuous mapping theorem,

$$\max_{\pi\in[\epsilon,1-\epsilon]}\max_{\theta_2} 2(\tilde{L}_n(\pi,\theta_2) - \tilde{L}_n(\pi,\hat{\theta}_{1,n}^n)) \Rightarrow \max_{\pi\in[\epsilon,1-\epsilon]} -3[\frac{1-\pi}{\pi}]^2[\Xi^{(4)}(\pi)]^{-1}\max[0,G_1]^2.$$

If $G_1 < 0$, then for any $\pi \in [\epsilon, 1-\epsilon]$, the RHS is achieved as zero. If $G_1 > 0$, then the RHS is maximized by maximizing $-3[\frac{1-\pi}{\pi}]^2[\Xi^{(4)}(\pi)]^{-1}$. By the definition of $\Xi^{(4)}$, $-3[\frac{1-\pi}{\pi}]^2[\Xi^{(4)}(\pi)]^{-1}$ is maximized when $\pi = 1/2$, as desired. ■

LEMMA B2: *Given A1, A2, A3(iv), A4(i, iii), and A5(v), , for* $i_1,\cdots,i_6 \in \{\theta_{01},\theta_{02},\cdots,\theta_{0r_0},\theta_1\}$,

(*i*) $n^{-1}\sum\{\nabla_{i_1} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)\}^4]$ *a.s.*;

(*ii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1}\nabla_{i_2} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)\}^4]$ *a.s.*;

(*iii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)\}^4]$ *a.s.*;

(*iv*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_4} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\cdots\nabla_{i_4} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)\}^2]$ *a.s.*;

(*v*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_5} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\cdots\nabla_{i_5} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)\}^2]$ *a.s.*;

(*vi*) $n^{-1}\sum\nabla_{i_1}\cdots\nabla_{i_6} f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n) \to E[\nabla_{i_1}\cdots\nabla_{i_6} f_t(\theta_0^*,\theta_*)/f_t(\theta_0^*,\theta_*)]$ *a.s.*

Proof of Lemma B2: (*i to vi*) The proofs are identical to that of Lemma B1. ■

LEMMA C2: *Given A1, A2, A3(iv), A4, A5(v), A7(iii) and* H_0, *for each* π,

(*i*) $(\hat{\hat{\theta}}_{0,n}^{(1)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(1)} + 1 - \pi) = 0$;

(*ii*) $(\hat{\hat{\theta}}_{0,n}^{(2)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(2)})' = -(\frac{1-\pi}{\pi})(\alpha',\ \beta)' + O_p(n^{-1/2})$;

(*iii*) $(\hat{\hat{\theta}}_{0,n}^{(4)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(4)})' = O_p(1)$;

(*iv*) $\hat{g}_t^{(1)} = 0$ *and* $\hat{m}_t^{(1)} = 0$;

(*v*) *for a sequence of random variables*, $\{\hat{q}_t\}$ *say, if* $\sum\hat{q}_t\hat{r}_t^{(1)} = O_p(n)$, *then* $\sum\hat{q}_t\hat{f}_t\hat{g}_t^{(2)} = O_p(n^{1/2})$;

(*vi*) $\sum\hat{m}_t\hat{g}_t^{(3)} = O_p(n^{1/2})$ *and* $\sum\hat{k}_t\hat{g}_t^{(3)} = O_p(n^{1/2})$;

(*vii*) $\sum\hat{m}_t\hat{g}_t^{(4)} = O_p(n^{1/2})$ *and* $\sum\hat{k}_t\hat{g}_t^{(4)} = O_p(n^{1/2})$.

(*viii*) $n^{-1/2}\sum\hat{r}_t^{(0,3)} \overset{\text{A}}{\sim} N(0,\Omega^{(3)})$.

Proof of Lemma C2: (*i*) Let $\hat{r}_t^{(0,2)} = \alpha'\hat{r}_t^{(1,0)} + \beta\hat{r}_t^{(0,1)}$ and iterate the proof of Lemma C1(i).

(ii) Using Lemma C2(i), if we iterate the proof of Lemma C1(ii),

$$\begin{bmatrix} n^{-1}\sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)}) & n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)'}) \\ n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)}) & n^{-1}\sum \hat{r}_*^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\hat{\theta}}_{0,n}^{(2)} \\ \pi\hat{\hat{\theta}}_{1,n}^{(2)} \end{bmatrix}$$
$$= \begin{bmatrix} (\frac{1-\pi}{\pi})n^{-1}\sum(\hat{r}_t^{(1,2)} - \hat{r}_t^{(1,0)(0,2)}) \\ -n^{-1}\sum\{(\frac{1-\pi}{\pi})\hat{r}_t^{(0,1)(0,2)} + (\frac{1-\pi}{\pi})^2\hat{r}_t^{(0,3)}\} \end{bmatrix}.$$

If we apply Lemmas B2(i to ii), then it's trivial to show that

$$\begin{bmatrix} R_*^{(1,0)(1,0)} & R_*^{(1,0)(0,1)'} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\hat{\theta}}_{0,n}^{(2)} \\ \pi\hat{\hat{\theta}}_{1,n}^{(2)} \end{bmatrix}$$
$$= -\left(\frac{1-\pi}{\pi}\right)\left\{\begin{bmatrix} R_*^{(1,0)(0,2)} \\ R_*^{(0,1)(0,2)} \end{bmatrix} - \begin{bmatrix} n^{-1}\sum\hat{r}_t^{(1,2)} \\ \left(\frac{1-\pi}{\pi}\right)n^{-1}\sum\hat{r}_t^{(0,3)} \end{bmatrix}\right\} + o_p(1).$$

Further, if we use the fact that $r_t^{(0,2)}(\theta_0^*, \theta_*) = \alpha' r_t^{(1,0)}(\theta_0^*, \theta_*) + \beta r_t^{(0,1)}(\theta_0^*, \theta_*)$, then it's not hard to derive that

$$\begin{bmatrix} R_*^{(1,0)(1,0)} & R_*^{(1,0)(0,1)'} \\ R_*^{(1,0)(0,1)} & R_*^{(0,1)(0,1)} \end{bmatrix} \left[\begin{bmatrix} \hat{\hat{\theta}}_{0,n}^{(2)} \\ \pi\hat{\hat{\theta}}_{1,n}^{(2)} \end{bmatrix} + \left[\frac{1-\pi}{\pi}\right]\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right] = -\frac{1-\pi}{\pi}\begin{bmatrix} n^{-1}\sum\hat{r}_t^{(1,2)} \\ \left(\frac{1-\pi}{\pi}\right)n^{-1}\sum\hat{r}_t^{(0,3)} \end{bmatrix} + o_p(1).$$

By Lemmas B2(i to iii) and A7(iii), the RHS is $O_p(n^{-1/2})$, leading to the desired result.

(iii) Using Lemma C2(i), if we rearrange $\tilde{M}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = 0$,

$$\begin{bmatrix} n^{-1}\sum(-\hat{r}_t^{(2,0)} + \hat{r}_t^{(1,0)(1,0)}) & n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)'}) \\ n^{-1}\sum(-\hat{r}_t^{(1,1)} + \hat{r}_t^{(1,0)(0,1)}) & n^{-1}\sum \hat{r}_*^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\hat{\theta}}_{0,n}^{(4)} \\ \pi\hat{\hat{\theta}}_{1,n}^{(4)} \end{bmatrix}$$
$$= \begin{bmatrix} \frac{(1-\pi)(1-3\pi+3\pi^2)}{\pi^3}(\bar{R}_n^{(1,4)} - \bar{R}_n^{(1,0)(0,4)}) \\ (\frac{1-\pi}{\pi})^4\bar{R}_n^{(0,5)} + \frac{2(1-\pi)^2(7\pi-5)}{\pi^3}\bar{R}_n^{(0,3)(0,2)} - \frac{(1-\pi)(1-3\pi+3\pi^2)}{\pi^3}\bar{R}_n^{(0,1)(0,4)} \end{bmatrix}$$
$$+6\begin{bmatrix} -\left(\frac{1-\pi}{\pi}\right)^2\bar{R}_n^{(0,2)(1,2)} \\ \left(\frac{1-\pi}{\pi}\right)^2\bar{R}_n^{(0,1)(0,2)(0,2)} \end{bmatrix} + O_p(1),$$

where $\bar{R}_n^{(k,l)(i,j)(m,q)} := n^{-1}\sum\hat{r}_t^{(k,l)}\hat{r}_t^{(i,j)}\hat{r}_t^{(m,q)}$ and the remainder is the collection of $O_p(1)$ terms given in Lemma B2(i to iv). We presented only the terms whose moment or derivative orders are highest. Given this, it's straightforward to obtain the desired result by Lemma B2(i to v).

(iv) This is obvious from the facts that $\hat{g}_t^{(1)} = (\pi-1)(\hat{h}_t - \hat{k}_t)(\hat{g}_t)^2$, $\hat{m}_t^{(1)} = (\pi\hat{\hat{\theta}}_{1,n}^{(1)} + 1 - \pi)\hat{f}_t^{(0,2)}$, $\hat{h}_t = \hat{k}_t$ and Lemma C2(i).

(v) From $\hat{r}_t^{(0,2)} = \alpha'\hat{r}_t^{(1,0)} + \beta\hat{r}_t^{(0,1)}$ and Lemma C2(i), $\hat{f}_t\hat{g}_t^{(2)} = -(\frac{1-\pi}{\pi}\alpha + \hat{\hat{\theta}}_{0,n}^{(2)})'\hat{r}_t^{(1,0)} - (\frac{1-\pi}{\pi}\beta + \pi\hat{\hat{\theta}}_{1,n}^{(2)})\hat{r}_t^{(0,1)}$. The given result follows by Lemma C2(ii).

(vi) Lemma C2(iv), $\tilde{M}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = 0$ imply that $\sum \hat{m}_t \hat{g}_t^{(3)} = -\sum \hat{m}_t^{(3)} \hat{g}_t$ and $\sum \hat{k}_t \hat{g}_t^{(3)} = -\sum \hat{k}_t^{(3)} \hat{g}_t - 3\sum \hat{k}_t^{(1)} \hat{g}_t^{(2)}$. Given these, $\sum \hat{m}_t^{(3)} \hat{g}_t$, $\sum \hat{k}_t^{(3)} \hat{g}_t$, and $\sum \hat{k}_t^{(1)} \hat{g}_t^{(2)}$ are $O_p(n^{1/2})$ by Lemmas C1(iii) and C2(i, ii, and v), implying the given conclusion.

(vii) Lemma C2(iv), $\tilde{M}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = 0$, and $\tilde{K}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = 0$ imply that $\sum \hat{m}_t \hat{g}_t^{(4)} = -\sum\{\hat{m}_t^{(4)} \hat{g}_t + 6\hat{m}_t^{(2)} \hat{g}_t^{(2)}\}$ and $\sum \hat{k}_t \hat{g}_t^{(4)} = -\sum\{\hat{k}_t^{(4)} \hat{g}_t + 6\hat{k}_t^{(2)} \hat{g}_t^{(2)} + 4\hat{k}_t^{(1)} \hat{g}_t^{(3)}\}$. Further, $\sum \hat{m}_t^{(4)} \hat{g}_t$, $\sum \hat{m}_t^{(2)} \hat{g}_t^{(2)}$, $\sum \hat{k}_t^{(4)} \hat{g}_t$ and $\sum \hat{m}_t^{(2)} \hat{g}_t^{(2)}$ are $O_p(n^{1/2})$ by Lemmas C1(iii) and C2(i to iii, and v). Finally, note that $\sum \hat{k}_t^{(1)} \hat{g}_t^{(3)} = \hat{\tilde{\theta}}_{1,n}^{(1)} \sum(\alpha' \hat{f}_t^{(1,0)} + \beta \hat{f}_t^{(0,1)}) \hat{g}_t^{(3)} = \hat{\tilde{\theta}}_{1,n}^{(1)} \sum(\alpha' \hat{m}_t \hat{g}_t^{(3)} + \beta \hat{k}_t \hat{g}_t^{(3)})$. Thus, $\sum \hat{k}_t^{(1)} \hat{g}_t^{(3)} = O_p(n^{1/2})$ by Lemma C2(vi).

($viii$) By the mean-value theorem, it is trivial to derive that

$$n^{-1/2} \sum \hat{r}_t^{(0,3)} = n^{-1/2} \sum r_t^{(0,3)}(\theta_0^*, \theta_*) + n^{-1/2} \sum \nabla_{\theta^1} r_t^{(0,3)}(\bar{\theta}_0, \bar{\theta}_1)[(\hat{\theta}_{n,0}^{n\prime}, \hat{\theta}_{n,1}^n) - (\theta_0^{*\prime}, \theta_*)]'$$

for some $(\bar{\theta}_0, \bar{\theta}_1)$. Given this, $n^{-1} \sum \nabla_{\theta^1} r_t^{(0,3)}(\,\cdot\,,\cdot\,)$ obeys the SULLN by A5(iv). Theorems 2(ii, iii) and A7(iii) complete the proof. ■

Proof of Lemma 3: (i) If we can combine Lemma C2(i) with the proof of Lemma 2(iii), then

$$\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = -\left[\frac{(1-2\pi 1)(1-\pi)}{\pi^2}\right] \sum \hat{r}_t^{(0,3)}.$$

The desired result follows from Lemma C2($viii$).

(ii) Note that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi) \sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) \hat{g}_t + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(2)}\}$ by applying Lemma C2(iv). We examine each element in the RHS. First,

$$\begin{aligned}\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) \hat{g}_t = &-\hat{\tilde{\theta}}_{1,n}^{(3)} \sum \hat{r}_t^{(0,2)} - 3\hat{\tilde{\theta}}_{1,n}^{(2)} \hat{\tilde{\theta}}_{1,n}^{(1)} \sum \hat{r}_t^{(0,3)} + (1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^3) \sum \hat{r}_t^{(0,4)} \\ &+ 3(\hat{\tilde{\theta}}_{0,n}^{(1)\prime} \hat{\tilde{\theta}}_{1,n}^{(2)} - \hat{\tilde{\theta}}_{0,n}^{(2)\prime} (\hat{\tilde{\theta}}_{1,n}^{(1)} - 1)) \sum \hat{r}_t^{(1,2)} + 3(1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^2) \hat{\tilde{\theta}}_{0,n}^{(1)\prime} \sum \hat{r}_t^{(1,3)} \\ &+ 3(1 - \hat{\tilde{\theta}}_{1,n}^{(1)}) \hat{\tilde{\theta}}_{0,n}^{(1)\prime} \sum \hat{r}_t^{(2,2)} \hat{\tilde{\theta}}_{0,n}^{(1)}.\end{aligned}$$

Note that each summand is $O_p(n^{1/2})$ by A7(iii), and we already showed that the coefficients of the summands in the RHS are $O_p(1)$ in Lemma C1(iii) and Lemmas C2(i to ii). Thus, $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) \hat{g}_t = O_p(n^{1/2})$. Next, trivially $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(2)} = O_p(n^{1/2})$ by Lemma C2(v). Thus, $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$.

(iii) Note that $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi) \sum\{(\hat{h}_t^{(4)} - \hat{k}_t^{(4)}) \hat{g}_t + 6(\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t^{(2)} + 4(\hat{h}_t^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(3)}\}$ by applying Lemma C2(iv). We examine each element in the RHS. First,

$$\begin{aligned}\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t = &- (3(\hat{\tilde{\theta}}_{1,n}^{(2)})^2 + 4\hat{\tilde{\theta}}_{1,n}^{(3)}(\pi^{-1}-1))\sum \hat{r}_t^{(0,3)} - 6\pi^{-2}(1-\pi)^2\hat{\tilde{\theta}}_{1,n}^{(2)}\sum \hat{r}_t^{(0,4)} \\ &- \pi^{-4}(1-2\pi)(1-2\pi+2\pi^2)\sum \hat{r}_t^{(0,5)} - 2\pi^{-1}(3\pi\hat{\tilde{\theta}}_{0,n}^{(2)}\hat{\tilde{\theta}}_{1,n}^{(2)} - 2\hat{\tilde{\theta}}_{0,n}^{(3)})'\sum \hat{r}_t^{(1,2)} \\ &+ 12\hat{\tilde{\theta}}_{0,n}^{(2)}\pi^{-1}(1-0.5\pi^{-1})\sum \hat{r}_t^{(1,3)}.\end{aligned}$$

By A7(iii), each summand is $O_p(n^{1/2})$. Further, Lemma C1(iii) and Lemmas C2(i to ii) imply that the coefficients of the summands in the RHS are $O_p(1)$. Thus, $\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t = O_p(n^{1/2})$. Second, it's trivial that $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(2)} + \sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(3)} = O_p(n^{1/2})$ by Lemma C2(v and vi). Therefore, $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$, as desired.
(iv) Note that $\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t + 10(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} + 10(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} + 5(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)}\}$. We examine each element in the RHS. First, $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} = O_p(n^{1/2})$ by Lemma C2(v). Second,

$$\begin{aligned}\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} =&(-1+2\pi)/(\pi^2)\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} - \hat{\tilde{\theta}}_{1,n}^{(2)}\sum \hat{f}_t^{(0,2)}\hat{g}_t^{(3)} \\ =&(-1+2\pi)/(\pi^2)\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} + o_p(n),\end{aligned}$$

where the last equality follows from that $\sum \hat{f}_t^{(0,2)}\hat{g}_t^{(3)} = \sum(\alpha'\hat{m}_t + \beta\hat{k}_t)\hat{g}_t^{(3)}$, which is $O_p(n^{1/2})$ by Lemma C2(vi). Note that

$$\begin{aligned}n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} =&(3(1-\pi)\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} - \hat{\tilde{\theta}}_{0,n}^{(3)})'\bar{R}_n^{(1,0)(0,3)} + (3(1-\pi)\beta\hat{\tilde{\theta}}_{1,n}^{(2)} - \pi\hat{\tilde{\theta}}_{1,n}^{(3)})\bar{R}_n^{(0,1)(0,3)} \\ &+ (1-\pi)(1-2\pi)/(\pi^2)\bar{R}_n^{(0,3)(0,3)}\end{aligned} \qquad (19)$$

from the fact that

$$\hat{f}_t\hat{g}_t^{(3)} = (3(1-\pi)\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} - \hat{\tilde{\theta}}_{0,n}^{(3)})'\hat{r}_t^{(1,0)} + (3(1-\pi)\beta\hat{\tilde{\theta}}_{1,n}^{(2)} - \pi\hat{\tilde{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,1)} + \left[\frac{1-\pi}{\pi}\right]\left[\frac{1-2\pi}{\pi}\right]\hat{r}_t^{(0,3)}.$$

Further, $\sum \hat{k}_t\hat{g}_t^{(3)} = o_p(n)$ and $\sum \hat{m}_t\hat{g}_t^{(3)} = o_p(n)$ by Lemma C2(vi), implying that

$$\begin{bmatrix}\bar{R}_n^{(1,0)(1,0)} & \bar{R}_n^{(0,1)(1,0)'} \\ \bar{R}_n^{(1,0)(0,1)} & \bar{R}_n^{(0,1)(0,1)}\end{bmatrix}\begin{bmatrix}(3(1-\pi)\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} - \hat{\tilde{\theta}}_{0,n}^{(3)}) \\ (3(1-\pi)\beta\hat{\tilde{\theta}}_{1,n}^{(2)} - \pi\hat{\tilde{\theta}}_{1,n}^{(3)})\end{bmatrix} + \left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right]\begin{bmatrix}\bar{R}_n^{(0,3)(1,0)} \\ \bar{R}_n^{(0,3)(0,1)}\end{bmatrix} = o_p(1).$$

From this, we can solve for $(3(1-\pi)\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} - \hat{\tilde{\theta}}_{0,n}^{(3)})$ and $(3(1-\pi)\beta\hat{\tilde{\theta}}_{1,n}^{(2)} - \pi\hat{\tilde{\theta}}_{1,n}^{(3)})$ and plug them back to (19). Then,

$$\begin{aligned}n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} &= \left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right]\left[\bar{R}_n^{(0,3)(0,3)} - \bar{R}_n^{(0,3)(1)}\bar{R}_n^{(1)(1)^{-1}}\bar{R}_n^{(1)(0,3)}\right] + o_p(1) \\ &= \left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right]\Omega^{(3)} + o_p(1)\end{aligned}$$

by applying Lemma B2(i to iii). Thus, it follows that $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} = -(1-\pi)(1-2\pi)^2/(\pi^4)\Omega^{(3)} + o_p(1)$. Third, note that

$$\sum(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t =(1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^5)\sum \hat{r}_t^{(0,6)} + 10(1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^3)\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\sum \hat{r}_t^{(1,4)} + 15(1 - \hat{\tilde{\theta}}_{1,n}^{(1)})\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(2)} + O_p(n^{1/2}), \tag{20}$$

where the remainder is the collection of $O_p(n^{1/2})$ terms implied by A7(iii). We can apply Lemmas B2(iv to vi) and C2(i) to obtain that

$$n^{-1}\sum(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t = \left[1 + \left(\frac{1-\pi}{\pi}\right)^5\right] R_*^{(0,6)} - 10\left[1 + \left(\frac{1-\pi}{\pi}\right)^3\right]\left(\frac{1-\pi}{\pi}\right)\alpha' R_*^{(1,4)} + 15\left[1 + \frac{1-\pi}{\pi}\right]\left(\frac{1-\pi}{\pi}\right)^2 \alpha' R_*^{(2,2)}\alpha + o_p(1).$$

Finally,

$$\begin{aligned}\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)} =& -(1 - \hat{\tilde{\theta}}_{1,n}^{(1)})\left[\frac{1-4\pi+6\pi^2-3\pi^3}{\pi^3}\right]\alpha'\sum \hat{r}_t^{(1,4)} - (1-\hat{\tilde{\theta}}_{1,n}^{(1)})\pi\alpha'\sum \hat{r}_t^{(2,1)}\hat{\tilde{\theta}}_{0,n}^{(2)} \\ & - 3\beta(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\sum \hat{r}_t^{(2,1)}\hat{\tilde{\theta}}_{0,n}^{(2)} + 6\hat{\tilde{\theta}}_{1,n}^{(1)}(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\alpha'\sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(2)} \\ & - 3(1-\hat{\tilde{\theta}}_{1,n}^{(1)})\left[-\frac{1-\pi}{\pi}\right]^2\sum \alpha'(\nabla_{\theta_0}\alpha'\nabla_{\theta_0}^2\hat{f}_t\alpha)/\hat{f}_t + o_p(n),\end{aligned}$$

where the remainder is the $o_p(n)$ terms implied by A7(iii). We can apply Lemmas B2(iii to v) and C2(i) to obtain that

$$\begin{aligned}n^{-1}\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)} =& -\left[\frac{1-4\pi+6\pi^2-3\pi^3}{\pi^4}\right]\alpha' R_*^{(1,4)} + \left[\frac{1-\pi}{\pi}\right]\alpha' R_*^{(2,1)}\alpha \\ & - \left[\frac{3\beta}{\pi}\right]\left[\frac{1-\pi}{\pi}\right]^2\alpha' R_*^{(2,1)}\alpha + \frac{6}{\pi}\left[\frac{1-\pi}{\pi}\right]^2\alpha' R_*^{(2,2)}\alpha \\ & - \frac{3}{\pi}\left[-\frac{1-\pi}{\pi}\right]^2 M_*^{(1,0)} + o_p(n),\end{aligned}$$

Summing up all the elements in the RHS according to $\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t + 10(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} + 10(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} + 5(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)}\}$ leads to the desired result. ■

Proof of Theorem 4: (i) By (17) and the continuous mapping theorem,

$$\max_{\theta_2} 2(\tilde{L}_n(\pi,\theta_2) - \tilde{L}_n(\pi,\hat{\theta}_{1,n}^n)) \Rightarrow \max_{\xi} -\frac{2}{3!}\left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right] G_2\xi^3 + \frac{1}{6!}\Xi^{(6)}(\pi)\xi^6,$$

and the maximum of the RHS with respect to ξ is attained when $\xi^3 = 60[\frac{(1-\pi)(1-2\pi)}{\pi^2}][\Xi^{(6)}(\pi)]^{-1}G_2$ as $-10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[\Xi^{(6)}(\pi)]^{-1}G_2^2$.

(ii) Further maximizing the RHS with respect to π, $LR_n^{(2)}(\epsilon) \Rightarrow -10[\frac{(1-\pi^\ddagger)(1-2\pi^\ddagger)}{(\pi^\ddagger)^2}]^2[\Xi^{(6)}(\pi^\ddagger)]^{-1}G_2^2$ by the definition of $\pi^\ddagger$ and the continuous mapping theorem. This completes the proof. ■

REMARKS: 1. Given that $\Xi^{(6)}(\pi) < 0$ for each π, $\lim_{\pi\to 1/2} -10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[\Xi^{(6)}(\pi)]^{-1} = 0$.
2. If $R_*^{(0,6)} = 0$, $\lim_{\pi\to 0} -10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[\Xi^{(6)}(\pi)]^{-1} = 0$.
3. If $R_*^{(0,6)} < 0$, $-10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2[\Xi^{(6)}(\pi)]^{-1}$ converges to a positive real number as π tends to zero or one.

LEMMA B3: *Given A1, A2, A3(v), A4(i, iii), and A5(vi), , for* $i_1, \cdots, i_8 \in \{\theta_{01}, \theta_{02}, \cdots, \theta_{0r_0}, \theta_1\}$,
(*i*) $n^{-1}\sum\{\nabla_{i_1} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^4]$ *a.s.*;
(*ii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1}\nabla_{i_2} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^4]$ *a.s.*;
(*iii*) $n^{-1}\sum\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^4]$ *a.s.*;
(*iv*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_4} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^4 \to E[\{\nabla_{i_1}\cdots\nabla_{i_4} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*v*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_5} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\cdots\nabla_{i_5} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*vi*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_6} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\cdots\nabla_{i_6} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*vii*) $n^{-1}\sum\{\nabla_{i_1}\cdots\nabla_{i_7} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)\}^2 \to E[\{\nabla_{i_1}\cdots\nabla_{i_7} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)\}^2]$ *a.s.*;
(*viii*) $n^{-1}\sum\nabla_{i_1}\cdots\nabla_{i_8} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) \to E[\nabla_{i_1}\cdots\nabla_{i_8} f_t(\theta_0^*, \theta_*)/f_t(\theta_0^*, \theta_*)]$ *a.s.*

Proof of Lemma B3: (*i to viii*) The proofs are identical to that of Lemma B1. ■

LEMMA C3: *Given A1, A2, A3(v), A4, A5(vi), A7(iv), and* H_{02}, *if* $\pi = 1/2$, *then*
(*i*) $(\hat{\hat{\theta}}_{0,n}^{(3)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(3)})' = 1.5(\alpha', \beta)' + O_p(n^{-1/2})$;
(*ii*) $(\hat{\hat{\theta}}_{0,n}^{(5)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(5)})' = O_p(1)$;
(*iii*) $(\hat{\hat{\theta}}_{0,n}^{(6)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(6)})' = O_p(1)$;
(*iv*) *for a sequence of random variables,* $\{\hat{q}_t\}$ *say, if* $\sum\hat{q}_t\hat{r}_t^{(1)} = O_p(n)$, *then* $\sum\hat{q}_t\hat{f}_t\hat{g}_t^{(3)} = O_p(n^{1/2})$;
(*v*) $\sum\hat{m}_t\hat{g}_t^{(5)} = O_p(n^{1/2})$ *and* $\sum\hat{k}_t\hat{g}_t^{(5)} = O_p(n^{1/2})$.

Proof of Lemma C3: (*i*) By Lemma C2(i), it follows that

$$\hat{f}_t\hat{g}_t^{(3)} = (3(1-\pi)\alpha - \hat{\hat{\theta}}_{0,n}^{(3)})'\hat{r}_t^{(1,0)} + (3(1-\pi)\beta - \pi\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,1)} + \left[\frac{1-\pi}{\pi}\right]\left[\frac{1-2\pi}{\pi}\right]\hat{r}_t^{(0,3)}, \qquad (21)$$

and the last term vanishes, if $\pi = 1/2$. Further, applying Lemma C2(vi) leads to that $\sum\hat{m}_t\hat{g}_t^{(3)} = O_p(n^{1/2})$ and $\sum\hat{k}_t\hat{g}_t^{(3)} = O_p(n^{1/2})$, so that it follows that $(1.5\beta - 0.5\hat{\hat{\theta}}_{1,n}^{(3)})\sum(\hat{r}_t^{(0,1)})^2 = O_p(n^{1/2})$ and $(1.5\alpha - \hat{\hat{\theta}}_{0,n}^{(3)})'\sum\hat{r}_t^{(1,0)}\hat{r}_t^{(1,0)\prime} = O_p(n^{1/2})$. Given these, $\sum(\hat{r}_t^{(0,1)})^2 = O_p(n)$ and $\sum\hat{r}_t^{(1,0)}\hat{r}_t^{(1,0)\prime} = O_p(n)$ by Lemma B3(ii), implying that $(\hat{\hat{\theta}}_{0,n}^{(3)\prime},\ \pi\hat{\hat{\theta}}_{1,n}^{(3)})' = 1.5(\alpha',\ \beta)' + O_p(n^{-1/2})$.

(*ii*) Rearranging $\tilde{M}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = 0$ leads to that

$$\begin{bmatrix} \bar{R}_n^{(1,0)(1,0)} & \bar{R}_n^{(0,1)(1,0)'} \\ \bar{R}_n^{(1,0)(0,1)} & \bar{R}_n^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(5)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(5)} \end{bmatrix} = - \begin{bmatrix} 0 \\ \bar{R}_n^{(0,6)} - 15\bar{R}_n^{(0,2)(0,4)} + 30\bar{R}_n^{(0,2)(0,2)(0,2)} \end{bmatrix} + O_p(1),$$

where $O_p(1)$ remainders are the collections of the $O_p(1)$ terms in Lemmas B2(i to iv) multiplied by other $O_p(1)$ terms in Lemmas C2(i, ii, iv) and C3(i). Lemmas B3(ii, iv, vi) complete the proof.

(*iii*) Rearranging $\tilde{M}_n^{(7)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(7)}(\pi, \hat{\theta}_{1,n}^n) = 0$ leads to that

$$\begin{aligned} &\begin{bmatrix} \bar{R}_n^{(1,0)(1,0)} & \bar{R}_n^{(0,1)(1,0)'} \\ \bar{R}_n^{(1,0)(0,1)} & \bar{R}_n^{(0,1)(0,1)} \end{bmatrix} \begin{bmatrix} \hat{\tilde{\theta}}_{0,n}^{(6)} \\ \pi\hat{\tilde{\theta}}_{1,n}^{(6)} \end{bmatrix} \\ &= \begin{bmatrix} 30\bar{R}_n^{(1,0)(0,2)(0,4)} + 90\bar{R}_n^{(1,2)(0,2)(0,2)} - 90\bar{R}_n^{(1,0)(0,2)(0,2)(0,2)} \\ 30\bar{R}_n^{(0,1)(0,2)(0,4)} + 90\bar{R}_n^{(0,2)(0,2)(0,3)} - 90\bar{R}_n^{(0,1)(0,2)(0,2)(0,2)} \end{bmatrix} \\ &+ \begin{bmatrix} \bar{R}_n^{(1,6)} - \bar{R}_n^{(1,0)(0,6)} - 15\bar{R}_n^{(1,4)(0,2)} - 15\bar{R}_n^{(1,2)(0,4)} \\ \bar{R}_n^{(1,4)(0,2)} - 15\bar{R}_n^{(1,2)(0,4)} \end{bmatrix} + O_p(1), \end{aligned}$$

where $\bar{R}_n^{(k,l)(i,j)(m,q)(x,y)} := n^{-1}\sum \hat{r}_t^{(k,l)}\hat{r}_t^{(i,j)}\hat{r}_t^{(m,q)}\hat{r}_t^{(x,y)}$ and the remainders are the collections of $O_p(1)$ terms verified by Lemmas B2(i to v) and multiplied by other $O_p(1)$ terms in Lemmas C2(i, ii, iv) and C3(i). Lemma B3(ii to vii) complete the proof.

(*iv*) This is obvious by (21) and Lemma C3(i).

(*v*) Given Lemma C3(iv), this is identical to the proof of Lemma C2(vi). ■

Proof of Lemma 4: (*i*) Note that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)}\}$ by applying Lemma C2(i). Given this, if we do some algebra, then it follows that

$$\begin{aligned} \tilde{L}_n^{(4)}(1/2, \hat{\theta}_{1,n}^n) &= \sum\left\{3\left(\hat{\tilde{\theta}}_{0,n}^{(2)} - \alpha\right)'\hat{r}_t^{(1,2)} - \left(3\hat{\tilde{\theta}}_{0,n}^{(2)}\right)'\hat{r}_t^{(2,0)}\alpha\right\} \\ &+ \sum\left\{\hat{r}_t^{(0,4)} + 3\left(0.5\hat{\tilde{\theta}}_{1,n}^{(2)} - \beta\right)\hat{r}_t^{(0,3)} - 3\left(0.5\alpha\hat{\tilde{\theta}}_{1,n}^{(2)} + \beta\hat{\tilde{\theta}}_{0,n}^{(2)}\right)'\hat{r}_t^{(1,1)}\right\}. \end{aligned}$$

We can now apply Lemma C2(ii). The desired result follows from this.

(*ii*) We can apply the mean value theorem given A7(iv). That is, for some $(\bar{\theta}_{n,0}, \bar{\theta}_{n,1})$,

$$n^{-1/2}\sum \hat{s}_t = n^{-1/2}\sum s_t + n^{-1}\sum \nabla_{\theta^1}\bar{s}_t\sqrt{n}[(\hat{\theta}_{n,0}^{n'}, \hat{\theta}_{n,1}^n) - (\theta_0^{*'}, \theta_*)]',$$

where $\nabla_{\theta^1}\bar{s}_t := \nabla_{\theta^1}\bar{r}_t^{(0,4)} - 6\nabla_{\theta^1}\beta\bar{r}_t^{(0,3)} + 6\nabla_{\theta^1}\alpha'\bar{r}_t^{(1,1)}\beta - 6\nabla_{\theta^1}\alpha'\bar{r}_t^{(1,2)} + 3\nabla_{\theta^1}\alpha'\bar{r}_t^{(2,0)}\alpha$ and

$$\bar{r}_t^{(i,j)} := f_t^{(i,j)}(\bar{\theta}_{n,0}, \bar{\theta}_{n,1})/f_t(\bar{\theta}_{n,0}, \bar{\theta}_{n,1}).$$

Given this, applying B3(i to v) and the NAMXG condition in A7(iv) delivers the desired result.

(iii) This is implied by Lemma 3(iii).

(iv) Using the proof of Lemma 3(iv),

$$\tilde{L}_n^{(6)}(1/2, \hat{\theta}_{1,n}^n) = 1/2 \sum \{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t + 10(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} + 5(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)}\},$$

and we verified in Lemma C2(v) that $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} = O_p(n^{1/2})$ regardless of π. Thus, if we further show that $\sum\{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t = O_p(n^{1/2})$ and $(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)} = O_p(n^{1/2})$, then the desired result follows. We show these one by one. First, by (20),

$$\begin{aligned}\sum(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t =& (1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^5) \sum \hat{r}_t^{(0,6)} + 10(1 - (\hat{\tilde{\theta}}_{1,n}^{(1)})^3)\hat{\tilde{\theta}}_{0,n}^{(2)\prime} \sum \hat{r}_t^{(1,4)} \\ &+ 15(1 - \hat{\tilde{\theta}}_{1,n}^{(1)})\hat{\tilde{\theta}}_{0,n}^{(2)\prime} \sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(2)} + O_p(n^{1/2}).\end{aligned}$$

Note that the coefficient in from of each summand is $O_p(1)$ by Lemma C2(i and ii), and each summand is $O_p(n^{1/2})$ by A7(iv). Thus, $\sum\{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t = O_p(n^{1/2})$. Second, at π equal to 1/2,

$$\begin{aligned}\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)} =& -2\alpha' \sum \hat{r}_t^{(1,4)} + \alpha' \sum \hat{r}_t^{(2,1)}\hat{\tilde{\theta}}_{0,n}^{(2)} - 6\beta\hat{\tilde{\theta}}_{0,n}^{(2)\prime} \sum \hat{r}_t^{(2,1)}\hat{\tilde{\theta}}_{0,n}^{(2)} \\ &- 12\alpha' \sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(2)} - 6\sum \alpha'(\nabla_{\theta_0}\hat{\tilde{\theta}}_{0,n}^{(2)\prime}(\nabla_{\theta_0}^2 \hat{f}_t)\hat{\tilde{\theta}}_{0,n}^{(2)})/\hat{f}_t.\end{aligned}$$

Applying Lemma C2(ii) and A7(iv) completes the proof.

(v) Note that $\tilde{L}_n^{(7)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(6)} - \hat{k}_t^{(6)})\hat{g}_t + 15(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t^{(2)} + 20(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(3)} + 15(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(4)} + 6(\hat{h}^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(5)}\}$ by Lemma C2(iv). We examine each element in the RHS. First, after some algebra,

$$\begin{aligned}\sum(\hat{h}_t^{(6)} - \hat{k}_t^{(6)})\hat{g}_t =& (-10(\hat{\tilde{\theta}}_{1,n}^{(3)})^2 - 15\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{\tilde{\theta}}_{1,n}^{(4)} + 6\hat{\tilde{\theta}}_{1,n}^{(5)}) \sum \hat{r}_t^{(0,3)} + (20\hat{\tilde{\theta}}_{1,n}^{(3)} - 45(\hat{\tilde{\theta}}_{1,n}^{(2)})^2) \sum \hat{r}_t^{(0,5)} \\ &+ (-15(\hat{\tilde{\theta}}_{1,n}^{(2)})^3 + 60\hat{\tilde{\theta}}_{1,n}^{(3)}\hat{\tilde{\theta}}_{1,n}^{(2)} - 15\hat{\tilde{\theta}}_{1,n}^{(4)}) \sum \hat{r}_t^{(0,4)} - 15(\hat{\tilde{\theta}}_{1,n}^{(2)}) \sum \hat{r}_t^{(0,6)} \\ &+ (-20\hat{\tilde{\theta}}_{1,n}^{(3)}\hat{\tilde{\theta}}_{0,n}^{(3)} - 15\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{\tilde{\theta}}_{0,n}^{(4)} - 15\hat{\tilde{\theta}}_{0,n}^{(2)}\hat{\tilde{\theta}}_{1,n}^{(4)} + 12\hat{\tilde{\theta}}_{0,n}^{(5)}) \sum \hat{r}_t^{(1,2)} \\ &+ (-45\hat{\tilde{\theta}}_{0,n}^{(2)}(\hat{\tilde{\theta}}_{1,n}^{(2)})^2 + 60\hat{\tilde{\theta}}_{0,n}^{(3)}\hat{\tilde{\theta}}_{1,n}^{(2)} + 60\hat{\tilde{\theta}}_{0,n}^{(2)}\hat{\tilde{\theta}}_{1,n}^{(3)}) \sum \hat{r}_t^{(1,3)} \\ &+ (40\hat{\tilde{\theta}}_{0,n}^{(3)} - 90\hat{\tilde{\theta}}_{0,n}^{(2)}\hat{\tilde{\theta}}_{1,n}^{(2)}) \sum \hat{r}_t^{(1,4)} + 120\hat{\tilde{\theta}}_{0,n}^{(2)\prime} \sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(3)} \\ &- 45\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{\tilde{\theta}}_{0,n}^{(2)\prime} \sum \hat{r}_t^{(2,2)}\hat{\tilde{\theta}}_{0,n}^{(2)}.\end{aligned}$$

Each summand is $O_p(n^{1/2})$ by A7(iv), and the coefficients of the summands are $O_p(1)$. Thus, $\sum(\hat{h}_t^{(6)} - \hat{k}_t^{(6)})\hat{g}_t = O_p(n^{1/2})$ by Lemmas C2(i, ii, iii) and C3(i, ii). Second, $\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t^{(2)}$ and $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(3)}$ are $O_p(n^{1/2})$ by Lemmas C2(v) and C3(iv) respectively. Third, $\sum(\hat{h}^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(5)} = O_p(n^{1/2})$

by Lemma C3(v), because $(\hat{h}^{(1)} - \hat{k}_t^{(1)}) = \pi^{-1}(\alpha' \hat{m}_t + \beta \hat{k}_t)$. Fourth, if $\pi = 1/2$, $(\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) = -\hat{\tilde{\theta}}_{1,n}^{(2)} \hat{f}_t^{(0,2)} = -\hat{\tilde{\theta}}_{1,n}^{(2)} \pi^{-1}(\alpha' \hat{m}_t + \beta \hat{k}_t)$, so that $\sum (\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t^{(4)} = O_p(n^{1/2})$ by Lemma C2(vii). All the elements consisting $\tilde{L}_n^{(7)}(\pi, \hat{\theta}_{1,n}^n)$ are $O_p(n^{1/2})$, leading to the given claim.

(vi) As given in the proof of Lemma 5(d), $\tilde{L}_n^{(8)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi) \sum \{ (\hat{h}_t^{(7)} - \hat{k}_t^{(7)}) \hat{g}_t + 21(\hat{h}_t^{(5)} - \hat{k}_t^{(5)}) \hat{g}_t^{(2)} + 35(\hat{h}_t^{(4)} - \hat{k}_t^{(4)}) \hat{g}_t^{(3)} + 35(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) \hat{g}_t^{(4)} + 21(\hat{h}^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t^{(5)} + 7(\hat{h}^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(6)} \}$. Given this, note that $\sum (\hat{h}_t^{(5)} - \hat{k}_t^{(5)}) \hat{g}_t^{(2)}$ and $\sum (\hat{h}_t^{(4)} - \hat{k}_t^{(4)}) \hat{g}_t^{(3)}$ are $O_p(n^{1/2})$ by Lemmas C2(v) and C3(iv). Further, $\sum (\hat{h}^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t^{(5)}$ is $O_p(n^{1/2})$ because if $\pi = 1/2$, $(\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) = -\hat{\tilde{\theta}}_{1,n}^{(2)} \hat{f}_t^{(0,2)} = -\hat{\tilde{\theta}}_{1,n}^{(2)} \pi^{-1}(\alpha' \hat{m}_t + \beta \hat{k}_t)$, so that $\sum (\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) \hat{g}_t^{(5)} = O_p(n^{1/2})$ by Lemma C3(v). Thus,

$$\tilde{L}_n^{(8)}(1/2, \hat{\theta}_{1,n}^n) = 1/2 \sum \{ (\hat{h}_t^{(7)} - \hat{k}_t^{(7)}) \hat{g}_t + 35(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) \hat{g}_t^{(4)} + 7(\hat{h}^{(1)} - \hat{k}_t^{(1)}) \hat{g}_t^{(6)} \} + o_p(1).$$

Further, from $\tilde{M}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = 0$ and $\tilde{K}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = 0$,

$$\begin{aligned} \sum \hat{k}_t \hat{g}_t^{(6)} &= -\sum (\hat{k}_t^{(6)} \hat{g}_t + 15 \hat{k}_t^{(4)} \hat{g}_t^{(2)} + 20 \hat{k}_t^{(3)} \hat{g}_t^{(3)} + 15 \hat{k}_t^{(2)} \hat{g}_t^{(4)} + 6 \hat{k}_t^{(1)} \hat{g}_t^{(5)}) \\ &= -\sum (\hat{k}_t^{(6)} \hat{g}_t + 15 \hat{k}_t^{(2)} \hat{g}_t^{(4)}) + O_p(n^{1/2}), \end{aligned}$$

and

$$\begin{aligned} \sum \hat{m}_t \hat{g}_t^{(6)} &= -\sum (\hat{m}_t^{(6)} \hat{g}_t + 15 \hat{m}_t^{(4)} \hat{g}_t^{(2)} + 20 \hat{m}_t^{(3)} \hat{g}_t^{(3)} + 15 \hat{m}_t^{(2)} \hat{g}_t^{(4)}) \\ &= -\sum (\hat{m}_t^{(6)} \hat{g}_t + 15 \hat{m}_t^{(2)} \hat{g}_t^{(4)}) + O_p(n^{1/2}), \end{aligned}$$

where the last equalities are implied by Lemmas C2(v) and C3(iv, v) because $(\hat{h}^{(1)} - \hat{k}_t^{(1)}) = \pi^{-1}(\alpha' \hat{m}_t + \beta \hat{k}_t)$. Also, the terms involving $m_t^{(1)}$ vanish to zero by Lemma C2(iv). Thus, at $\pi = 1/2$, it follows that $\sum (\hat{k}_t^{(1)} - \hat{h}_t^{(1)}) \hat{g}_t^{(6)} = 2(\alpha' \sum \hat{m}_t \hat{g}_t^{(6)} + \beta \sum \hat{k}_t \hat{g}_t^{(6)}) = -2 \sum (\alpha' \hat{m}_t^{(6)} \hat{g}_t + \beta \hat{k}_t^{(6)} \hat{g}_t) - 30 \sum (\alpha' \hat{m}_t^{(2)} \hat{g}_t^{(4)} + \beta \hat{k}_t^{(2)} \hat{g}_t^{(4)}) + o_p(n)$. This implies that

$$\begin{aligned} \tilde{L}_n^{(8)}(1/2, \hat{\theta}_{1,n}^n) =& 0.5 \sum \{ (\hat{h}_t^{(7)} - \hat{k}_t^{(7)}) - 14(\alpha' \hat{m}_t^{(6)} + \beta \hat{k}_t^{(6)}) \} \hat{g}_t \\ & + 35 \sum \{ 0.5(\hat{h}_t^{(3)} - \hat{k}_t^{(3)}) - 3(\alpha' \hat{m}_t^{(2)} + \beta \hat{k}_t^{(2)}) \} \hat{g}_t^{(4)} + o_p(n). \end{aligned} \tag{22}$$

Given this, some algebra shows that

$$\begin{aligned} & 0.5 \sum \{ (\hat{h}_t^{(7)} - \hat{k}_t^{(7)}) - 14(\alpha' \hat{m}_t^{(6)} + \beta \hat{k}_t^{(6)}) \} \hat{g}_t = \sum \hat{r}_t^{(0,8)} - 28 \sum \hat{r}_t^{(0,7)} \beta + 367.5 \alpha' \sum \hat{r}_t^{(1,5)} \beta \\ & -28 \alpha' \sum \hat{r}_t^{(1,6)} - 1260 \alpha' \sum \hat{r}_t^{(2,3)} \alpha \beta + 210 \alpha' \sum \hat{r}_t^{(2,4)} \alpha + 420 \beta \sum \alpha' \nabla_{\theta_1} (\alpha' \nabla_{\theta_0}^2 \hat{f}_t \alpha) / \hat{f}_t \\ & -420 \sum (\nabla_{\theta_1}^2 (\alpha' \nabla_{\theta_0}^2 \hat{f}_t \alpha)) / \hat{f}_t + 105 \sum \alpha' (\nabla_{\theta_0}^2 (\alpha' \nabla_{\theta_0}^2 \hat{f}_t \alpha)) \alpha / \hat{f}_t + o_p(n). \end{aligned}$$

where the $o_p(n)$ remainder term is implied by Lemma B2(i, ii). Thus,

$$0.5n^{-1}\sum\{(\hat{h}_t^{(7)}-\hat{k}_t^{(7)})-14(\alpha'\hat{m}_t^{(6)}+\beta\hat{k}_t^{(6)})\}\hat{g}_t = R_*^{(0,8)}-28R_*^{(0,7)}\beta+367.5\alpha'R_*^{(1,5)}\beta$$
$$-28\alpha'R_*^{(1,6)}-1260\alpha'R_*^{(2,3)}\alpha\beta+210\alpha'R_*^{(2,4)}\alpha+420\alpha'(M_*^{(1,1)}\beta-M_*^{(1,2)})+105\alpha'M_*^{(2,0)}\alpha+o_p(n).$$

Also, after some algebra, it follows that

$$(0.5[\hat{h}_t^{(3)}-\hat{k}_t^{(3)}]-3[\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)}])/\hat{f}_t = (6\beta^2-0.5\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}+(1.5\hat{\hat{\theta}}_{1,n}^{(2)}-3\beta)\hat{r}_t^{(0,3)}+\hat{r}_t^{(0,4)}$$
$$-3(\beta\hat{\hat{\theta}}_{0,n}^{(2)}+0.5\alpha\hat{\hat{\theta}}_{1,n}^{(2)})'\hat{r}_t^{(1,1)}+3(\hat{\hat{\theta}}_{0,n}^{(2)}-\alpha)'\hat{r}_t^{(1,2)}-3\alpha'\hat{r}_t^{(2,0)}\hat{\hat{\theta}}_{0,n}^{(2)}$$
$$=\hat{s}_t+(6\beta^2-0.5\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}+o_p(1)=\hat{s}_t+(6\beta^2-0.5\hat{\hat{\theta}}_{1,n}^{(3)})(\alpha'\hat{r}_t^{(1,0)}+\beta\hat{r}_t^{(0,1)})+o_p(1),$$

where the last two equalities follow from Lemmas C2(i, ii) and the definition of $\hat{s}_t$. By Lemma C2(vii), it is also true that

$$\sum\hat{m}_t\hat{g}_t^{(4)}=\sum\hat{r}_t^{(1,0)}\hat{f}_t\hat{g}_t^{(4)}=o_p(n), \tag{23}$$

$$\sum\hat{k}_t\hat{g}_t^{(4)}=\sum\hat{r}_t^{(0,1)}\hat{f}_t\hat{g}_t^{(4)}=o_p(n), \tag{24}$$

implying that

$$\sum\{0.5(\hat{h}_t^{(3)}-\hat{k}_t^{(3)})-3(\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)})\}\hat{g}_t^{(4)}=\sum\hat{s}_t\hat{f}_t\hat{g}_t^{(4)}+o_p(n). \tag{25}$$

By some tedious algebra,

$$\begin{aligned}\hat{f}_t\hat{g}_t^{(4)} =& -0.5\hat{\hat{\theta}}_{1,n}^{(4)}\hat{r}_t^{(0,1)}-(1.5(\hat{\hat{\theta}}_{1,n}^{(2)})^2-2\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}-3\hat{\hat{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,3)}-\hat{r}_t^{(0,4)}-\hat{\hat{\theta}}_{0,n}^{(4)\prime}\hat{r}_t^{(1,0)}-3\hat{\hat{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,1)}\hat{\hat{\theta}}_{1,n}^{(2)}\\ &-6\hat{\hat{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,2)}-3\hat{\hat{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(2,0)}\hat{\hat{\theta}}_{0,n}^{(2)}+1.5(\hat{\hat{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,1)}+2\hat{r}_t^{(0,2)}+2\hat{\hat{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,0)})^2\\ =& -\hat{s}_t-0.5\hat{\hat{\theta}}_{1,n}^{(4)}\hat{r}_t^{(0,1)}-(1.5(\hat{\hat{\theta}}_{1,n}^{(2)})^2-2\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}-\hat{\hat{\theta}}_{0,n}^{(4)\prime}\hat{r}_t^{(1,0)}+o_p(1)\\ =& -\hat{s}_t-(6\alpha\beta^2-2\alpha\hat{\hat{\theta}}_{1,n}^{(3)}+\hat{\hat{\theta}}_{0,n}^{(4)})'\hat{r}_t^{(1,0)}-(6\beta^3-2\beta\hat{\hat{\theta}}_{1,n}^{(3)}+0.5\hat{\hat{\theta}}_{1,n}^{(4)})\hat{r}_t^{(0,1)}+o_p(1),\end{aligned}$$

where the last two equalities follow from Lemmas C2(i, ii) and the definition of $\hat{s}_t$. Thus, we can rephrase (23), (24), and (25) as

$$\begin{bmatrix}\sum\hat{m}_t\hat{g}_t^{(4)}\\ \sum\hat{k}_t\hat{g}_t^{(4)}\end{bmatrix}=-\begin{bmatrix}\sum\hat{s}_t\hat{r}_t^{(1,0)}\\ \sum\hat{s}_t\hat{r}_t^{(0,1)}\end{bmatrix}-\begin{bmatrix}\sum\hat{r}_t^{(1,0)}\hat{r}_t^{(1,0)\prime} & \sum\hat{r}_t^{(0,1)}\hat{r}_t^{(1,0)}\\ \sum\hat{r}_t^{(1,0)\prime}\hat{r}_t^{(0,1)} & \sum\hat{r}_t^{(0,1)}\hat{r}_t^{(0,1)}\end{bmatrix}\begin{bmatrix}\upsilon_{1,n}\\ \upsilon_{2,n}\end{bmatrix}=o_p(n) \tag{26}$$

and

$$\sum\{0.5(\hat{h}_t^{(3)}-\hat{k}_t^{(3)})-3(\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)})\}\hat{g}_t^{(4)}=-\sum\hat{s}_t^2-\upsilon_n'\sum\hat{s}_t\hat{r}_t^{(1)}+o_p(n) \tag{27}$$

respectively, where

$$\upsilon_n := \begin{bmatrix} \upsilon_{1,n} \\ \upsilon_{2,n} \end{bmatrix} := \begin{bmatrix} \hat{\tilde{\theta}}^{(4)}_{0,n} + 6(\beta^2 - \beta)\alpha \\ 0.5\hat{\tilde{\theta}}^{(4)}_{1,n} + 6(\beta^2 - \beta)\beta \end{bmatrix}.$$

This therefore implies that

$$\upsilon_n = -\begin{bmatrix} \sum \hat{r}^{(1,0)}_t \hat{r}^{(1,0)\prime}_t & \sum \hat{r}^{(0,1)}_t \hat{r}^{(1,0)}_t \\ \sum \hat{r}^{(1,0)\prime}_t \hat{r}^{(0,1)}_t & \sum \hat{r}^{(0,1)}_t \hat{r}^{(0,1)}_t \end{bmatrix} \begin{bmatrix} \sum \hat{s}_t \hat{r}^{(1,0)}_t \\ \sum \hat{s}_t \hat{r}^{(0,1)}_t \end{bmatrix} + o_p(1),$$

by (26). If this is substituted to (27), then $\sum\{0.5(\hat{h}^{(3)}_t - \hat{k}^{(3)}_t) - 3(\alpha' \hat{m}^{(2)}_t + \beta \hat{k}^{(2)}_t)\}\hat{g}^{(4)}_t$ is equal to

$$-\sum \hat{s}^2_t + \begin{bmatrix} \sum \hat{s}_t \hat{r}^{(1,0)}_t \\ \sum \hat{s}_t \hat{r}^{(0,1)}_t \end{bmatrix}' \begin{bmatrix} \sum \hat{r}^{(1,0)}_t \hat{r}^{(1,0)\prime}_t & \sum \hat{r}^{(0,1)}_t \hat{r}^{(1,0)}_t \\ \sum \hat{r}^{(1,0)\prime}_t \hat{r}^{(0,1)}_t & \sum \hat{r}^{(0,1)}_t \hat{r}^{(0,1)}_t \end{bmatrix} \begin{bmatrix} \sum \hat{s}_t \hat{r}^{(1,0)}_t \\ \sum \hat{s}_t \hat{r}^{(0,1)}_t \end{bmatrix} + o_p(n).$$

Given this, we can apply Lemma B3(i to iv). Then

$$n^{-1}\sum\{0.5(\hat{h}^{(3)}_t - \hat{k}^{(3)}_t) - 3(\alpha' \hat{m}^{(2)}_t + \beta \hat{k}^{(2)}_t)\}\hat{g}^{(4)}_t = -\Omega^{(s)} + o_p(1).$$

This completes the proof. ■

Proof of Theorem 5: (*i*) We can substitute $-10[\frac{(1-\pi)(1-2\pi)}{\pi^2}]^2\Omega^{(3)}$ to $\Xi^{(6)}(\pi)$ in the proof of Theorem 4(*i*).
(*ii and iii*) These are obvious. ■

References

ANDREWS, D. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

BICKEL, P. AND CHERNOFF, H. (1993): "Asymptotic Distribution of the Likelihood Ratio Statistic in a Prototypical Nonregular Problem," in J. Ghosh, S. Mitra, K. Parthasarathy and B. Prakasa Rao, eds. *Statistics and Probability*: *A Raghu Raj Bahadur Festschrift*, 83–96. New York: Wiley.

BIERENS, H. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58, 1443–1458.

CHEN. H. AND CHEN, J. (2001): "The Likelihood Ratio Test for Homogeneity in the Finite Mixture Models," *Canadian Journal of Statistics*, 29, 201–216.

Chernoff, H. and Lander, E. (1995): "Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 19–40.

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

Dacunha-Castelle, D. and Gassiat, E. (1999): "Testing the Order of a Model Using Locally Conic Parameterization: Population Mixtures and Stationary ARMA Processes," *The Annals of Statistics*, 27, 1178–1209.

Davies, R. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

Davies, R. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

Ghosh, J. and Sen, P. (1985): "On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results," in L. Le Cam and R. Olshen, eds. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, 2. Berkeley: University of California Press, pp. 789–806.

Goffinet, B., Loisel, P., and Laurent, B. (1992): "Testing in Normal Mixture Models When the Proportions are Known," *Biometrika*, 79, 842-846.

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

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

Heckman, J. and Singer, B. (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models of Duration Data," *Econometrica*, 52, 271–320.

Lindsay, B. (1995): *Mixture Models*: *Theory, Geometry and Applications*. Hayward: Institute for Mathematical Statistics.

McLachlan, G. and Basford, K. (1988): *Mixture Models*: *Inference and Applications to Clustering*. New York: Marcel Dekker.

Neyman, J. (1959): "Optimal Asymptotic Tests of Composite Statistical Hypotheses," *Probability and Statistics*, *The Harold Cramér Volume.* Ed. U. Genander, 213–234. Wiley, New York.

Neyman, J. and Scott, E. (1966): "On the Use of C(α) Optimal Tests of Composite Hypothesis," *Bulletin of the Institute of International Statististics*, 41, 477–497.

Ranga Rao, R. (1962): "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659-680.

Scott, D. (1973): "Central Limit Theorems for Martingales and for Processes with Stationary Increments Using a Skorokhod Representation Approach,“ *Advances in Applied Probability*, 5, 119–137.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

White, H. (1994): *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

White, H. (2001): *Asymptotic Theory for Econometricians*. Orlando: Academic Press.